agriculture

MDPI

Review

Multi-Target Harvesting Robots: Key Advances and Future Trends in Perception, Decision-Making, and Task Scheduling

Meng Wang, Wei Ji *, Donghao Wu, Zijie Shi, Hongjie Weng and Bo Xu

The School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China; 2112507181@stmail.ujs.edu.cn (M.W.); 15157701728@163.com (D.W.); s2438165101@163.com (Z.S.); 2222407033@stmail.ujs.edu.cn (H.W.); xubo@ujs.edu.cn (B.X.)

* Correspondence: jiwei@ujs.edu.cn

Abstract

Multi-target harvesting is an important direction for improving the efficiency of orchard harvesting robots and advancing their engineering deployment. Compared with single-target harvesting, multi-target harvesting is constrained by dense fruit distribution, branch and leaf occlusion, illumination variation, manipulator accessibility, harvesting-sequence planning, and multi-agent conflicts. These factors impose higher requirements on visual perception, state representation, decision-making, and task scheduling. This paper focuses on multi-target harvesting tasks in complex orchard environments. It first analyzes task characteristics and constraint sources and then reviews advances in visual perception and scene understanding for target recognition, attribute identification, spatial localization, and spatiotemporal consistency maintenance. Typical decision-making methods are summarized, including rule- and experience-based heuristics, cost-model-based optimization, and data-driven learning approaches. Progress in multi-arm and multi-robot harvesting systems is further reviewed in terms of perception sharing, task allocation, cooperative operation, and conflict avoidance. Continuous and stable perception under complex occlusion, unified scheduling-oriented state representation, real-time decision-making under strong constraints, collision avoidance in shared multi-agent workspaces, and distributed state synchronization remain key challenges. Overall, multi-target harvesting has shown potential to improve orchard operation efficiency, but efficient collaboration and engineering deployment in complex orchard environments still require further breakthroughs.

Keywords: multi-target harvesting; harvesting robots; visual perception; decision-making; task scheduling; multi-robot collaboration

Check for updates

Academic Editor: Maohua Xiao

Received: 5 August 2026
Revised: 31 August 2026
Accepted: 2 September 2026
Published: 6 September 2026

Copyright: © 2026 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.

1. Introduction

Global agriculture is undergoing an accelerated transformation toward sustainability, intensification, and intelligence, while pressures on food security, agricultural labor shortages, and rising labor costs are becoming increasingly prominent [1]. In fruit and vegetable harvesting, picking is a labor-intensive and time-sensitive operation that is evolving from traditional manual harvesting toward mechanized, intelligent, and collaborative approaches. Although manual harvesting has strong environmental adaptability and fruit-protection ability, it relies heavily on human experience and suffers from high work intensity, low efficiency, and poor stability, making it difficult to meet the demand for large-scale, efficient, and low-cost harvesting in modern orchards [2]. Mechanized harvesting can improve operational efficiency to some extent, but its perception and manipulation

capabilities remain limited, making it difficult to adapt to complex orchard scenarios characterized by scattered fruit distribution, severe occlusion, and heterogeneous maturity [3]. With the development of artificial intelligence, machine vision, and robotics, harvesting equipment has further evolved toward single-arm robots with visual perception, operation planning, and autonomous execution capabilities and has gradually expanded to more efficient paradigms such as multi-arm parallel harvesting and multi-robot collaborative harvesting [4,5]. This technological evolution not only provides a new approach to alleviating agricultural labor shortages, but also offers key support for efficient, standardized, and intelligent orchard harvesting. Therefore, harvesting robots have become a current research hotspot in agricultural robotics [6–8].

Although harvesting robot prototypes have demonstrated promising operational capabilities in laboratory environments, their engineering deployment still faces severe challenges in unstructured natural orchards [9]. Unlike structured grasping tasks on industrial production lines, natural orchard environments are characterized by strong uncertainty and complexity. Naturally grown fruits are often distributed in clusters or scattered patterns, accompanied by branch and leaf occlusion, drastic illumination changes, and wind disturbance, which introduce considerable uncertainty into target recognition, spatial localization, and stable execution. In addition, due to the kinematic structure of the robotic arm and physical obstacles in orchard spaces, the reachable workspace of the manipulator often cannot cover all visible targets. Under the combined effects of perceptual uncertainty and actuator constraints, robots cannot simply harvest all detected targets randomly, but must have decision-making and scheduling capabilities for multi-target tasks [10]. Thus, multi-target harvesting is not a simple repetition of single-target harvesting. It is a continuous operating process in which a robotic system evaluates, selects, and organizes multiple candidate fruits under shared temporal, spatial, and operational constraints to improve overall harvesting returns within a limited time budget. Without global understanding and dynamic decision-making capability in multi-target scenarios, robots may suffer from repeated operations, local congestion, or even collisions when facing dense fruit clusters due to improper target allocation [11–13]. Therefore, achieving reliable perception, reasonable prioritization, and efficient scheduling of multiple targets under limited spatiotemporal constraints has become a key issue restricting harvesting robots from moving from technical verification toward engineering application.

Existing studies generally follow two research directions. One focuses on target acquisition and state estimation in complex orchard scenarios, improving the reliability of target information through visual perception and scene understanding. The other emphasizes target selection and task organization under multiple constraints, aiming to improve continuous harvesting efficiency through decision-making strategies. From the perspective of engineering-oriented continuous operation, multi-target harvesting requires more than improving the recognition accuracy of a single perception algorithm or optimizing the motion path of a single robotic arm. Perception results must also be organized into state representations that support target prioritization, path selection, and task scheduling. Specifically, multiple candidate targets need to be evaluated jointly under factors such as maturity, occlusion level, localization confidence, manipulator accessibility, collision-avoidance constraints, and time cost. To increase operational throughput, multi-target harvesting can be implemented through different execution configurations. In single-arm continuous multi-fruit harvesting, one manipulator processes multiple fruit targets sequentially. Multi-arm harvesting enables parallel operation by coordinating multiple manipulators on the same robotic platform, whereas multi-robot harvesting extends perception sharing, task allocation, and coordinated execution across independent platforms. Accordingly, multi-target harvesting encompasses different execution forms ranging from sequential single-arm op-

eration to parallel multi-arm operation and cross-platform multi-robot collaboration. With the development of parallel and collaborative operating forms, perception sharing, task allocation, conflict suppression, and dynamic rescheduling have become new challenges that must be addressed for large-scale application.

Reviews in the field of robotic harvesting have covered a broad range of topics, including end-effector design, visual perception, fruit detection and localization, motion planning, and execution control [14–16]. However, the scope of individual reviews is still typically limited to a specific technical module or operational stage. In particular, vision-oriented reviews take visual perception and localization as their technical mainline, focusing mainly on detection accuracy, localization accuracy, and robustness. End-effector reviews focus on harvesting mechanisms and fruit–robot interaction, with emphasis on mechanical adaptability and fruit protection. Planning and control reviews, in contrast, center on motion generation, trajectory tracking, and execution, emphasizing motion feasibility and operational safety. Consequently, existing reviews provide limited synthesis of how perception information and operational constraints span different technical stages to jointly support multi-target decision-making, collaborative task organization, and dynamic scheduling. This review focuses on continuous multi-target harvesting in complex orchard environments and takes the coupling of perception, state representation, decision-making, and task scheduling as its technical mainline. It emphasizes the intrinsic relationships among task constraints, information flow, and multi-agent collaboration. The distinctive contribution of this review is an integrated system-level synthesis of the cross-stage technical chain from perception-information acquisition to collaborative scheduling, thereby complementing existing reviews that focus on individual technical topics or operational stages.

Standardized reporting of literature identification and screening can improve the transparency and reproducibility of the review process [17]. To review research progress on multi-target harvesting robots in perception, decision-making, task scheduling, and task allocation, this study conducted literature retrieval and screening using all databases on the Web of Science platform. The platform provides broad interdisciplinary coverage and supports standardized filtering by research area. The search was completed on 25 February 2026 and covered publications from 2016 to 25 February 2026. The exact search string entered into the database was (multi-target OR perception OR dense fruit OR dual-arm OR multi-robot decision making OR task scheduling OR task allocation OR multi-arm harvesting) AND (fruit picking robot OR harvesting robot). A total of 527 records were initially retrieved. Records were screened sequentially by publication year (11 excluded), language (4 excluded), document source (142 excluded), and research area (3 excluded), leaving 367 records for manual eligibility assessment. During the manual assessment, studies were retained if they directly addressed at least one core stage of the multi-target harvesting workflow and provided methodological or experimental evidence suitable for comparative analysis. Using this same criterion, 79 studies were excluded during manual assessment, and 288 articles were ultimately retained for this review. The detailed literature-identification and screening process, together with the screening criteria, is shown in Figure 1.

The overall framework of this review is shown in Figure 2. Section 2 analyzes the task characteristics and constraint sources of multi-target harvesting, including the effects of unstructured environments, fruit attributes, and robot-related constraints, and discusses the coupling between perception and decision-making. Section 3 reviews visual perception and scene understanding methods for multi-target harvesting, with emphasis on target recognition, attribute identification, spatial localization, and spatiotemporal consistency maintenance. Section 4 summarizes decision-making methods under multi-constraint conditions and compares the applicability of heuristic, optimization-based, and learning-

based strategies. Section 5 further discusses collaborative perception, task allocation, and conflict avoidance mechanisms in multi-arm and multi-robot harvesting systems. Section 6 summarizes the key challenges and future research directions. Finally, the paper concludes with a summary and the limitations of this review.

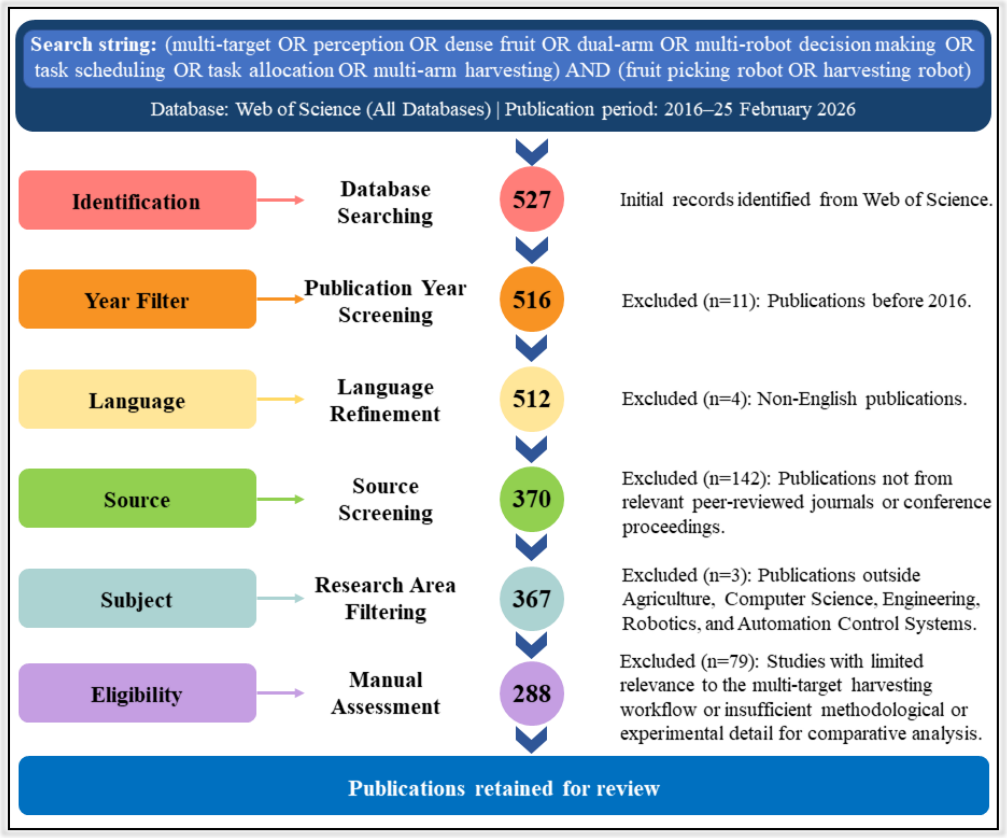


Figure 1. Literature search and screening process.

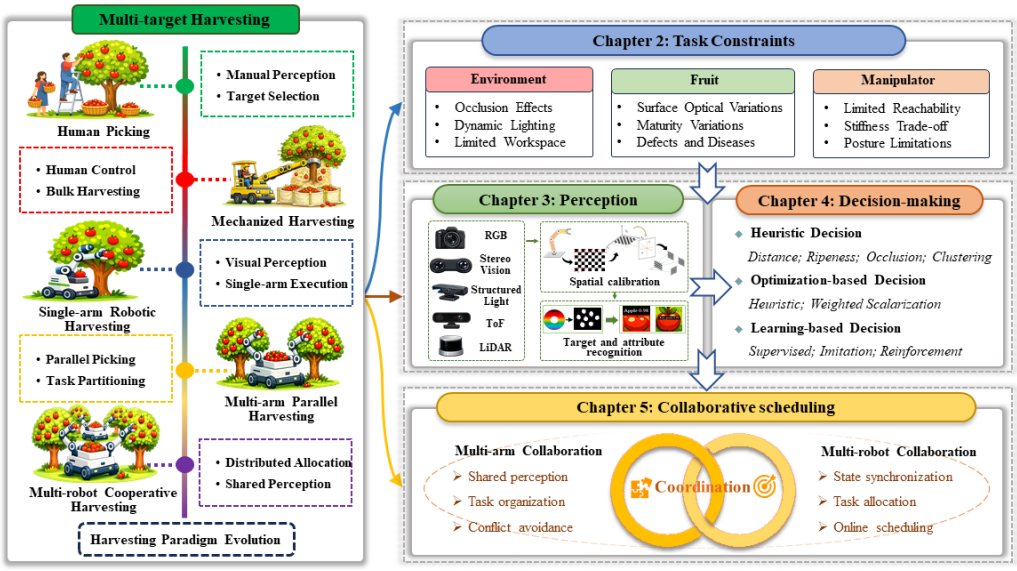


Figure 2. Overall research framework diagram.

2. Characteristics and Constraints of Multi-Target Harvesting Tasks

The efficiency of multi-target harvesting depends not only on the recognition accuracy of individual fruits or the success rate of single grasping actions, but also on whether robots can reasonably select, prioritize, and organize multiple candidate targets under

complex constraints. In natural orchard environments, fruit spatial distribution, branch and leaf occlusion, illumination variation, maturity differences, manipulator accessibility, and collision-avoidance requirements jointly form the constraint space of multi-target harvesting. These constraints affect not only the acquisition of target states by the perception system, but also whether a target is harvestable, reachable, and suitable for a given harvesting sequence. Accordingly, this section considers the operating characteristics of end-effectors, fruit characteristics, and manipulator constraints as task-level factors and analyzes how they shape the state representation, feasible decision space, and scheduling requirements of multi-target harvesting.

2.1. Harvesting Methods and Task Characteristics

Unlike the grasping of standardized rigid components in industrial settings, fruits to be harvested exhibit pronounced biological variability. Their pedicel attachment, peel resistance to damage, and maturity state affect not only end-effector operating strategies but also the expected execution time, success probability, and damage risk of harvesting tasks [18,19]. At present, harvesting methods can mainly be summarized into two categories: "clamping–twisting" and "suction/clamping–cutting", as shown in Figure 3.

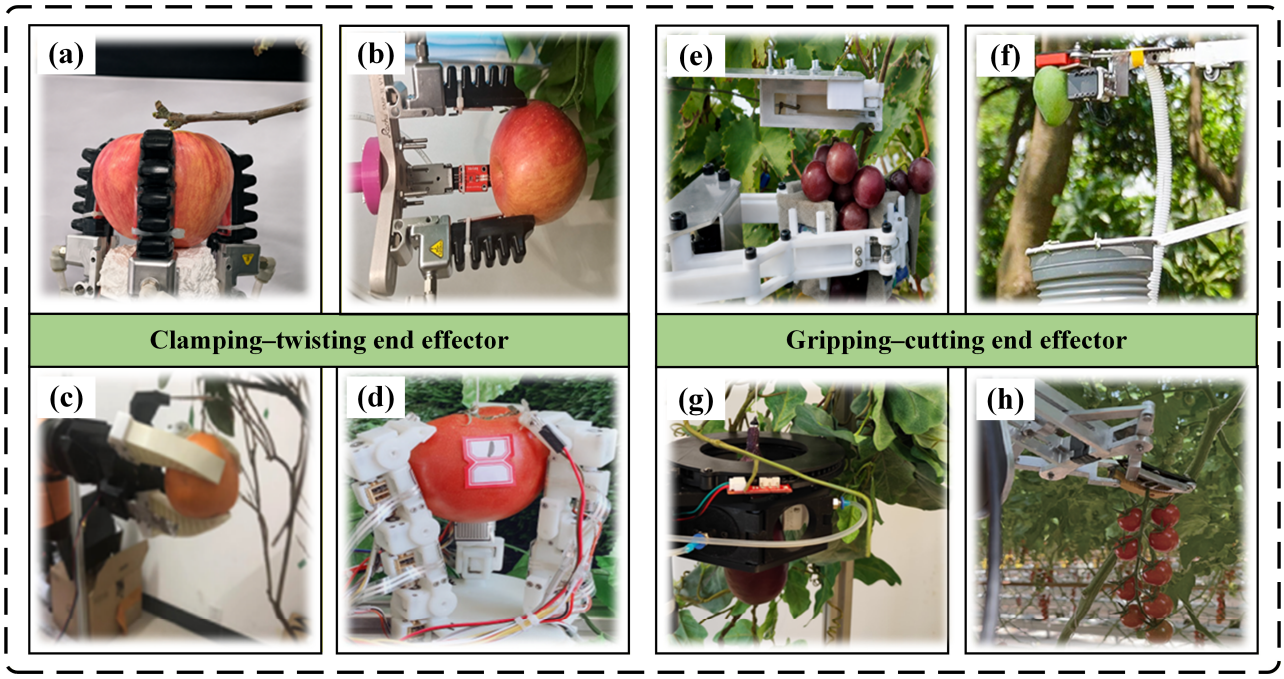


Figure 3. End-effectors used for different harvesting targets under two harvesting modes. (**a**–**d**) are examples of the application of clamping–twisting end-effectors in the harvesting of apples, pears, citrus, and tomatoes [19–22]; (**e**–**h**) are examples of the application of clamp-cutting end-effectors in the harvesting of grapes, mangoes, cucumbers and cherry tomatoes [23–26]. (Subfigures (**a**–**h**) were adapted from the corresponding references under the Creative Commons Attribution 4.0 License (CC BY 4.0)).

Different harvesting methods are mainly determined by crop stem connection characteristics. For spherical fruits such as apples and citrus, whose stems are strongly connected and whose peels are relatively tough, twisting-based harvesting is commonly used; its key lies in separating the stem through gripping friction and wrist torque while controlling the separation force and angle [27]. In contrast, crops such as tomatoes, grapes, and strawberries, which have soft peels or short stems, are more suitable for a "fix first and then cut" mode, requiring higher accuracy in picking-point localization, end-effector intervention, and stem recognition [28]. To improve multi-target harvesting efficiency, some studies have

attempted synchronous or continuous harvesting through mechanical vibration or specialized harvesting mechanisms [29,30], while others have integrated separation, feeding, and collection through in-situ harvesting devices [31]. However, due to the strict requirements of high-value fruits for appearance quality and damage control, refined single-fruit cyclic operation remains the mainstream mode in current automated harvesting.

Harvesting tasks are also constrained by the risk of physical damage. Fruits are viscoelastic biological objects; their internal tissues may be damaged before visible peel damage occurs, and their resistance to external force decreases with increasing maturity. Therefore, fixed gripping-force parameters are difficult to apply to all targets [32]. Non-destructive harvesting usually requires the integration of fruit biomechanical characteristics, flexible end-effector structures, and force feedback regulation, while damage risk can be reduced through gripping-force threshold optimization and slip suppression [33–35]. Meanwhile, robots need to recognize maturity state and infer relevant physical properties based on spatial localization, so as to dynamically adjust gripping force or operation strategy [36,37]. In multi-target scenarios, the delay or failure of each perception, localization, separation, and placement step in single-fruit cyclic operation can accumulate into substantial efficiency loss. The harvesting mode, expected execution time, success probability, and damage risk associated with each target should be incorporated into target-priority determination and harvesting-sequence planning. Therefore, task efficiency depends not only on mechanical structure and operation mode, but also on the real-time performance of environmental perception and decision planning [38].

2.2. Unstructured Environmental Characteristics and Visual Constraints

The influence of unstructured orchard environments on visual perception in multi-target harvesting is mainly reflected in two aspects: occlusion and illumination variation [39]. First, occlusion caused by leaves, branches, supporting structures, and neighboring fruits can lead to incomplete fruit contours, reduced visible surfaces, and even invisible key operation parts such as stems or cutting points. This partial observability reduces the accuracy of target centroid, pose, and picking-point estimation, while increasing the collision risk during end-effector approach [40]. In multi-target scenarios, a single viewpoint is often insufficient to capture complete information of all candidate targets, making multi-view perception an important way to alleviate occlusion. When evaluating a multi-camera system in a V-trellis apple orchard, Villacrés et al. [41] found that the visible-fruit coverage of a single-camera view was only 88.3%, whereas dual-camera cooperation increased the coverage to 97.5%, indicating that viewpoint configuration directly affects fruit visibility and localization reliability. Figure 4 shows the spatial layout of this multi-camera configuration.

When the task shifts from fruit detection to picking-point localization and action execution, occlusion becomes more critical. Key structures such as stems and cutting points are small and easily occluded by leaves or other fruits, and their detection failure directly affects the intervention path and operational safety of the end-effector [42]. To improve target visibility, some studies have reduced leaf interference through active de-occlusion or environmental interaction. He et al. [43] integrated a fan system in strawberry harvesting to reduce occlusion from leaves and vines, increasing the harvesting success rate from 58.1% to 73.9%. Yao et al. [44] exposed occluded regions through a safe leaf-pushing strategy, thereby improving the stability of shape completion and pose estimation. In addition, mutual occlusion among fruits can cause boundary adhesion and instance confusion, further affecting target recognition, localization, and harvesting-sequence judgment [45]. The relative relationship between the occluder and the occluded target also affects recogni-

tion stability [46]. In close-range operation scenarios, mutual occlusion can also prolong localization and harvesting time, reducing continuous operation efficiency [47].

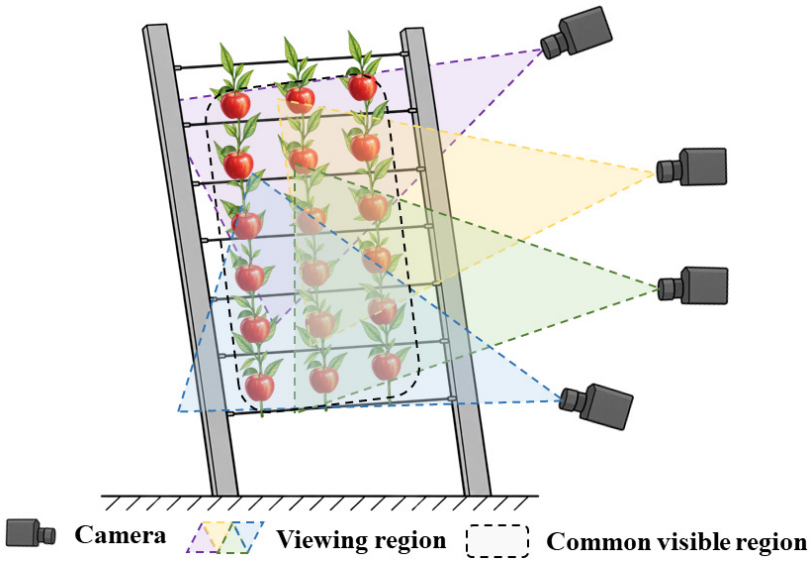


Figure 4. Graphical representation of the spatial layout of multiple cameras.

Second, illumination variation can cause target appearance drift and visual localization fluctuation. Direct sunlight, shadows, overexposure, specular reflection, and color-temperature changes in natural orchards can alter the contrast between fruits and backgrounds, reducing model robustness across different times and weather conditions [48]. Illumination nonuniformity should be regarded as a long-term background condition rather than occasional interference, because it continuously affects the detectability and localization stability of mature fruits [49]. Zuo et al. [50] found in field target detection that different illumination conditions significantly changed target surface features and background contrast, thereby increasing detection instability. Zhuang et al. [51] also found in citrus orchard field tests that false detections could still occur in regions with drastic illumination changes, indicating that dynamic lighting disrupts visual feature consistency. Since visual perception results are usually directly involved in grasping-pose and operation-point generation, illumination-induced localization errors may further propagate to the execution stage [52]. In a field evaluation of an orange harvesting robot, Zeeshan et al. [53] reported that poor lighting conditions increased fruit damage risk, which was associated with end-effector grasping-orientation deviation caused by unstable visual feedback.

Overall, occlusion in unstructured environments mainly causes geometric information loss and invisibility of key structures, whereas illumination variation mainly leads to appearance drift and localization fluctuation. Together, these factors reduce the stability of target detection, picking-point localization, and continuous execution in multi-target harvesting.

2.3. Fruit Properties and Biological Constraints

As non-standardized biological objects, fruits vary in surface optical properties, ripening progression, and health status. These differences jointly affect perception stability, harvesting-value assessment, and target-priority determination in multi-target harvesting. Naturally grown fruits show substantial intraclass variation in color, texture, gloss, wax bloom, and clustered morphology, introducing uncertainty into target recognition, maturity determination, and abnormality detection. Table 1 summarizes the attribute characteristics of typical harvesting objects and their effects on perception.

Table 1. Different fruit attributes and their impact on perception.

Feature Category	Fruit Type	Main Color/Gloss Characteristics	Ripeness Change Characteristics	Common Diseases or Abnormalities
Specular reflection	Apple	Smooth peel, pronounced specular highlights, red/yellow/green background color.	Color transition from green to yellow or red, improved color uniformity.	Scuffing, bruising, depressions, spots, browning, surface defects [54].
	Cherry	Smooth peel, strong highlights, bright red to deep red coloration at maturity.	Progressive darkening from light red to deep red, increased color saturation.	Cracking, scuffing, surface defect spots [55].
	Tomato	Smooth peel, pronounced highlights, green at immature stage, red at mature stage.	Progression from green to breaker to pink to red.	Cracking, lesions, rot spots, surface damage [56]
	Sweet Pepper	Waxy peel, pronounced highlights, green/yellow/orange/red coloration.	Gradual transition from green to yellow or orange and then red.	Sunscald, scuffing, lesions, rot [57]
Complex Texture	Strawberry	Prominent achenes, fine surface texture, fragmented texture pattern, more pronounced red coloration at maturity.	Color transition from white or light red to red and deep red, enhanced surface contrast.	Bruising, mold rot, lesions, malformed fruit [58]
	Citrus	Distinct oil-gland texture, predominantly diffuse reflection, orange–yellow or yellow–green coloration.	Color transition from green to orange–yellow, often with uneven coloration.	Peel blemishes, disease spots, sunken defects [59]
	Cucumber	Ridges, grooves, spines, wart-like protrusions, predominantly green coloration, uneven gloss.	Transition from dark green to light green or yellow with aging, rougher peel texture.	Disease spots, mildew, rot spots, malformation, peel damage [60]
Clustered with Bloom	Grape	Clustered berries, severe mutual occlusion, wax bloom, cultivar-dependent coloration.	Color transition from green to red-purple or black during veraison, uneven ripening within clusters.	Spoiled berries, mold rot, compression damage, occlusion-related missed detection [61]
	Blueberry	Bluish-gray matte appearance, wax bloom, small berry size, high susceptibility to occlusion.	Color transition from green to purple and blue, freshness associated with bloom integrity.	Bloom damage, darkened appearance, rotten berries, uneven ripening [62]

As shown in Table 1, different fruit attributes lead to different visual constraints. Smooth fruits such as apples and tomatoes tend to produce specular reflection under strong illumination, which can obscure surface details and interfere with boundary localization [63]. Fruits such as grapes and blueberries are often covered with wax bloom and distributed in clusters, making them prone to blurred edges, berry occlusion, and instance-segmentation difficulty [62]. Green crops such as cucumbers and peppers have colors similar to leaves and branches, making it difficult to distinguish targets from backgrounds

using color features alone; texture or geometric features are often needed to improve recognition reliability [64].

Maturity and health status further determine whether a target has harvesting value. Related studies have shown that fruit growth and ripening are often accompanied by changes in morphology or posture, and different cultivars exhibit inconsistent changes in color, size, texture, and gloss, making fixed thresholds or single-feature criteria difficult to apply in complex orchard scenarios [65]. For some crops, maturity is mainly reflected in color change, whereas for others it is more associated with size growth, color deepening, or changes in surface optical properties [66]. For example, increased gloss in mature tomatoes and translucency during grape veraison indicate that maturity assessment often requires the integration of multiple cues, such as color, texture, and optical characteristics [67]. Since maturity states lack clear boundaries, color-transition stages and uneven ripening can easily lead to misclassification, further increasing the difficulty of harvestability assessment [68]. Meanwhile, abnormal conditions such as mold spots, cracking, bruising, malformation, and rot can change fruit color distribution and morphological features, while directly reducing harvesting value [69–71]. Therefore, multi-target harvesting robots need not only determine whether fruits are present, but also identify their maturity, health status, and harvesting value. These attributes determine both whether a target should be selected and how it should be prioritized relative to other candidate targets, thereby providing direct evidence for target screening and task-priority ranking.

2.4. Robotic Arm Kinematics and Spatial Structure Constraints

In modern orchards, harvesting robots usually move along fixed row paths, and the motion trajectory of the mobile base is constrained, making it difficult to flexibly adjust the operating position as human pickers do. Therefore, when the mobile base position is relatively fixed, target approach and grasping mainly depend on the robotic arm. However, branches, leaves, and supporting structures in orchards divide the manipulator workspace into multiple irregular regions, causing the originally continuous operating space to become discretized. Even if some fruits are within the maximum geometric reach of the manipulator, they may still be difficult to access in practice if no collision-free path is available. Such spatial constraints significantly reduce the effective workspace of the manipulator and increase collision risk when the robotic arm operates deep inside the canopy [72,73].

To reach fruits located inside the canopy or at higher positions, the manipulator usually needs a large working range. However, increasing arm length often reduces structural rigidity, increases system inertia, and weakens dynamic stability, making the end-effector more prone to vibration during high-speed motion or sudden stopping, thereby reducing positioning accuracy [74]. Different manipulator configurations show clear differences in reachable workspace, structural rigidity, control complexity, and applicable scenarios, as shown in Table 2.

Table 2. Comparison of accessibility characteristics of different types of robotic arm structures.

Manipulator Configuration	Range of Degrees of Freedom	Reachable Workspace Morphology	Structural Rigidity	Control Complexity	Typical Application Scenarios
Serial	6–7	Large, continuous	Moderate	High	Fruit harvesting in orchards with wide distribution and complex occlusion [28]
Parallel	3–6	Medium to small	High	Moderate	Fixed-point, high-precision harvesting [75]

Table 2. *Cont.*

Manipulator Configuration	Range of Degrees of Freedom	Reachable Workspace Morphology	Structural Rigidity	Control Complexity	Typical Application Scenarios
SCARA	4	Planar	High vertical rigidity	Low	Harvesting of low-growing crops in greenhouses [76]
Hybrid	6–8	Customizable, suitable for complex spaces	Relatively high	High	Multi-manipulator collaboration in dynamic environments [77]

As shown in Table 2, different manipulator configurations have distinct characteristics in terms of reachable workspace, structural rigidity, and control complexity. Serial manipulators have a large range of motion and can flexibly enter the canopy for obstacle avoidance, making them suitable for orchard harvesting scenarios with wide fruit distribution and complex occlusion. Parallel and SCARA manipulators have higher rigidity and positioning accuracy, but their workspaces are relatively constrained, making them more suitable for greenhouses, low-growing crops, or fixed-point high-precision operations. Hybrid manipulators attempt to achieve a compromise between flexibility and rigidity, but their structural design and control complexity are relatively high. Meanwhile, to adapt to diverse fruit growth postures, harvesting manipulators usually require higher degrees of freedom. However, increased degrees of freedom also introduce problems such as joint limits, posture replanning, and kinematic singularities. When the manipulator approaches joint limits or enters special configurations, its motion capability may decrease or even become locally "locked", affecting path execution during continuous harvesting [78]. Therefore, in multi-target harvesting, robots need to determine not only whether a target is visible, but also whether it is truly harvestable by considering manipulator reachability, obstacle-avoidance space, and pose feasibility [79]. These kinematic and spatial constraints determine not only whether a target is executable, but also which execution unit should handle it when multiple manipulators or robots are available and in what order it should be processed. They therefore directly constrain target screening, task allocation, and scheduling in multi-target harvesting and constitute important physical boundaries of multi-target harvesting decision-making.

2.5. Analysis of the Coupling Relationship Between Perception and Decision-Making

In traditional robotic systems, perception and decision-making are usually divided into relatively independent modules. The system first obtains environmental and target states and then makes task decisions based on the perception results. This serial mode is clear in structured environments or single-target tasks, but it becomes limited in multi-target harvesting scenarios. The operating characteristics of end-effectors, fruit attributes, and manipulator constraints jointly determine the task states, feasible target set, and execution costs on which subsequent decision-making and scheduling rely. Because candidate fruits are numerous and widely distributed and are affected by occlusion, attribute variations, and localization errors, perception results are no longer merely inputs to decision-making. Detected targets, attributes, three-dimensional locations, picking points, and local obstacle information cannot be directly used for task decision-making. Instead, they need to be organized into decision-oriented state representations. These state representations associate each candidate target with its harvestability, spatial relationships, accessibility, and estimated execution cost, thereby defining the feasible candidate-target set and decision space for target selection, harvesting-sequence planning, task allocation, and path generation. For example, target position errors may make manipulator paths infeasible,

while incorrect maturity or harvestability estimation may lead to ineffective operations or reduced efficiency [80–82].

Meanwhile, the decision-making process also changes the system's demand for perceptual information. In the global task-planning stage, the robot focuses more on target number, spatial distribution, and accessibility, whereas the execution stage requires finer information such as target pose, picking-point position, and surrounding geometric constraints. In natural orchards, plant deformation, illumination fluctuations, and changes in viewing angle can cause perception results to drift over time. Therefore, multi-target harvesting requires a closed loop in which environmental observations are continuously updated, transformed into task-relevant states, and refined according to decision-making requirements. As shown in Figure 5, perception and decision-making form a dynamic coupling relationship: perception results determine the reliability of the decision space, while decision strategies determine the type, accuracy, and update frequency of required perceptual information [83–86].

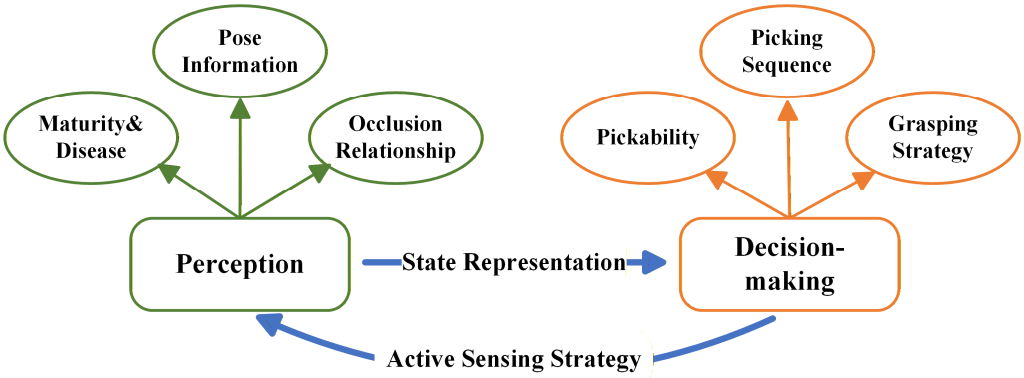


Figure 5. Schematic diagram of the coupling relationship between perception and decision-making in the harvesting robot.

3. Visual Perception and Scene Understanding

Visual perception in multi-target harvesting needs to provide target attributes, spatial positions, operation poses, and environmental constraints to support target selection, path planning, and task scheduling [87–89]. Early methods mainly relied on manually designed color and texture features and showed limited adaptability to illumination variation and occlusion. With the development of deep learning, three-dimensional vision, and multi-target tracking, perception has expanded from static two-dimensional recognition to scene understanding, operation-point localization, and spatiotemporal consistency maintenance, thereby better supporting continuous harvesting.

3.1. Perception Architecture and Sensor Configuration

The perception capability of harvesting robots depends on sensor configuration, imaging modality, and coordinate mapping accuracy. In complex agricultural environments, sensor selection and calibration directly determine whether visual outputs can be reliably transformed into spatial information for target localization and manipulator control.

3.1.1. Sensor Categories and Imaging Modes

The selection of imaging modality requires a trade-off among information richness, environmental adaptability, measurement accuracy, and system cost. RGB cameras provide rich color and texture information and are widely used for fruit recognition and attribute classification, but they cannot provide depth information for manipulator localization. RGB-D cameras combine color and distance information and have therefore become common in harvesting robots [90–92]. Among RGB-D sensors, structured-light and ToF cameras

provide high accuracy and real-time performance under controlled illumination, but their active infrared signals are easily disturbed by strong outdoor sunlight [93]. Binocular stereo vision is more adaptable to outdoor illumination because it is a passive sensing method, although weak-texture fruit surfaces and occlusion may still cause matching failure [94]. LiDAR and thermal imaging can further improve geometric perception or target discrimination in specific scenarios, but their cost and integration complexity limit large-scale deployment [95]. The characteristics of mainstream sensors are summarized in Table 3.

Table 3. Comparison of performance characteristics and applicable scenarios of mainstream vision sensors for harvesting robots.

Sensor Type	Imaging Principle	Core Advantages	Main Limitations	Environmental Adaptability	Typical Application Scenarios
RGB camera	Optical imaging	Rich color texture information, high resolution, low cost	No depth information, easily affected by illumination changes	Moderate	Fruit recognition, maturity classification, disease detection
Binocular stereo vision	Stereo vision principle	Passive sensing, strong resistance to illumination interference, large measurement range	Difficult depth estimation in weak-texture regions or under occlusion, high computational cost	High	Outdoor large-area navigation, fruit localization
Structured-light camera	Active projection and decoding	High precision, good edge detail, strong real-time performance	Sensitive to strong illumination interference, short measurement range	Low	Indoor greenhouse precision grasping, short-range obstacle avoidance
ToF camera	Time-of-flight	High frame rate, low computational cost, no texture dependence	Low resolution, susceptible to multipath interference and strong illumination	Moderate	Rapid obstacle avoidance, real-time hand–eye servoing
LiDAR	Laser ranging	Long sensing range, insensitive to illumination, high accuracy	Lack of texture information, high cost, sparse data	Very high	Picking navigation, geometric modeling of tree rows

Table 3 summarizes the imaging principles, strengths, limitations, environmental adaptability, and typical deployment scenarios of mainstream vision sensors for harvesting robots. The performance trade-offs of these sensors are closely related to their operating conditions. Structured-light cameras rely on active projection and are most effective under controlled illumination and short working distances; they are therefore assigned a Low environmental adaptability rating. RGB cameras and ToF cameras are rated Moderate because their practical performance can be constrained by illumination variation, strong sunlight, multipath interference, or sensing-range limitations. By contrast, binocular stereo vision is rated High owing to its passive sensing mechanism and better suitability for outdoor illumination, although weak texture and occlusion can still compromise depth estimation. LiDAR receives a very high rating because it provides illumination-insensitive ranging over a broader sensing range and remains effective in unstructured orchard environments. These differences are also reflected in their typical applications: RGB cameras are widely used for fruit recognition and maturity classification, structured-light and ToF cameras are more suitable for indoor or greenhouse precision operations, binocular vision is commonly

used for outdoor fruit localization, and LiDAR is particularly suitable for orchard navigation and geometric modelling. Consequently, no single sensor can simultaneously satisfy the requirements of target recognition, depth measurement, illumination robustness, and real-time deployment. Therefore, sensor configuration for multi-target harvesting should be selected according to crop type, operating environment, and task requirements, so that perception outputs can support target localization, occlusion judgment, and subsequent decision planning.

3.1.2. Sensor Calibration and Coordinate System Mapping

Accurate coordinate mapping is the basis for transforming visual perception results into manipulator motion commands. Visual data are usually represented as pixel coordinates or relative depth values and need to be mapped into the robot base or end-effector coordinate system through camera imaging models and robot coordinate transformations [96]. Therefore, vision system calibration mainly includes camera intrinsic and distortion parameter estimation, as well as extrinsic or hand–eye calibration between the camera and robot coordinate systems.

In engineering applications, camera intrinsic calibration is usually performed using planar calibration boards. The planar template calibration framework proposed by Zhang [97] estimates camera intrinsics, distortion parameters, and view-dependent extrinsics by observing checkerboard corners from multiple views and minimizing reprojection error, as shown in Figure 6. However, in orchard environments, terrain unevenness, branch occlusion, and illumination variation make it difficult to deploy calibration boards. Therefore, some studies combine pre-calibration with multi-view parameter optimization [98], while others use SfM and BA frameworks to estimate focal length, distortion, and camera pose without dedicated calibration objects [99]. Recent self-supervised video-based calibration methods also provide new possibilities for calibration-object-free deployment in agricultural scenarios [100].

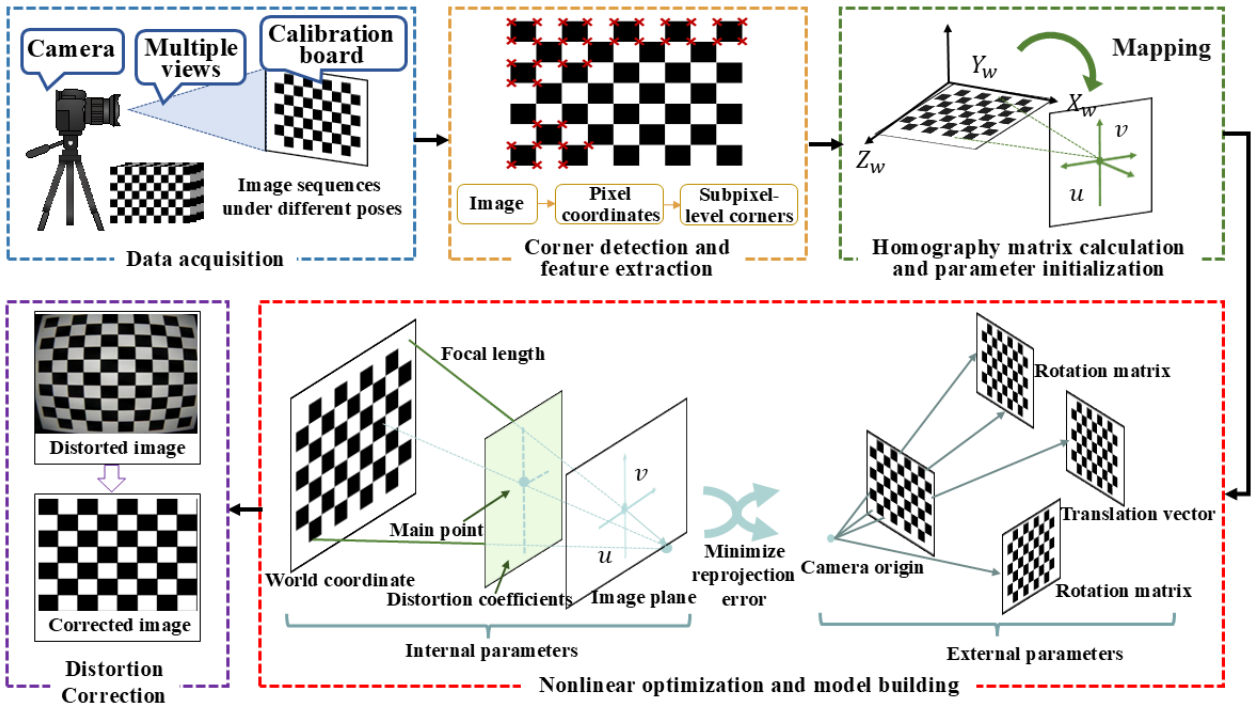


Figure 6. Planar calibration framework.

Hand–eye calibration establishes the spatial relationship between the camera coordinate system and the robot kinematic coordinate system, and its core problem can be

formulated as the classical AX = XB equation [101]. For harvesting robots, researchers have combined calibration methods with specific sensors and task scenarios. For example, Zhang et al. [102] proposed a ToF-camera-based hand–eye calibration method to reduce the effects of illumination and depth noise. Jiang et al. [103] used a reference block to obtain the hand–eye pose relationship through contact and visual measurement, as shown in Figure 7.

Compared with industrial scenarios, orchard harvesting is often accompanied by ground bumps, manipulator vibration, and long-term extrinsic drift. Existing studies therefore introduce ICP constraints, spatiotemporal joint calibration, or online adaptive visual servoing to improve robustness under field conditions [104–107]. Overall, visual calibration for harvesting robots is shifting from static offline calibration toward robust optimization and online adaptation for long-term field operation.

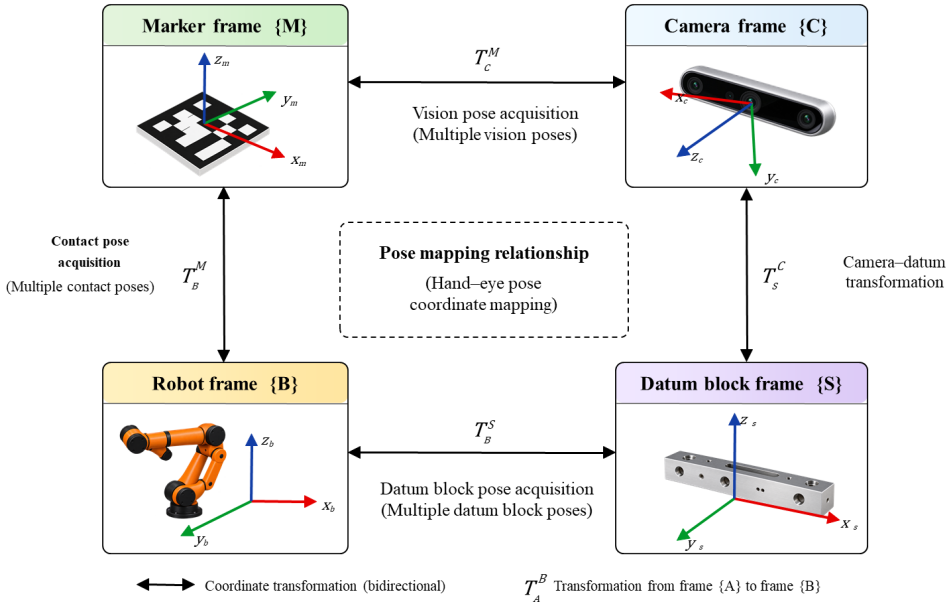


Figure 7. Schematic diagram of hand–eye pose coordinate mapping based on reference block.

3.2. Fruit Identification and Attribute Classification

Visual perception is a key link between the physical environment and harvesting decision-making. For multi-target harvesting robots, the perception system needs not only to identify fruit targets from complex backgrounds, but also to determine their maturity, health status, and key operation parts, thereby providing a basis for target screening, spatial localization, and harvesting-sequence planning.

3.2.1. Harvesting Target Identification Methods

Illumination fluctuation, target overlap, and background interference in unstructured orchard environments require harvesting target recognition to consider real-time performance, robustness, and semantic representation simultaneously. According to the perceived object, existing harvesting target recognition tasks can be extended from whole-fruit detection to key-structure recognition, as shown in Figure 8.

Early harvesting target recognition mainly relied on handcrafted features such as color thresholds and shape descriptors, which were effective in scenes with clear contrast between fruits and backgrounds, but showed limited generalization under similar colors, illumination variation, or severe occlusion [108,109]. To improve foreground–background separation, some studies introduced preprocessing methods such as image enhancement or color-space transformation; however, these pixel-level methods still lack stability under drastic natural illumination changes [110]. With the development of convolutional

neural networks, deep learning methods based on frameworks such as YOLO, SSD, and R-CNN have gradually become mainstream, improving fruit detection capability in complex backgrounds through hierarchical feature extraction [111–113].

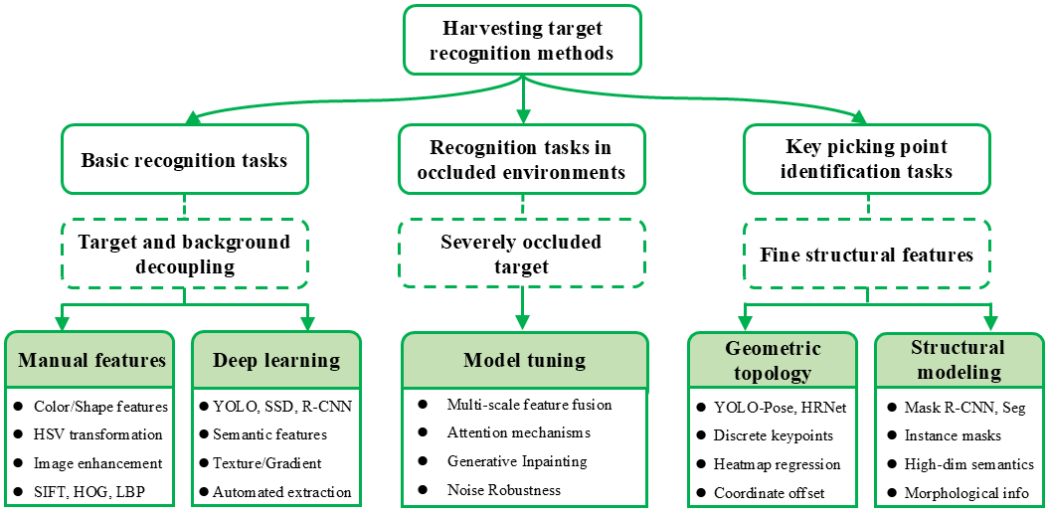


Figure 8. Different task types for harvesting target identification.

To address local feature loss caused by branch and leaf occlusion, researchers have further enhanced model adaptability to incomplete targets and complex backgrounds through multi-scale feature fusion, attention mechanisms, and generative completion [114–116]. These methods can improve the detection stability of occluded fruits to some extent, but in dense multi-target scenarios, boundary adhesion, instance confusion, and missed detection of small targets still affect subsequent localization and task ordering.

In addition to whole-fruit detection, harvesting tasks also require the recognition of key operation parts such as fruit stems, calyxes, and cutting points, because these structures directly affect the approach direction and operation mode of the end-effector [9]. For fruit stems that are small, slender, and easily confused with the branch and leaf background, existing studies have mainly formed two technical routes. One is geometric topology modeling represented by keypoint detection, such as improved YOLO-Pose and HRNet, which predicts discrete keypoints such as connection points and cutting points to provide constraints for cutting-position calculation [117–119]. The other is pixel-level structure modeling represented by instance segmentation, such as Mask R-CNN and YOLO-seg, which generates masks of fruit stems or target regions to provide more complete morphological information and helps recover the spatial continuity of key structures under complex occlusion [120–122]. Overall, harvesting target recognition is shifting from simple fruit detection toward joint perception of targets and operation-related key structures.

3.2.2. Fruit Attribute Identification Methods

After fruit targets are accurately detected, further determining their maturity and health status is important for target screening and priority determination in multi-target harvesting. Unlike target detection, which mainly focuses on whether fruits are present, attribute recognition focuses on whether fruits are mature, healthy, and worth harvesting and is therefore usually used as a preliminary filtering condition for harvesting decisions. Among fruit attributes, maturity determination is the most widely used direction. Visible-image-based methods usually distinguish mature and immature targets using color, texture, and morphological features. For example, Wang et al. [123] proposed ASFA-net to generate fruit masks after strawberry detection and used the proportion of red pixels calculated from the hue component in HSV space to quantify strawberry maturity, as shown in Figure 9.

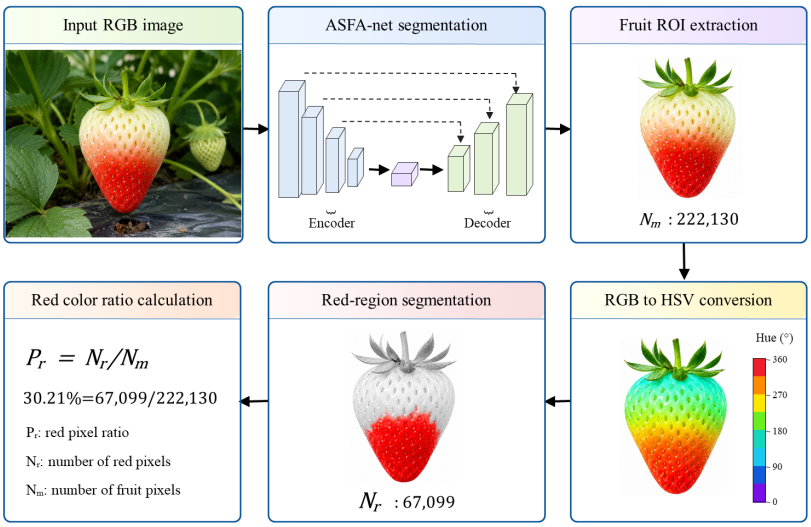


Figure 9. Calculation of the red ratio of strawberry skin based on the ASFA-net model and HSV space.

In addition to color features, texture features can also be used for maturity evaluation. Wang et al. [124] extracted and enhanced citrus surface texture features using LBP- and R-LBP-based image processing methods, thereby improving the discrimination of fruit ripeness, as shown in Figure 10. Liu et al. [125] combined color and gradient texture features from RGB images with HOG descriptors to identify mature tomatoes. However, such visible-feature-based methods are easily affected by illumination variation, occlusion, and cultivar differences and are often more suitable for judging whether a target is harvestable than for stably describing a continuous ripening process.

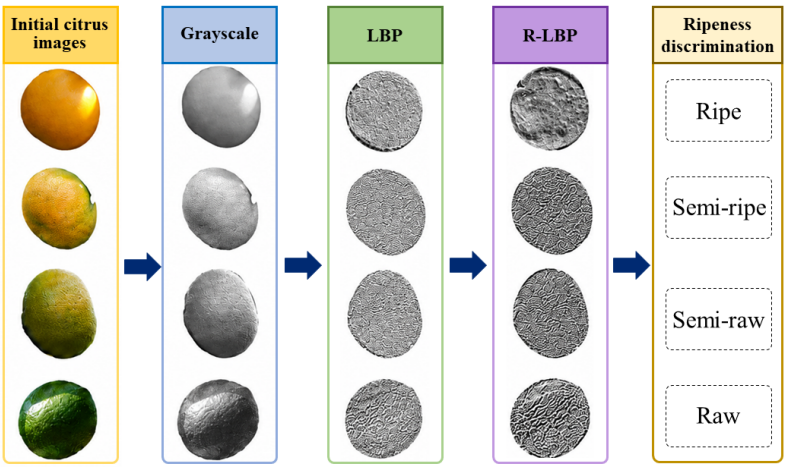


Figure 10. Citrus ripeness discrimination based on LBP and R-LBP texture features. (adapted from Ref. [124] under the Creative Commons Attribution 4.0 License (CC BY 4.0)).

To improve the stability of maturity perception, some studies have introduced non-destructive detection technologies such as multispectral, hyperspectral, and near-infrared spectroscopy, using reflectance characteristics at different wavelengths to characterize maturity-related indicators such as sugar content, pigment content, and firmness [126–128]. For example, Lauretti et al. [129] integrated a VIS-NIR multispectral device on a selective harvesting robot platform to identify multiple tomato maturity stages and verify robotic harvesting. Bertoglio et al. [130] used close-range snapshot hyperspectral imaging to model grape ripening-related indicators and verified its potential for non-destructive assessment. Meng et al. [131] used hyperspectral imaging combined with chemometric models to predict soluble solids content, flesh firmness, and other physicochemical indices of kiwifruit at

different maturity stages, and further visualized their spatial distributions, as shown in Figure 11. However, due to equipment cost, high data dimensionality, online computational burden, and strict calibration requirements, real-time deployment of these methods on mobile harvesting robots remains challenging.

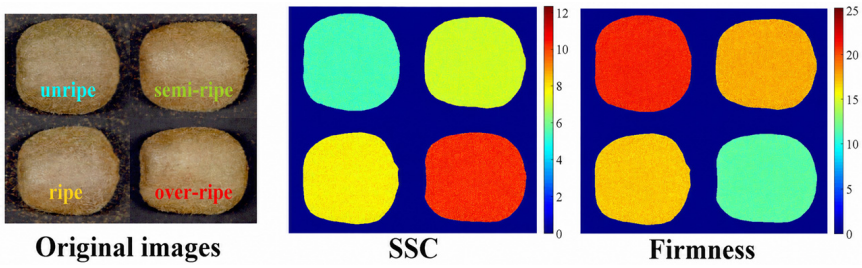


Figure 11. Visualization of kiwifruit maturity-related spatial distributions based on hyperspectral imaging (adapted from Ref. [131] under the Creative Commons Attribution 4.0 License (CC BY 4.0)).

In addition to maturity, fruit surface health status also affects harvesting value. Early studies mainly used local color and texture features to detect disease spots or defects on citrus surfaces for fruit quality screening [132,133]. With the development of learning-based models, surface quality perception has gradually shifted from overall classification to defect-region localization and quantification. Liang et al. [134] achieved apple surface defect grading through semantic segmentation and area statistics. Miao et al. [135] fused multi-tone spatial components to segment kiwifruit targets and defect regions and used Grad-CAM to improve the interpretability of fine-grained defect perception. Overall, fruit attribute recognition provides key information such as maturity, health status, and harvesting value for multi-target harvesting, but still faces challenges in generalization and real-time deployment under complex illumination, occlusion, and cross-cultivar conditions.

3.3. Environmental Modeling and Spatial Positioning

Fruit identification and attribute classification provide semantic information about targets, whereas manipulator execution requires three-dimensional spatial states of both targets and surrounding obstacles. For multi-target harvesting, the vision system needs to localize fruits, describe the spatial distribution of leaves, branches, and supporting structures, and provide geometric constraints for accessibility evaluation, picking-point localization, and collision-free path planning.

3.3.1. Obstacle Detection and Environmental Modeling

In harvesting robots, target recognition results from two-dimensional images are usually combined with RGB-D depth data and camera intrinsic models to generate three-dimensional point clouds through back-projection [110,136]. Since raw point clouds often contain fruits, leaves, branches, supporting structures, and sensor noise, they cannot be directly used for path planning. Obstacle detection and environmental modeling are therefore required to clarify the manipulator's operable space and potential collision regions.

After three-dimensional point clouds are generated, environmental modeling first needs to separate non-target obstacles from mixed scenes. For spatially separable structures such as trunks, branches, supporting wires, and leaf clusters, Euclidean clustering, density-based clustering, and region growing can group obstacle points according to spatial neighborhood relationships [137,138]. These methods are suitable for extracting entities with spatial continuity, such as trunk point clouds in orchard robot navigation [139]. However, dense point-cloud clusters alone are insufficient for real-time harvesting planning, because excessive data increase the computational burden of collision detection and path

search. Therefore, obstacle point clouds usually need to be further abstracted into computable geometric representations. For inclined branches or slender structures, PCA-based oriented bounding boxes can describe the principal-axis direction more accurately than ordinary bounding boxes [140]. For irregular canopy point clouds, three-dimensional convex hulls can be used to represent outer boundary envelopes [141].

Beyond local obstacle modeling, multi-target harvesting also requires a global description of traversable and occupied regions. Octree-based voxel maps probabilistically update spatial occupancy states and can maintain relatively stable environmental representations under sensor noise [142]. For example, OctoMap has been used to construct three-dimensional occupancy maps for orchard navigation and path planning [143], and occupancy information can also support viewpoint planning in greenhouse crop reconstruction [144]. However, geometric occupancy maps only indicate whether a space is occupied, but cannot distinguish leaves, trunks, or supporting structures. Since these objects pose different levels of risk to manipulator motion, semantic information is increasingly introduced into environmental modeling [145]. Semantic point-cloud segmentation, layered cost maps, and differentiated collision costs can distinguish obstacle types and reduce unnecessary avoidance of low-risk leaf regions [146–148]. For multi-target harvesting, such semantic spatial representations are useful for evaluating accessible regions, collision risks, and feasible harvesting paths around candidate targets.

To facilitate the understanding of environmental modeling methods in orchard scenarios, Table 4 summarizes the applicable objects, advantages, and limitations of typical technical levels.

Table 4. Typical technical levels and their characteristics in orchard environment modeling [137–148].

Technical Level	Representative Methods	Advantages	Limitations
Obstacle extraction	Euclidean clustering, DBSCAN, region growing	Facilitates the separation of individual obstacles	Sensitive to occlusion and parameter settings
Geometric abstraction	AABB, OBB, convex hull	Reduces the computational burden of motion planning	Large geometric approximation errors
Occupancy mapping	OctoMap, voxel grid	Suitable for obstacle avoidance and path planning	Lack of semantic attributes
Semantic mapping	Semantic voxels, hierarchical cost maps	Enables discrimination of obstacle attributes	Dependent on the accuracy of front-end classification

Overall, orchard environmental modeling has evolved from local obstacle extraction to occupancy mapping and semantic mapping. Its role is to provide continuous, computable, and semantically discriminative spatial constraints for target pose estimation and harvesting decisions.

3.3.2. Target Pose Estimation Method

The visual perception system ultimately needs to provide executable spatial poses for the end-effector, including fruit coordinates, approach direction, grasping posture, cutting position, or stem-axis orientation [149]. In practical systems, two-dimensional pixel coordinates are transformed into the robot base coordinate system using depth information

and hand–eye calibration matrices. However, geometric centroids alone are insufficient under complex occlusion and diverse crop morphologies. Therefore, pose estimation methods need to consider target biological characteristics, end-effector operation modes, and local environmental constraints [150]. Typical methods and applicable scenarios are summarized in Table 5.

Table 5. Classification and characteristics of target pose estimation methods for harvesting.

Method Category	Representative References	Core Method	Pose Representation	Advantages and Applicable Scenarios
Conventional pose estimation	Gao et al. [151]	RANSAC-based spherical fitting combined with line fitting	Vertical vector from the fruit center to the main branch line	Corrects centroid offset and provides a collision-free approach direction
	Niu et al. [152]	Berry-region identification combined with geometric connection	Direction of the line connecting the fruit center and the berry-cluster center	Effective robustness to local point-cloud noise
	Kang et al. [153]	Three-dimensional spherical Hough transform	Euler angles computed from spatial distribution	Voting-based mechanism, improved robustness under partial occlusion
	Lehnert et al. [154]	Superellipsoid geometric modeling combined with nonlinear optimization	Full six-degree-of-freedom pose	Effective representation of complex non-spherical morphologies with rounded contours, such as plants
	Luo et al. [155]	Locally weighted regression smoothing	Tangential direction of the fruit stem centerline	Guides the cutting mechanism to align accurately with the cutting point
	Zhang et al. [156]	Skeleton thinning extraction combined with topological feature analysis	Vector between the picking-point cutting slope and key points	Suitable for greenhouse crop branch structures and guides accurate cutter alignment
Learning-based pose estimation	Wang et al. [157]	Geometric perception network A3N combined with PointNet	Direct regression of the predicted manipulator approach angle	Capable of handling natural illumination and severe occlusion, achieving centimeter-level accuracy
	Ci et al. [158]	Keypoint R-CNN combined with 3D point-cloud fusion	Spatial pose of fruit stem nodes	Combines 2D features with 3D space for accurate localization
	Park et al. [159]	YOLOv8n-seg combined with local curvature features	Target stem-axis direction	High real-time performance, balancing ripeness assessment and pose estimation in occluded scenes

Traditional pose estimation methods mainly rely on explicit geometric models. For regular spherical fruits such as apples and citrus, fruit centers, surface normals, and approach directions are usually estimated through sphere fitting, ellipsoid fitting, or geometric correction combined with branch direction [151,152]. When point clouds are incomplete, three-dimensional spherical Hough transform can further improve the robustness of center estimation [153]. For crops with complex shapes or clustered distributions, such as sweet peppers, grapes, and tomatoes, pose estimation focuses more on stems, branches, or cutting

points. Superellipsoid models, locally weighted regression, skeleton thinning, and topological analysis have been used to estimate fruit pose, stem direction, or cutting posture for different crop structures [154–156].

Traditional geometric methods are interpretable, but they are sensitive to missing point clouds, depth noise, and illumination variation. In recent years, learning-based methods have begun to estimate approach angles, keypoints, and local poses directly from images, point clouds, or multimodal features. For example, geometric perception networks combined with PointNet can predict manipulator approach angles under occlusion [157]. Methods combining Keypoint R-CNN or YOLOv8n-seg with three-dimensional point clouds can balance key-structure localization, real-time performance, and pose-estimation accuracy in clustered scenes [158–160]. For multi-target harvesting, pose estimation affects not only single-target grasping or cutting, but also target accessibility, collision-free path generation, and harvesting-sequence arrangement. Therefore, future methods need to improve pose stability under occlusion while meeting real-time deployment requirements.

3.4. Spatiotemporal Correlation and Perceptual Consistency

In studies on fruit recognition and pose estimation, visual perception methods usually focus on parsing single-frame images or point clouds. However, orchard harvesting has clear temporal continuity. Robots need not only to recognize targets at a single moment, but also to maintain stable target identities and spatial states during continuous operation. Affected by branch and leaf occlusion, illumination variation, and sensor noise, frame-by-frame detection and localization results may suffer from ID switches, short-term target loss, and three-dimensional position fluctuations, which further affect target counting, repeated-harvesting avoidance, harvesting-sequence updating, and manipulator control.

To model target consistency across time steps, detection-based multi-target tracking methods usually take frame-by-frame detection results as input and maintain target identity continuity through cross-frame matching. Representative studies on matching strategies and identity-preservation mechanisms are summarized in Table 6.

As shown in Table 6, different tracking methods involve trade-offs among real-time performance, occlusion robustness, and identity preservation. Motion- or geometry-consistency-based methods have low computational cost and are suitable for real-time harvesting scenarios, but they are prone to ID switches under dense fruit overlap or short-term occlusion [161–165]. Appearance-enhanced methods improve target re-identification after occlusion, but their performance can be affected by illumination variation and similar fruit appearance [166–169]. Multi-cue fusion methods provide stronger identity preservation in complex scenes, but they also increase system complexity [170,171].

Another key issue in perceptual consistency is spatial-state stability. In unstructured orchards, depth noise, occlusion changes, viewpoint motion, and coordinate-transformation errors can cause target three-dimensional positions to fluctuate across frames. Without stabilization, picking-point generation, path planning, and end-effector alignment may all be affected. Therefore, related studies have proposed spatial-state stabilization methods from the perspectives of temporal smoothing, multi-source fusion, calibration compensation, and observation-quality improvement, as summarized in Table 7.

As shown in Table 7, recursive filtering methods can suppress three-dimensional position jitter with relatively low computational cost, making them suitable for real-time state updating during continuous harvesting, but they depend on motion models and noise assumptions [172–174]. Multi-source fusion and extrinsic compensation methods improve the consistency of multi-view or RGB-D perception results in a unified coordinate system, but their effectiveness depends on calibration accuracy and synchronization quality [102,175–177]. Depth-map restoration, point-cloud completion, active perception,

and laser-assisted measurement can improve three-dimensional data quality from the observation source, but usually increase hardware or computational burden [178–180]. Compared with single-frame detection accuracy, spatiotemporal consistency is more directly related to state reliability during continuous harvesting and is a key basis for multi-target harvesting to progress from target detection to stable task execution.

Table 6. Representative studies on detection-based multi-target tracking in identity continuity modeling.

Method Category	Research Problem	Target Object	Research Method	Core Principle	Reference
Motion/geometric consistency-driven	Stringent requirements for low latency and real-time performance in closed-loop control	General multi-target scenarios	SORT	Online data association based on Kalman filtering prediction and Hungarian matching, emphasizing real-time performance and algorithmic simplicity	Bewley et al. [161]
	Identity discontinuity and target overlap caused by occlusion	Mango	MangoYOLO + Kalman + Hungarian	Frame-by-frame detection combined with motion prediction and Hungarian assignment for cross-frame matching, alleviating target identity loss caused by occlusion	Wang et al. [162]
	Tracking failure caused by morphological deformation of clustered targets	Grape clusters	Lightweight YOLOv5s + tracking-and-counting pipeline	Designed for morphological variation and occlusion in clustered targets, achieving robust identity continuity through detection-driven cross-frame association	Shen et al. [163]
	ID drift and mismatching in complex backgrounds	Tomato	YOLOv5-4D (RGB-D fusion) + ByteTrack + region counting	RGB-D fusion improves detection robustness; ByteTrack performs cross-frame association, while region-based counting suppresses ID drift and identity switching	Rong et al. [164]
	Identity index jumps caused by dense target overlap	Citrus	OrangeYolo + OrangeSort	Detection-driven tracking maintains identity continuity through cross-frame matching and reduces ambiguity caused by dense overlap	Zhang et al. [165]
Appearance-enhanced association	Failure of purely geometric matching under dense occlusion	General multi-target scenarios	DeepSORT	Deep appearance feature metrics are introduced on top of the motion model, and appearance similarity is exploited to reduce ID switches in dense scenes	Wojke et al. [166]
	Feature degradation and identity loss caused by drastic illumination changes	Citrus	Improved YOLOv5s + DeepSORT	Detection is combined with appearance-feature association to improve cross-frame consistency and state stability under occlusion and illumination variation	Lin et al. [167]
	Insufficient association confidence in deeply complex scenes	Banana	RGB-D + DeepSORT	DeepSORT is used to address visual re-identification, while additional constraints are introduced to improve the reliability of cross-frame association	Zhou et al. [168]
	Risk of trajectory fragmentation in long-term monitoring	Tomato	YOLO-DeepSORT	Detection and appearance-based association are combined to maintain identity continuity across frames for long-term monitoring and stable target-state estimation	Ge et al. [169]

Table 6. *Cont.*

Method Category	Research Problem	Target Object	Research Method	Core Principle	Reference
Multi-cue fusion and robust association	Effects of illumination interference and sensor noise on stability	Strawberry	StraTracker (feature association + adaptive filtering + counting module)	The problem is reformulated as frame-by-frame tracking, integrating feature association and adaptive filtering to reduce illumination interference and ID switching	An et al. [170]
	Viewpoint mutation and jitter caused by manipulator motion	Green pepper	Detection + improved DeepSORT (appearance matching + trajectory optimization)	Appearance matching and trajectory optimization are introduced to reduce ID switches and improve state estimation accuracy	Du et al. [171]

Overall, the role of visual perception in multi-target harvesting is to transform unstructured orchard environments into structured state representations that can support decision-making. From sensor configuration and coordinate mapping to target recognition, attribute identification, spatial modeling, and spatiotemporal consistency maintenance, existing studies have formed a relatively complete perception pipeline. Through these methods, complex operating scenes can be abstracted into candidate target sets with identity information, attributes, spatial positions, and stable states, thereby reducing the uncertainty of perception outputs. On this basis, the key issue in multi-target harvesting further shifts to how these environmental and target states can be used for target screening, priority ranking, and harvesting decision-making.

Table 7. Representative studies on methods for target spatial-state stabilization and perceptual consistency enhancement.

Method Category	Target Object and Characteristics	Research Method	Core Principle	Reference
Filtering and probabilistic estimation	Target fruits in multi-view occluded environments	Particle-filter-based multi-camera fruit localization	The 3D state is treated as a latent variable and updated by Bayesian inference based on multi-view observations; temporal filtering is used to suppress measurement jitter	Mehta et al. [172]
	Orchard multi-target fruits under camera vibration	Dynamic Kalman-filter-based target state updating (combined with detection and tracking)	Adaptive noise weighting is introduced in the prediction–update stage to mitigate the amplification of observation noise and manipulator-induced vibration in state estimation	Zhai et al. [173]
	Strawberry under close-range continuous harvesting	Systematic optimization of 3D localization methods and error modeling	Localization uncertainty is quantified from the perspectives of measurement sources and sensing chains, and an error-distribution model is established to provide a theoretical basis for filter parameter setting	Ge et al. [174]

Table 7. *Cont.*

Method Category	Target Object and Characteristics	Research Method	Core Principle	Reference
Fusion and calibration compensation	Multiple fruit types under strong outdoor illumination	Systematic error analysis and characterization of depth-camera systems	Different error sources, including noise distribution and distance drift, are modeled, and multi-sensor fusion and compensation strategies are proposed to reduce instability in 3D output	Neupane et al. [175]
	General targets viewed from an end-effector-mounted ToF sensor	ToF-based hand–eye calibration optimization	The hand–eye transformation matrix in the calibration chain is optimized to reduce the accumulated effect of extrinsic drift and coordinate-transformation errors on 3D localization accuracy	Zhang et al. [102]
	Kiwifruit with calyx interference	Binocular localization based on fruit and calyx detection	Calyx keypoint matching and geometric constraints are introduced to reduce mismatches and improve the consistency of stereo disparity estimation along the depth (Z-axis) direction	Gao et al. [176]
	Strawberry plants under a multi-view acquisition architecture	Calibration-enhanced multi-view RGB-D perception and fusion	High-precision extrinsic calibration is used to support cross-view point-cloud fusion, alleviating target-surface incompleteness and localization jumps caused by depth loss in single-view observations	Hu et al. [177]
Noise suppression/active Sensing	Kiwifruit with depth loss and occlusion	Regional filtering + end-to-end stereo matching network (full-resolution depth)	Data cleaning and network optimization are used to improve dense-depth quality and fill holes caused by occlusion and weak texture, thereby enhancing 3D stability from the input stage	Jing et al. [178]
	Grapes with irregular cluster distribution	Harvest-point 3D localization based on depth filtering and irregular completion	Adaptive filtering and completion are performed for clustered depth loss and outliers, enhancing point-cloud structural continuity and improving the stability of harvest-point spatial estimation	Shen et al. [179]
	Apples under drastic natural illumination changes	Active laser–camera scanning and calibration	Active measurement is used to replace passive depth estimation that is highly sensitive to environmental changes, and laser triangulation is employed to obtain millimeter-level stable depth-based positional output	Zhang et al. [180]

3.5. Perception Performance Characteristics

The practical performance of perception systems in orchard environments is jointly affected by sensor configuration, perception methods, operating conditions, and computational efficiency. Table 8 therefore compares representative systems in terms of sensor configuration, validation setting, semantic perception, spatiotemporal performance, and processing efficiency.

Table 8. Representative perception systems and their reported performance in harvesting-related tasks.

Perception Task	Sensor Configuration	System and Task	Validation Level and Setting	Semantic Perception Performance	Spatiotemporal Performance	Processing Efficiency	Ref.
Target recognition	RGB camera	Apple recognition based on ShufflenetV2-YOLOX	Natural-orchard validation: apple recognition	AP 96.76%, precision 95.62%, recall 93.75%, and F1-score 0.95	—	65 FPS	[2]
Attribute recognition	RGB camera	Four-stage tomato maturity recognition using an improved YOLOv5 model	Greenhouse/semi-structured validation: tomato images	Precision 95.58% and mAP 97.42%	—	9.2 ms/image	[56]
Target and key-structure recognition	RGB camera	Strawberry and keypoint recognition using STRAW-YOLO	Dataset-based validation: strawberry and keypoint dataset	Precision 91.6%, recall 91.7%, and mAP@0.5 96.0%	Provides keypoint information required for harvesting operations	62.6 FPS	[119]
Target recognition and localization	RGB-D camera	Apple recognition and picking-point localization using an improved YOLOX model	Natural-orchard and dataset validation: apple localization	F1-score 93.0% and mAP 94.09%	Localization errors in the X, Y, and Z directions were below 7 mm, 7 mm, and 5 mm, respectively	167.43 FPS	[90]
Instance segmentation and localization	RGB-D camera	Apple recognition and picking-point localization using an improved SOLOv2 model	Dataset-based validation: orchard images	F1-score 88.5%, mAP 90.1%, and mIoU 83.2%	Localization errors in the X, Y, and Z directions were 0–3.95 mm, 0–5.16 mm, and 0–1 mm, respectively	29.5 FPS	[91]
Depth measurement for fruit localization	Stereo, active infrared stereo, and ToF depth cameras	Comparison of eight depth cameras for fruit localization and size measurement	Controlled outdoor validation: sunlight, 1.5 m	—	Azure Kinect achieved an RMSE of 6 mm, a bias of 17 mm, a standard deviation of 2 mm, and a depth fill rate of 100%	—	[175]
Recognition and pose estimation	Stereo vision	YOLO-Broccoli-Seg and point-cloud-based broccoli pose estimation	Field validation: complex background and inclined broccoli	Mask mAP@0.5 of 0.973 and Mask mAP@0.5:0.95 of 0.683	Pose-estimation R^2 of 0.934	Approximately seven times the FPS of Mask R-CNN	[160]
Fruit and calyx localization	Stereo vision	Kiwifruit and calyx localization using YOLOv5x recognition and stereo matching	Natural-orchard validation: kiwifruit with calyx interference	mAP of 93.5%	Mean deviations in the X, Y, and Z directions were 7.9 mm, 6.4 mm, and 4.8 mm, respectively	108 ms/image	[176]
Orchard semantic mapping	32-line LiDAR with an external IMU	LiDAR-inertial mapping and 3D-ODN semantic perception	Multi-setting validation: greenhouse and natural orchards	Tree-detection mIoU of 0.812	Semantic maps supported tree localization and traversable-area analysis	Feature extraction: 0.1681 s; prediction: 0.0268 s	[148]
Orchard point-cloud segmentation	LiDAR point cloud	Tree and background segmentation using DFSNet	Dataset-based validation: orchard point clouds	Overall accuracy 89.43% and mIoU 74.05%	Achieved semantic separation of orchard elements	—	[146]
High-precision fruit localization	Active laser-camera system	Apple localization based on red-line laser triangulation	Controlled validation: 1.0–1.6 m, varied occlusion	—	In most cases, localization errors in the X, Y, and Z directions were within 0.4 mm, 0.8 mm, and 3 mm, respectively	—	[180]

Table 8. *Cont.*

Perception Task	Sensor Configuration	System and Task	Validation Level and Setting	Semantic Perception Performance	Spatiotemporal Performance	Processing Efficiency	Ref.
Temporal tracking and counting	RGB video	Grape-cluster tracking and counting using pruned YOLOv5s combined with SORT	Field-video validation: clustered and occluded targets	Detection mAP of 82.3%	Counting accuracy 84.9% and correlation coefficient of 0.9905 with manual counts	50.4 FPS	[163]
Temporal tracking and counting	RGB video	Citrus tracking using OrangeYolo and OrangeSort	Field-video validation: two orchards, varied lighting	—	Counting MAE of 0.081 and standard deviation of 0.08	—	[165]
Attribute recognition	RGB-D camera + VIS–NIR multispectral sensor	Tomato ripeness classification and selective harvesting platform	Controlled validation: six ripeness stages, nine target positions	Average ripeness-classification accuracy: 93.72%	Maximum position and orientation errors: 8 mm and 0.09 rad	—	[129]
Attribute recognition	Hyperspectral imaging (400–1000 nm)	Kiwifruit ripeness evaluation using PLS/PLS-DA	Controlled postharvest validation	SSC: R^2 0.85–0.94, RMSE 0.73–1.10 °Brix; classification sensitivity: 93–97%	—	—	[131]

Table 8 shows that different perception configurations have distinct information-acquisition capabilities and performance emphases. RGB systems mainly provide rapid recognition of fruits, maturity levels, and key structures, thereby supporting efficient candidate-target screening. RGB-D and stereo-vision systems preserve semantic recognition capability while further providing three-dimensional information on target locations, picking points, and poses. Their localization accuracy is jointly affected by depth-measurement methods, occlusion levels, target texture, and calibration quality. LiDAR point-cloud systems are more suitable for describing the spatial distribution of trees, ground surfaces, and obstacles and for constructing environmental semantic representations for traversability and collision-risk analysis. In addition, active laser-camera systems can provide higher-precision depth localization under complex illumination and occlusion conditions, whereas video-tracking systems maintain target identity and state continuity through cross-frame association. The multispectral and hyperspectral systems listed in Table 8 can further use spectral features for maturity and internal-quality assessment, thereby supplementing attribute screening in selective harvesting. Their representative results are mainly obtained under controlled validation conditions, and more systematic evidence is still needed regarding online processing efficiency and robustness under complex natural illumination.

In addition to performance metrics, the validation environment is an important basis for assessing the engineering maturity of a perception system. Results obtained from datasets, laboratory experiments, or other controlled environments mainly reflect the recognition and spatial-measurement capabilities of perception methods under relatively controllable conditions. Greenhouse or semi-structured environments begin to introduce practical factors such as crop morphology, operating space, and illumination variations. Natural-orchard validation further examines system stability under conditions including occlusion, illumination fluctuations, background interference, and uneven target distributions. Therefore, the same accuracy or processing-efficiency metric should be interpreted together with its validation level, and field results can more directly reflect the readiness of a perception system for continuous harvesting deployment.

From an operational perspective, perception-system selection should not be determined solely by a single recognition accuracy or localization error. Continuous multi-target harvesting requires simultaneous information on target identity, maturity, spatial location, operable pose, surrounding obstacles, and cross-frame state changes. RGB systems provide high processing efficiency for scenarios dominated by rapid target screening. For tasks requiring precise localization, accessibility assessment, and picking-pose planning, RGB-D, stereo vision, or active laser measurement can provide more direct spatial constraints. LiDAR point clouds and semantic maps are particularly valuable for large-scale orchard-environment modeling and shared spatial information among multiple execution units. Perception outputs should be further organized into candidate-target sets containing target attributes, spatial relationships, accessibility, execution costs, and state stability to support target-priority determination, harvesting-sequence planning, and task allocation.

The processing efficiencies reported in Table 8 normally correspond to specific perception tasks and cannot independently represent the end-to-end response time of a harvesting system. In practical deployment, visual inference must also be coupled with depth estimation or point-cloud processing, coordinate transformation, target-state maintenance, and subsequent online scheduling, and the computational overhead of these stages accumulates within a single harvesting cycle. Therefore, a high frame rate of an individual recognition model does not necessarily mean that the full perception-to-decision chain can meet real-time requirements. For mobile harvesting systems, on-board hardware or edge-computing platforms require sufficient computing capacity to maintain the update frequencies needed for perception, spatial-state construction, and online scheduling, while reserving computational resources for motion control and safety monitoring.

4. Multi-Target Harvesting Decision-Making Under Constraints

Three-dimensional scene reconstruction, target recognition, and state estimation provide harvesting robots with environmental and target information. However, the key issue in multi-target harvesting is not only whether fruits can be perceived, but also how to determine target priorities, harvesting sequences, and task arrangements under limited operation time, manipulator workspace, and collision-avoidance constraints. Therefore, multi-target harvesting decision-making needs to balance efficiency, safety, and execution reliability. Existing studies generally recognize that perception results cannot directly determine harvesting behavior. Instead, information abstraction and constraint representation are required to transform environmental states into target sets, cost indicators, and task constraints for decision reasoning. According to differences in decision-generation mechanisms and modeling forms, related methods can be summarized as rule- and experience-based heuristic methods, optimization-based methods aiming at overall performance improvement, and learning-based methods that generate strategies through data-driven approaches.

4.1. Decision Information Abstraction and Constraint Representation

In multi-target harvesting tasks, visual perception systems can usually output fruit detection results, three-dimensional positions, and keypoints. However, these fragmented perception results are difficult to use directly as inputs for task-level decision-making. Before entering the decision-making process, these outputs need to be associated across coordinate frames and temporally synchronized and then screened by considering confidence and operational constraints so that they can be organized into updatable, decision-oriented state representations. This state representation links target attributes and spatial information with harvestability, accessibility, interference risk, and estimated execution requirements, thereby transforming environmental observations into computable decision inputs. Existing studies have pointed out that an information organization and abstraction process is

required between perception and decision-making, so that transient perception outputs can be transformed into stable and structured task descriptions to support harvesting-sequence planning and task allocation [181,182].

From existing task-planning studies on harvesting robots, decision-making usually does not rely on a single perceptual quantity, but integrates target information, spatial relationships, and task states. In decision-oriented state representations, this information is organized at the target, relationship, and task levels. In this review, this three-level organization serves as a conceptual analytical framework for integrating the information required for multi-target harvesting decisions. The boundaries between these levels are analytical rather than rigid, and the specific state variables assigned to each level may vary with the sensing, planning, and execution requirements of different harvesting systems. First, target-level information such as fruit position, picking or grasping point, maturity, and size can be used to construct candidate target sets and form initial priority or cost evaluation. In clustered or high-density fruit scenarios, such information also needs to be combined with local spatial neighborhood structures to reduce unnecessary back-and-forth motion and improve operational efficiency [11]. Second, relationship-level information such as reachability, distance, occlusion or interference risk can further support task allocation, conflict avoidance, and local sequence adjustment. For continuous harvesting tasks, task-level information should include the remaining target set, target availability, execution progress, retry status, and manipulator availability. These states provide the execution context required for event-triggered dynamic rescheduling when task conditions or resource availability change. Together, these three levels provide a structured analytical basis for candidate-target evaluation, priority or cost construction, task allocation, and replanning. Representative decision-oriented state information and its roles are presented in Table 9.

Table 9. Decision-oriented state representations and their roles in multi-target harvesting.

Information Level	Specific Content	Role in Decision-Making	Representative References
Target-level information	3D position, picking/grasping point, maturity, size	Construction of candidate target sets; priority or cost evaluation; generation of the initial harvesting sequence	Kang et al. [183]; Bhattacharyya et al. [184]; Zhu et al. [185]; Xie et al. [186]
Relationship-level information	Reachability, distance, occlusion/interference risk, workspace overlap	Task allocation; conflict avoidance; sequence adjustment and local reordering	Han et al. [187]; Cho et al. [188]; He et al. [43]
Task-level information	Remaining target set, execution progress, retry status, and manipulator availability	Event-triggered dynamic rescheduling: task-state updates, retry/deferral, partial reassignment, and local sequence adjustment after failures or other abnormal events	Xie et al. [189]; Liu et al. [12]

To support subsequent harvesting decisions, the environmental states provided by the sensing system are typically organized into a unified, computable representation to serve as direct input to the decision-making algorithm. Target-level states can be used to evaluate candidate targets and construct priority scores or costs. Relationship-level states provide the accessibility and interference constraints required for sequence adjustment and conflict avoidance. Task-level states provide the execution context required for task allocation and dynamic rescheduling. A common approach in existing research is to use a graph structure, treating fruits as nodes and characterizing the connections between nodes using distance, operational costs, or reachability relationships. Alternatively, a

state space or scheduling model centered on state variables and constraint sets can be constructed to describe the current available actions and resource usage. In this way, different types of decision-making methods can reason and solve problems based on the same information, thereby generating specific harvesting sequences and operational plans. To support dynamic rescheduling, the representation should allow event-triggered state transitions. When a harvesting failure, target disappearance, newly confirmed target, or temporary manipulator unavailability occurs, the system updates the affected target states, task ownership, and available resources and then performs constraint-aware dynamic rescheduling through retrying, local sequence adjustment, task reallocation, or replanning.

4.2. Heuristic Decision-Making Methods

Faced with the real-time requirements of multi-target harvesting tasks in unstructured environments, heuristic methods usually do not pursue strict global optimality. Instead, they use manually designed rules, empirical criteria, or evaluation functions to quickly generate executable harvesting sequences under limited computational resources. These methods usually establish target priorities based on information such as target position, maturity, occlusion level, harvestability, spatial distribution, and movement cost. Because of their intuitive logic, low computational cost, and ease of engineering implementation, heuristic decision-making methods have been widely used in early harvesting robot systems [190].

Early heuristic strategies usually determine harvesting order based on one or a few key factors. Among them, spatial distance is the most intuitive criterion, as it can reduce unnecessary back-and-forth manipulator motion among targets. Park et al. [10] designed a harvest ordering module in a cucumber harvesting system, where a global camera obtained the spatial positions of multiple fruits and generated the harvesting sequence according to their relative positions, as shown in Figure 12. Similar studies represent detected targets as spatial candidate sets and generate harvesting sequences according to distance or spatial proximity [191]. These distance-priority strategies are simple and real-time, but they mainly focus on movement cost and give limited consideration to fruit value, occlusion, and harvestability.

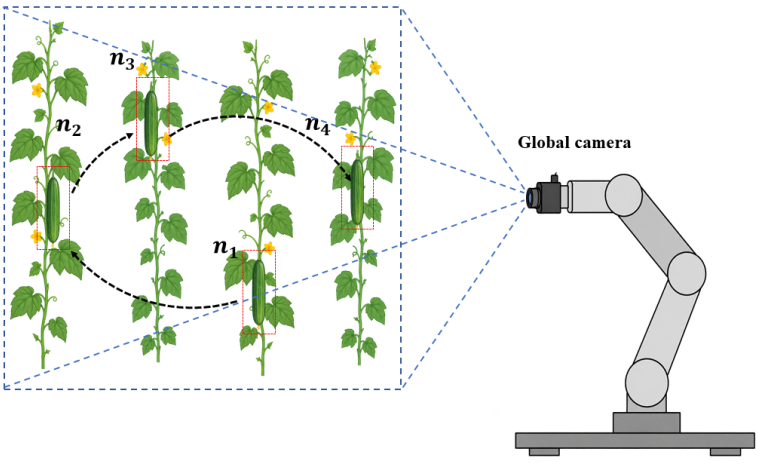


Figure 12. Schematic diagram of multi-target harvesting order based on spatial proximity relationship.

Different from distance-priority strategies, maturity-priority strategies emphasize harvesting value and usually prioritize fruits that have reached the harvesting standard. Rong et al. [192] used tomato maturity as the basis for selective harvesting decisions, and Hou et al. [193] determined cherry tomato harvesting priority according to bunch structure and the spatial distribution of mature fruits. Zhang et al. [194] further proposed a

graded strawberry harvesting strategy that classifies mature fruits into priority categories according to their spatial structures and updates the remaining target categories after high-priority fruits are harvested, as shown in Figure 13. These methods make the harvesting sequence more consistent with agricultural production requirements, but when mature fruits are dense or severely occluded, spatial relationships and accessibility still need to be considered.

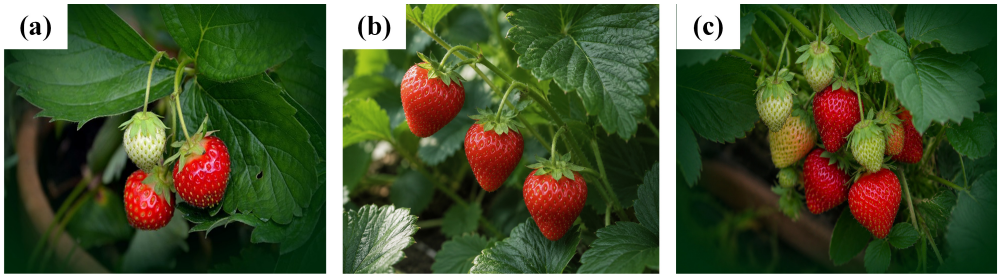


Figure 13. Examples of classification of mature strawberries under different spatial structures. (**a**) Classification of adjacent mature fruits; (**b**) classification under linear arrangement; (**c**) classification under clustered arrangement.

In heavily occluded or dense-target scenarios, relying only on distance or maturity may lead to invalid grasping or frequent path jumps. Therefore, some studies further introduce harvestability judgment and spatial zoning. He et al. [43] proposed a strategy that prioritizes unobstructed or slightly occluded strawberries in complex occlusion environments to reduce invalid grasping and improve efficiency. Karkee et al. [195] introduced a harvestability discrimination mechanism for field strawberry harvesting, prioritizing targets that satisfy accessibility and operational conditions. When targets are clustered, spatial zoning is more suitable than single-target ranking for organizing continuous operation. Dai et al. [11] proposed the 3MSP2 sequential harvesting planning framework based on spatial zoning. The method first divides tomato harvesting clusters using density clustering and then determines inter-cluster and intra-cluster harvesting sequences, as shown in Figure 14. Wang et al. [196] used K-means clustering to generate two-level tomato harvesting sequences between and within task regions, while Ning et al. [197] proposed the AYDY framework for high-density sweet pepper harvesting, generating collision-free harvesting sequences through clustering and intra-cluster ordering. Compared with single-rule methods, spatial zoning is more suitable for dense-target scenarios, but its performance still depends on clustering results and manually designed ordering logic.

On this basis, evaluation-function-driven heuristic methods further map multiple factors into computable scores or costs. Compared with single-rule strategies, evaluation functions can jointly consider movement cost, operational feasibility, target value, success probability, and operational safety, thereby providing a quantitative basis for harvesting-sequence generation. Van Henten et al. [198] constructed an evaluation function incorporating travel time and operation cost for greenhouse cucumber harvesting. Bac et al. [199] further noted that harvesting systems for high-value crops usually determine harvesting priority by jointly evaluating distance, success probability, and fruit value. In practice, Silwal et al. [200] calculated continuous scores for apple candidate targets and selected the target with the best evaluation value at each step to gradually construct the harvesting sequence. Ji et al. [201] combined improved DBSCAN spatial clustering with multidimensional weighted evaluation, using Gaussian distance weight and saliency level to quantify apple harvesting priority. Lehnert et al. [202] introduced task-cost evaluation in sweet pepper harvesting, allowing sequence decisions to balance efficiency and success rate. Therefore, evaluation-function-based methods can be regarded as an extension from rule-based heuristics to multi-factor quantitative decision-making.

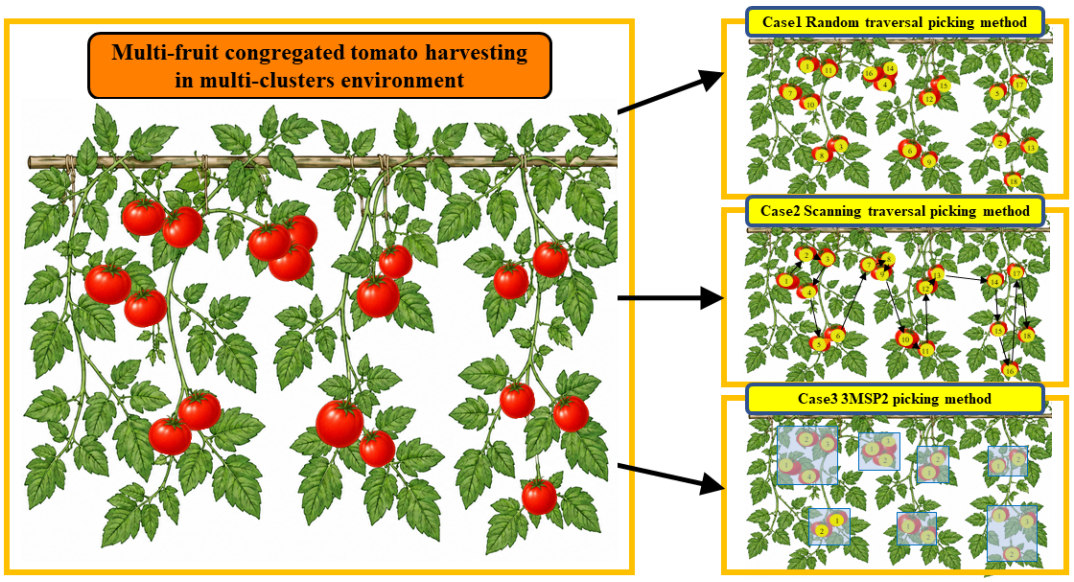


Figure 14. Schematic diagram of different harvesting decision-making methods.

Overall, heuristic decision-making methods have advantages in real-time performance, simplicity, and interpretability, making them suitable for harvesting tasks with limited target numbers, relatively simple environments, or constrained computational resources. However, their performance depends on manually designed rules, empirical weights, or predefined evaluation functions, and is sensitive to environmental changes and parameter design. When target scale increases, occlusion becomes more severe, or constraints such as reachability, collision risk, and failure retry become strongly coupled, heuristic methods often struggle to achieve unified global trade-offs [203,204].

4.3. Optimization Decision-Making Methods

Compared with heuristic methods, optimization-based decision-making emphasizes global trade-offs under explicit mathematical models. Its core idea is to formulate harvesting-sequence planning, target selection, or task arrangement as constrained combinatorial optimization problems, with objectives such as operation efficiency, energy consumption, harvesting benefit, or execution reliability [205]. In multi-target harvesting scenarios, robots need not only to reduce unnecessary movement, but also to consider manipulator reachability, collision avoidance, operation time, and target value. Therefore, optimization-based methods usually express harvesting tasks through objective functions and constraint sets to obtain harvesting sequences or task arrangements with better overall performance [206]. A typical optimization-based multi-target harvesting decision-making process is shown in Figure 15.

In the modeling stage, the key issue is how to define decision variables, objective functions, and constraints. In addition to basic costs such as path length, harvesting time, and energy consumption, some studies further introduce occlusion level, collision risk, and grasping failure probability into the objective function to improve executability in complex environments [207]. Other studies model the execution cost of perception tasks, allowing the system to coordinate perception benefit and harvesting efficiency [204]. Compared with heuristic methods, optimization-based modeling can explicitly represent multiple constraints and conduct overall trade-offs under a unified objective. However, its performance depends more strongly on the design of cost functions and the accuracy of environmental state representation.

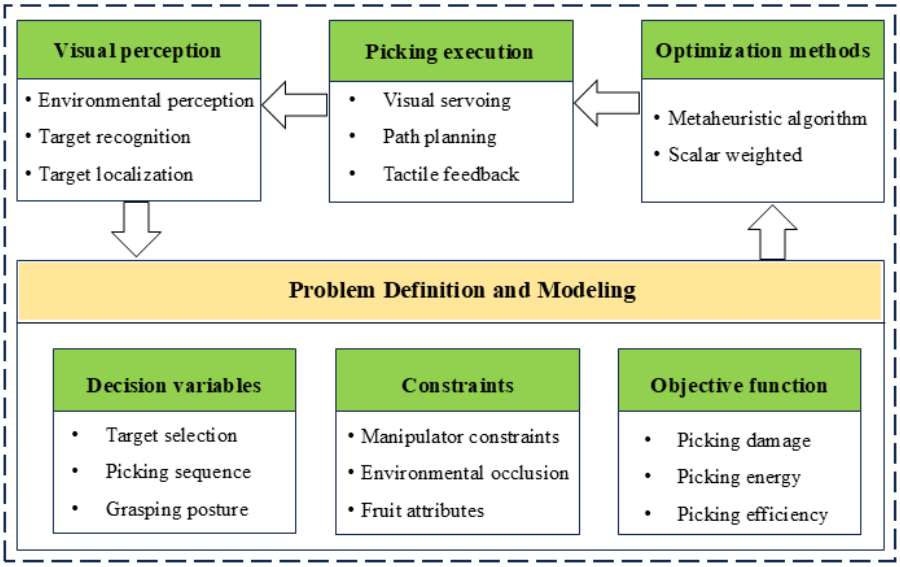


Figure 15. Typical optimization-based decision framework for multi-target harvesting.

Since harvesting-sequence planning is a typical combinatorial optimization problem, exact solution difficulty increases rapidly as the number of candidate targets grows. Therefore, metaheuristic algorithms such as genetic algorithms, ant colony optimization, and particle swarm optimization are widely used to obtain near-optimal harvesting sequences under limited computational resources [208]. Different algorithms emphasize different aspects. Zhang et al. [209] proposed a SOM-GA hybrid algorithm to improve global optimization of apple grasping sequences through structured target division. Li et al. [210] used the pheromone updating mechanism of ant colony optimization to enhance path-search guidance in multi-citrus harvesting spaces. Sun et al. and Xiong et al. [211,212] applied particle swarm optimization to harvesting-sequence generation, obtaining better sequences within limited time through swarm-based iterative search. These methods have stronger global search capability than simple heuristics, but their computational burden and parameter sensitivity remain significant when target numbers or constraints increase.

In addition to stochastic search methods, deterministic weighted models have also been used in multi-target harvesting decisions. Ma et al. [213] proposed a multi-weighted decision scheme for Korla fragrant pear harvesting-sequence planning, using the rated power of each manipulator joint as weight factors to establish an energy-consumption evaluation function and calculate trajectory weights for the pre-harvesting, harvesting, and unloading stages. Such methods have clear objective functions and stable solution processes, making them suitable for tasks with explicit energy or time indicators. However, their adaptability may be limited when uncertainty such as occlusion, failure retry, and dynamic disturbance becomes strong.

Optimization-based decision-making is therefore more suitable for harvesting scenarios with larger target scales, higher efficiency requirements, or constraints that need explicit modeling. Its strength lies in coordinating path cost, execution risk, and operation benefit within a unified framework, but it also faces high computational complexity, dependence on prior information, and model mismatch caused by environmental uncertainty [214].

4.4. Learning-Based Decision-Making Methods

With the expansion of perceptual data and artificial intelligence methods, learning-based decision-making has become an important direction in multi-target harvesting. Unlike heuristic methods, it learns target screening, priority judgment, and strategy generation from operational data rather than relying only on manually designed rules or fixed weights.

Compared with optimization-based methods, it can partly reduce the difficulty of explicit cost modeling in complex scenarios [215]. However, practical deployment performance depends not only on model performance, but also on whether the training data and observed operating states adequately cover variations in orchard environments, target states, and execution constraints. Existing studies mainly involve supervised learning, imitation learning, and reinforcement learning, which correspond to target scoring, expert-policy learning, and long-term reward optimization, respectively.

Supervised learning methods usually learn indicators related to harvesting benefit or execution reliability, such as grasping quality, harvestability, or local risk, and use them to support target ranking. Wang et al. [157] used a geometric perception network to predict fruit grasping pose and grasping quality, which can further support harvestability evaluation. Zhou et al. [216] combined target selection with obstacle perception by using semantic segmentation, depth information, and visible fruit area to determine candidate target priorities. These methods replace simple manually designed weights with data-learned evaluation functions, but their performance still depends on training-data distribution and annotation quality. Insufficient training-data coverage of severely occluded fruits, dense fruit clusters, maturity variations, localization uncertainty, or changes in manipulator reachability may bias learned scores or harvestability estimates. Such bias can reduce the reliability of subsequent target ranking.

Imitation learning learns target selection and action preferences from expert demonstrations, allowing robots to form implicit priority strategies closer to human experience. Related studies have used learning from demonstration to train continuous actions such as approach, picking, and placement, or to improve dynamic grasping and viewpoint selection in occluded scenes [217–219]. Such methods can better reflect human experience in target selection and execution reliability, but they depend on demonstration coverage and may suffer from error accumulation in unseen scenarios. Furthermore, policies learned from expert demonstrations do not inherently satisfy constraints related to kinematic feasibility, collision avoidance, or fruit-damage control. Their practical deployment therefore still requires feasibility verification, risk assessment, or rule-based safety mechanisms. In unstructured pepper harvesting, imitation learning has also shown practical potential for strategy generation [220].

Reinforcement learning formulates multi-target harvesting as a sequential decision-making problem and optimizes target selection or sensing actions through long-term rewards. Deep reinforcement learning has been used for Next-Best-View planning to improve target localization reliability [221], and self-supervised reinforcement learning has been applied to viewpoint adjustment for occluded stems or cutting points [222]. These methods are suitable for dynamic updating and long-term reward optimization. However, transferring policies trained in simulation to real orchards remains challenging because simulations cannot fully reproduce factors such as illumination variations, leaf and branch occlusion, wind-induced target motion, contact uncertainty, and perception errors. The acquisition of large-scale real interaction data is also costly, particularly for samples involving harvesting failures, rare disturbances, and high-risk operating states. At the same time, online trial and error cannot be directly applied to practical harvesting if it may cause collisions, fruit damage, or infeasible manipulator motions. Therefore, in addition to effective reward design and policy learning, practical deployment requires representative training data, safety-aware action constraints, and feasibility or risk verification during execution. Table 10 summarizes the decision-modeling forms, advantages, limitations, and representative studies of different learning-based methods.

Table 10. Comparison of learning-based methods for multi-target harvesting priority decision-making.

Category	Decision Modeling Strategy	Advantages	Limitations	Representative References
Supervised learning	Learns target scores or feasibility-related evaluation values from labeled data, and outputs target ranking or pickability prediction to support target selection and sequence decision-making	Stable training, fast inference, easy integration with heuristic optimization or decision fusion	Strong dependence on labeled data; limited cross-scenario generalization; performance sensitive to evaluation design	Alaauddeen et al. [223]; Zhao et al. [150]; Kim et al. [28]; Li et al. [224]
Imitation learning	Learns target selection and action policies from expert demonstrations, directly outputting the next target or picking-order preference to generate harvesting decisions end-to-end	Decisions closer to human experience; suitable for reordering among densely clustered targets; no need for explicit reward design	High demonstration-collection cost; errors may accumulate in closed-loop execution; limited adaptability to scenarios beyond the training distribution	Fei et al. [225]; Porichis et al. [226]; Pagliarani et al. [227]; Mahmoudi et al. [228]
Reinforcement learning	Formulates multi-target harvesting as a sequential decision-making problem, and learns target selection and sequencing policies through long-horizon reward optimization, enabling online decision-making and dynamic updating	Adaptive to dynamic changes; jointly considers efficiency and success rate; suitable for closed-loop updating	High training cost; limited sample efficiency; complex reward design; sim-to-real gap; safety constraints	Williams et al. [229]; Zheng et al. [230]; Patil et al. [231]; Zhang et al. [232]

As shown in Table 10, supervised learning is suitable for target scoring or harvestability evaluation, imitation learning is useful for learning human-like selection preferences, and reinforcement learning is more appropriate for dynamic updating and long-term reward optimization. However, learning-based decision-making does not necessarily replace rules or optimization models. For all three learning paradigms, data coverage and cross-scenario distribution shifts affect the reliability of learned target evaluations, action preferences, and decision policies, while execution safety requires learning outputs to be verified against task and motion constraints. A more practical direction is to use learning models to provide target scores or candidate policies and then use rule constraints, risk assessment, feasibility verification, and optimization planning to filter out unsafe actions and coordinate global task objectives. Such a hybrid architecture can reduce reliance on unconstrained exploration and enable learned decision policies to better satisfy the requirements for stability, interpretability, and safety in orchard deployment.

The three categories of decision-making methods exhibit distinct characteristics in terms of computational demand, scalability as the number of targets and coupled constraints increases, real-time performance, and adaptability to dynamic environments. Because the reviewed studies differ in task scale, hardware platforms, implementation settings, and evaluation metrics, a unified experimental comparison is not feasible. Table 11 therefore provides a comprehensive qualitative assessment of the relative characteristics of these methods from implementation-oriented perspectives.

Table 11. Comparison of multi-target harvesting decision-making methods across key implementation dimensions.

Decision-Making Method	Computational Demand	Scalability	Real-Time Performance	Dynamic Adaptability	Key Characteristics
Heuristic methods	Low	Moderate, limited under strongly coupled constraints	High	Low to moderate	Rule-driven, interpretable, fast response, parameter-sensitive
Optimization-based methods	Moderate to high	Moderate, affected by growth in combinatorial scale	Moderate	Moderate	Explicit constraints, global trade-offs, re-optimization burden
Learning-based methods	High during training, low to moderate online	Moderate, constrained by state-space complexity and data coverage	Moderate to high	Moderate to high	Data-driven adaptability, training cost, sim-to-real gap

The comparative characteristics summarized in Table 11 reflect the different formulations and treatment of multi-target harvesting decision problems adopted by the three method categories. Heuristic methods typically rely on predefined priority rules, weighted scoring, or sequence-selection logic. They therefore have low computational demand and fast response, but their ability to handle strongly coupled target relationships, multiple task objectives, and frequent state changes is constrained by the completeness of the rules and parameter settings. Optimization-based methods explicitly formulate target selection, harvesting sequence, path cost, collision constraints, and resource availability as decision variables or constraints. They provide a clearer basis for multi-objective trade-offs and conflict resolution in coupled decisions, but their solution space expands as the numbers of candidate targets, execution units, and constraints increase. Learning-based methods can learn decision policies from historical data or interaction experience and update decisions according to changes in target states. Although online inference after training can be efficient, practical applicability still depends on training cost, data coverage, cross-orchard generalization, and the ability to ensure decision safety and stability under unknown conditions. Therefore, hybrid approaches that combine rule constraints, optimization models, and learning policies are expected to better balance real-time performance, adaptability, and engineering reliability in multi-target harvesting.

Decision-making methods likewise have different levels of validation. Simulation or offline target-set evaluation is suitable for examining the effects of decision rules, objective functions, and constraint modeling under controllable conditions. Laboratory, greenhouse, or semi-structured harvesting tests can further examine whether the generated decisions can be implemented in coordination with actual perception and execution systems. Natural-orchard validation can more directly reflect the ability of a method to address perception uncertainty, execution disturbances, and changes in target availability. Therefore, results such as path length, completion time, or decision quality should also be interpreted together with their corresponding validation levels, distinguishing idealized target-set evaluation from integrated harvesting experiments.

From the perspective of practical deployment, the computational cost of a decision-making method should not be evaluated solely by its independent solution time, but should also be matched to the update period of the task state. The computational time available for online scheduling is jointly constrained by upstream visual inference and three-

dimensional state construction. As the numbers of targets, constraints, or execution units increase, even a scheduling method that produces high-quality solutions may be unable to support effective closed-loop operation if it cannot complete the solution process within the state-update period. Therefore, practical systems need to match decision complexity and update frequency to the available computational resources, and adopt local adjustment or hierarchical scheduling when global optimization cannot be completed in time.

5. Perception and Coordinated Scheduling in Multi-Unit Harvesting Systems

Multi-unit harvesting extends decision-making from a single execution unit to coordinated operation across shared or distributed workspaces. This requires shareable scheduling states that describe target ownership, execution-unit availability, workspace constraints, task progress, and resource utilization. Multi-arm systems primarily involve workspace partitioning, motion coordination, and inter-arm collision avoidance, whereas multi-robot systems further require perception sharing, state synchronization, task allocation, and dynamic scheduling. These coordination requirements constitute the main focus of this section.

5.1. Perception and Collaborative Decision-Making in Multi-Arm Harvesting

Multi-arm harvesting improves efficiency through parallel operation and is suitable for dense-target scenarios or tasks limited by single-arm cyclic operation. However, this efficiency gain depends on shared perception, task division, and safe coordination rather than simply increasing the number of manipulators. Multiple arms need to allocate targets based on a unified target set and spatial coordinate representation, while avoiding repeated assignment, missed targets, and inter-arm conflicts. Therefore, research on multi-arm harvesting mainly focuses on collaborative perception, workspace partitioning, cooperative operation organization, and safety-constraint mechanisms [233]. Task allocation needs to be coupled with trajectory feasibility and collision-avoidance requirements so that task-level efficiency objectives can be translated into executable multi-arm operations.

5.1.1. Multi-Arm Collaborative Perception Strategy

Multi-arm collaborative perception aims to provide shareable and spatially aligned target states for task allocation. Existing visual layouts in multi-arm harvesting systems can be divided into centralized and distributed configurations.

Centralized layouts use global cameras or fixed-view sensors to cover the workspace and provide a unified candidate target set and coordinate reference for multiple manipulators. RGB-D cameras, ToF-RGB-D sensors, and active laser scanning modules have been used in dual-arm grape, apple, and tomato harvesting systems to obtain unified target information and reduce allocation bias caused by inconsistent target representations [234–237]. These methods are convenient for unified scheduling, but fixed viewpoints may provide insufficient local observations under severe occlusion or during close-range operation.

Distributed layouts equip each manipulator with arm-mounted or eye-in-hand visual sensors, allowing each arm to obtain local observations during target approach. Related studies have used RGB-D cameras or eye-in-hand architectures in dual-arm apple and mango harvesting to improve target discovery and local point-cloud acquisition under occlusion [238,239]. Compared with centralized layouts, distributed perception is more suitable for fine localization, but it also requires reliable multi-camera calibration, data synchronization, and cross-arm coordinate consistency.

To enable task allocation based on the same target set, cross-manipulator spatial consistency is essential. Fixed RGB-D or ToF cameras can provide consistent target-localization

results for dual arms within a unified coordinate representation [240,241]. Multi-camera stereo fusion can further construct a global fruit map for multi-arm parallel operation. In field trials of a multi-arm apple-harvesting system, Li et al. [242] reported that their integrated multi-view perception system reduced the median target-localization error by up to 44.43%. This result indicates that complementary viewpoints help improve the reliability of the shared target states on which multi-arm task allocation relies. However, this advantage depends on reliable multi-camera calibration, temporal synchronization, and cross-manipulator coordinate alignment, thereby increasing system-integration requirements. Overall, multi-arm collaborative perception is shifting from single-view target detection toward shared target representation and cross-arm spatial consistency maintenance for task allocation.

5.1.2. Task Area Division for Multi-Arm Harvesting

Task-region partitioning is the basis for parallel operation in multi-arm multi-target harvesting. Since multiple manipulators share the same workspace, unclear target ownership or region boundaries can lead to two arms approaching the same target, local path crossing, or motion interference. Therefore, multi-arm systems usually need to spatially allocate target sets first, allowing different manipulators to work in relatively independent regions while using coordination mechanisms to handle conflicts near shared boundaries.

Existing region-partitioning methods mainly include static spatial partitioning and dynamic load allocation. Static spatial partitioning divides the field of view or workspace into fixed subregions according to platform structure and workspace geometry. Its advantages are clear rules, simple implementation, and low collision risk. Depending on manipulator arrangement, static partitioning can take the form of left–right symmetric partitioning [234], upper–lower layered partitioning [243], left–right fixed partitioning [236], or array-based rectangular cell partitioning [244]. These methods essentially reduce cross-region motion and inter-arm interference by predefining the operating range of each manipulator. Typical region-partitioning modes are shown in Figure 16. Static partitioning is suitable for regular structures and relatively balanced target distributions, but it may cause workload imbalance when fruit density differs across regions.

To improve adaptability to uneven target distribution, dynamic allocation methods introduce load-balancing ideas and adjust task ownership according to target number, spatial position, or real-time manipulator capability. Lei et al. [245] proposed a load-balanced workspace partitioning method for four-arm cooperative harvesting, using fruit three-dimensional coordinates and the Euclidean distance from targets to manipulator bases as workload measures, as shown in Figure 17. Compared with fixed partitioning, such methods are more adaptive to changes in fruit spatial distribution, but they require additional computation for region boundaries and manipulator workload states.

In dual-arm systems, dynamic allocation can further use overlapping workspaces for task adjustment. Yan et al. [246] addressed workload imbalance in dual-arm apple harvesting using an adaptive partitioning method based on the “U-tube” principle. The method minimized the travel difference between the two arms in a semi-overlapping workspace, as shown in Figure 18. In field trials, the dynamic partitioning strategy achieved a dual-arm parallel-operation ratio of 85.7–93.3%, higher than the 50.7–72.4% obtained with the equal-partitioning strategy. The total operating time under dynamic partitioning was 17.8–22.3 s, compared with 22.8–26.2 s under equal partitioning. However, in comparable experiments, dynamic partitioning produced a total dual-arm trajectory length that was 46–359 mm longer than that of equal partitioning. This result indicates that dynamic partitioning improves parallel utilization and task-completion efficiency by prioritizing workload balance rather than simply minimizing cumulative travel distance. Patra et al. [247] further

proposed an online task-allocation method based on the real-time capabilities of the actuators. The method enabled manipulators to share task loads according to their current motion capabilities, thereby reducing trajectory-tracking errors and avoiding actuator saturation. Therefore, static partitioning, with its simple coordination rules and low scheduling overhead, is better suited to regular scenarios with stable target distributions. Dynamic allocation is more suitable when target density or manipulator capabilities are unbalanced, but its online reallocation process needs to control the additional path, switching, and computational costs.

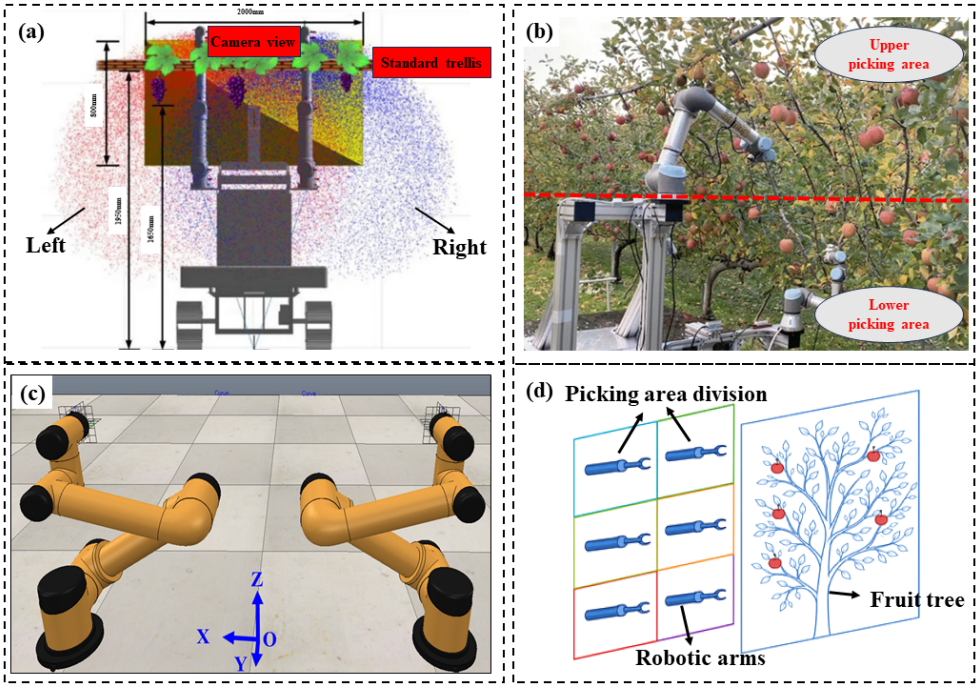


Figure 16. Typical region division and task allocation patterns of multi-arm harvesting robots. (**a**) Symmetrical left–right division [234]; (**b**) hierarchical division [243]; (**c**) fixed left–right partitioning [236]; (**d**) array-type rectangular unit division. (Subfigures (**a**–**c**) were adapted from the corresponding references under the Creative Commons Attribution 4.0 License).

Overall, task-region partitioning needs to balance implementation complexity, workload balance, path continuity, and collision safety. Static partitioning is suitable for regular scenarios and stable workflows, whereas dynamic allocation is more suitable for uneven target density or significant differences in manipulator capability. After region boundaries are determined, the system still needs cooperative operation strategies to arrange action timing and role division among arms, reducing execution interference in shared workspaces.

At this stage, the task-allocation mechanism determines how multi-arm systems respond to changes in target distribution and execution constraints. Fixed partitioning and predefined priorities are suitable for relatively stable workspaces, whereas heuristic and learning-based methods support adjustments of task ownership and operating sequences as task states change [248]. In practice, these adjustments must be coupled with trajectory-feasibility and collision checks within a unified state-update period to maintain stable multi-arm operation.

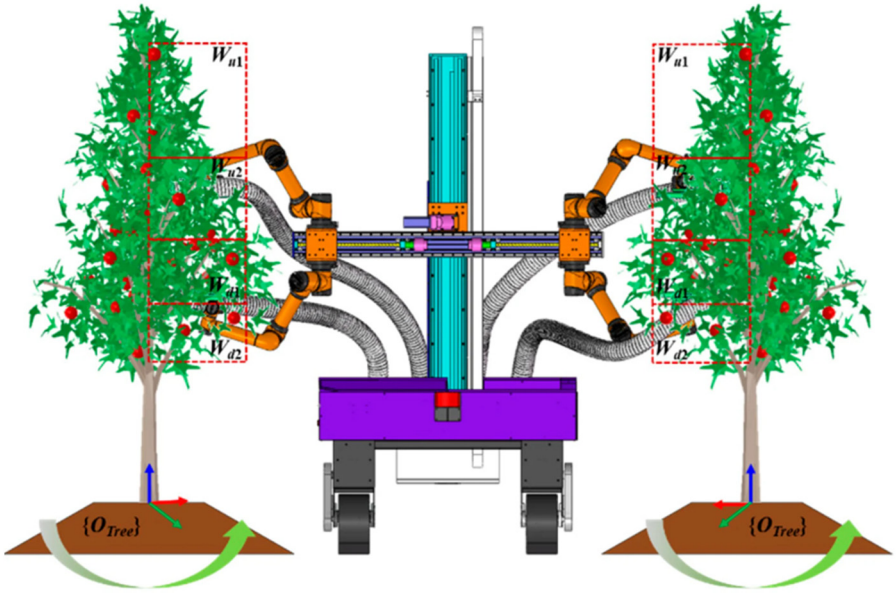


Figure 17. Schematic diagram of workspace partitioning for four robotic arms based on load balancing constraints (adapted from Ref. [245] under the Creative Commons Attribution 4.0 License (CC BY 4.0)).

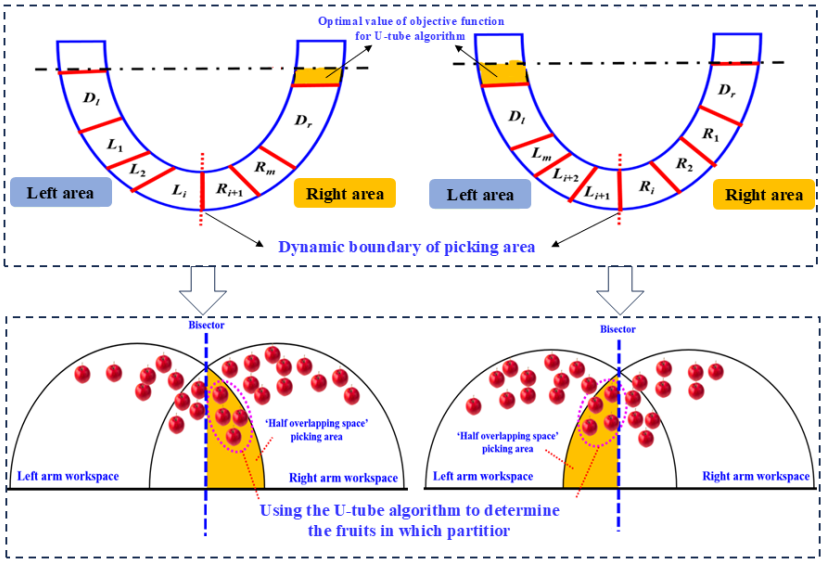


Figure 18. Schematic diagram of the adaptive zoning strategy for fruit picking using a dual-arm system based on the "U-tube" algorithm (adapted from Ref. [246] under the Creative Commons Attribution 4.0 License (CC BY 4.0)).

5.1.3. Collaborative Operation Strategy for Multi-Arm Harvesting

Task-region partitioning can clarify target ownership in multi-arm systems, but during execution, manipulators may still interfere with each other because of shared workspaces, neighboring targets, or crossing paths. Therefore, multi-arm harvesting requires not only task-level target allocation, but also execution-level organization of action timing and role relationships, so that multiple manipulators can cooperate stably within the same workspace.

A common strategy is to reduce motion conflicts in shared spaces through temporal coordination. This type of method does not change the harvesting action itself, but avoids multiple manipulators entering high-risk regions at the same time through master–slave priority, staggered entry, or alternating execution. Chen et al. [237] divided dual-arm tomato harvesting into multiple subtasks and incorporated inter-arm collision avoidance and end-effector pose constraints into coordinated planning. On a physical dual-arm experimental

platform, the method achieved a harvesting success rate of 88.9% and an average harvesting time of 9.5 s/fruit. By comparison, the RRT method achieved a harvesting success rate of 80.6% and an average time of 18.4 s/fruit. Conventional RL achieved a harvesting success rate of 55.5% and an average time of 12.4 s/fruit. Zhu et al. [241] introduced feedback-based rhythm adjustment and alternating execution for execution failures and shared-resource constraints in dual-arm apple harvesting, reducing waiting and blocking. Huang et al. [240] adopted a multi-apple harvesting mode in which the main arm executed key actions first and the auxiliary arm participated at suitable moments, improving continuous operation under dense targets or limited space. However, when target states change or harvesting failures occur, relying only on predefined priorities or fixed action sequences may not meet task requirements. Xie et al. [189] compared DRL-based dynamic task planning with heuristic search strategies in multi-arm harvesting simulations. For static layouts containing 25 and 50 fruits, the DRL strategy reduced execution time by 10.7% and 3.1%, respectively. In dynamic tests involving an initial harvesting failure, heuristic scheduling left four unharvested targets in every 25-target trial and nine in every 50-target trial. The dynamic strategy reduced these numbers to zero and one to two, respectively. However, because the system reallocated manipulators to attempt failed targets, the dynamic strategy resulted in longer total operating time in some trials. These results indicate that fixed temporal coordination is suitable for scenarios with stable task states and low scheduling-overhead requirements, whereas dynamic collaboration is more valuable when task-completion completeness and disturbance adaptability are prioritized [249].

When occlusion, branch interference, or fruit swing affects harvestability, multi-arm systems can improve success rates through functional complementarity. Unlike single-arm systems that repeatedly adjust posture, multi-arm platforms can assign auxiliary actions such as stabilization, leaf pushing, or branch support to one manipulator, while another manipulator performs harvesting. Sepúlveda et al. [250] used one manipulator to move obstructing leaves aside and the other to grasp and harvest eggplants. In laboratory trials simulating typical harvesting conditions, the system achieved a harvesting success rate of 91.67% and an average harvesting cycle of 26 s/fruit. Liu et al. [251] used one manipulator to stabilize the pedicel and the other to grasp and twist the fruit for separation in avocado harvesting. In ten physical trials, the pedicel-stabilizing manipulator achieved a success rate of 100%, whereas the fruit-grasping manipulator achieved a success rate of 80%, indicating the importance of stable auxiliary actions for subsequent fruit separation. In a three-arm system, Liu et al. [252] further decoupled pedicel stabilization from fruit harvesting through role specialization. The three-arm system achieved a single-attempt success rate of 90% in laboratory trials, and field trials further verified the feasibility of the hierarchical coordination framework in natural environments. Overall, functional complementarity is particularly valuable when target accessibility or operational stability is the primary constraint. When targets are relatively easy to access, multi-arm systems can instead adopt more mutually independent parallel harvesting to maintain continuous operating efficiency.

Post-harvest coordination mainly concerns fruit support, collection, and transfer after separation. For cutting-based or damage-sensitive harvesting, lack of support may cause fruit dropping or collision; if the same arm performs both harvesting and placement serially, the continuous operation cycle is also slowed down. Stavridis et al. [253] proposed a dual-arm harvesting scheme in which one arm supported the crop body while the other performed cutting, synchronizing separation and support. Gursoy et al. [254] further incorporated cutting, collection, and transfer into a dual-arm cooperative task description, reducing the time for which the harvesting arm is occupied by post-harvest handling. Lammers et al. [235] integrated fruit handling and boxing units into a dual-arm apple harvesting

robot, reducing the influence of post-harvest handling on continuous harvesting rhythm at the system level. Thus, multi-arm cooperative operation not only improves parallel harvesting efficiency, but also reduces execution risks caused by occlusion, swinging, and post-harvest damage through role division.

5.1.4. Inter-Arm Collision Avoidance Mechanisms in Multi-Arm Harvesting Systems

Multi-arm harvesting improves parallel efficiency, but it also increases safety risks in shared workspaces. Multiple manipulators move simultaneously within limited canopy spaces, and targets are often densely distributed. Without effective constraints, inter-arm collisions, path interference, or motion interruptions may occur. Inter-arm collision avoidance is not only related to system safety, but also affects the feasibility of task allocation and the efficiency of parallel operation. It therefore needs to be considered together with collaborative decision-making.

Existing inter-arm collision avoidance methods can be roughly divided into four categories: safety-distance constraints, priority and yielding coordination, predictive collision detection and online correction, and velocity-scaling avoidance. Safety-distance constraints monitor the minimum distance between links, joints, or end-effectors in real time and trigger deceleration, pausing, or posture adjustment when the distance approaches a threshold. These methods are simple and real-time, but may be conservative in dense operations [255]. Priority and yielding coordination assigns action priorities and allows some manipulators to slow down or wait when potential conflicts occur, so that key actions can be completed first and frequent stopping has less impact on the operation rhythm [256]. Predictive collision detection and online correction identify potential conflicts from short-term motion trends and trigger local path adjustment or online replanning. These methods can reduce efficiency loss caused by excessive conservatism, but require higher computational capacity and control stability [257,258]. Velocity-scaling avoidance continuously adjusts motion speed according to inter-arm proximity without changing the overall task plan, allowing manipulators to slow down smoothly near risky regions. It is relatively easy to implement, but with narrow spaces or frequent crossing motions, it usually needs to be combined with other mechanisms [259,260]. The typical implementations, advantages, and limitations of these methods are shown in Table 12.

Table 12. Comparison of inter-arm collision avoidance mechanisms in multi-arm systems.

Method	Typical Implementation Strategy	Advantages	Limitations	Representative References
Safety-distance constraint	The minimum inter-arm distance is monitored in real time; when a predefined threshold is approached, deceleration, temporary stop, or pose adjustment is triggered	Simple implementation, strong real-time performance, easy engineering deployment	Conservative; may increase waiting time in dense operation scenarios	Lei et al. [255]; Prakash et al. [261]; Wang et al. [262]
Priority and yielding coordination	When a conflict occurs, yielding and staggered execution are scheduled based on predefined rules or priority levels, reducing the probability of multiple arms entering the conflict region simultaneously	Good operational continuity, reduced frequent stopping	Sensitive to rules and parameter settings; may lead to suboptimal waiting under complex and crowded conditions	Yang et al. [263]; Wei et al. [264]; Li et al. [265]

Table 12. *Cont.*

Method	Typical Implementation Strategy	Advantages	Limitations	Representative References
Predictive collision avoidance and online correction	Potential conflicts are identified through short-horizon prediction, and local adjustment or task reallocation is triggered in advance to avoid entering high-risk states	Less conservative, higher parallel efficiency	High computational cost; difficult to balance real-time performance and stability	Wang et al. [266]; Ying et al. [267]; Dong et al. [268]
Velocity-scaling-based avoidance	The motion pace is continuously adjusted according to the degree of inter-arm proximity, enabling smooth avoidance without changing the task organization	Smooth motion, easy engineering implementation	May still be insufficient in crowded intersections or narrow spaces; often needs to be combined with rule-based coordination	Liu et al. [260]; Luo et al. [269]; Wu et al. [259]

As shown in Table 12, the collision avoidance strategy of multi-arm harvesting systems needs to balance operational stability and harvesting efficiency. While existing methods can reduce the risk of inter-arm collisions in most scenarios, they may still encounter problems such as conservative waiting, coordination interruption, or efficiency reduction under conditions of dense targets, significant overlap in shared space, or frequent cross-arm movements. Future inter-arm collision avoidance mechanisms for multi-arm harvesting should improve coordination efficiency and adaptability while ensuring safety. They should reduce reliance on fixed thresholds and manual rules and improve the rapid handling of dynamic conflicts. These advances would better support stable operation in high-density, multi-target harvesting.

Beyond ensuring safe operation, the practical value of multi-arm harvesting also lies in the improved harvesting efficiency resulting from coordinated execution. Shared perception, task partitioning, coordinated operation organization, and collision avoidance jointly determine whether multiple manipulators can sustain effective parallel operation. Table 13 summarizes the reported performance of representative multi-arm harvesting systems and their comparisons with single-arm operation.

Table 13 shows that representative multi-arm systems can improve harvesting performance through parallel execution, expanded operational coverage, or functional complementarity. The table includes simulations, controlled experiments, and natural-orchard validations. Simulations and controlled experiments help isolate the effects of parallel execution, task organization, or functional complementarity, whereas natural-orchard results further reflect system performance under practical conditions involving target distribution, occlusion, and operational disturbances. Depending on system configuration and task characteristics, these mechanisms can reduce harvesting time per fruit, increase fruit-picking throughput, or improve harvesting success rates under the corresponding test conditions. However, increasing the number of manipulators does not necessarily produce proportional efficiency gains. Actual benefits are also affected by target density, task distribution, shared-workspace conflicts, and coordination strategies. Therefore, improved performance in multi-arm harvesting depends not only on the number of execution units, but also on the effective coupling among shared perception, task organization, and safe coordination mechanisms.

Table 13. Reported performance of representative multi-arm harvesting systems and comparison with single-arm operation.

System	Coordination Mode	Validation Level and Setting	Reported Performance	Comparison with Single-Arm Operation	Ref.
Up to 12 Cartesian manipulators	Joint scheduling and trajectory planning	Simulation: orchard scenario, 100 fruits/m^2	Throughput: 2.21 fruits/s with 12 arms	Single-arm: 0.17 fruits/s; 12 arms: 2.21 fruits/s, 12.7-fold improvement	[191]
Dual-arm grape-harvesting robot	Symmetric partitioning and master–slave coordination	Field validation: horizontal-trellis vineyard	9 s/cluster; recognition rate: 88%; harvesting success rate: 83%	Conventional single-arm system described in the study: approximately 25 s/cycle	[234]
Dual-arm apple-harvesting robot	Parallel operation in upper and lower workspaces	Natural-orchard validation: apple orchard	Success rate: 71.2%; dual-arm mode: 4.6 s/fruit	Single-arm mode: 5.8 s/fruit	[235]
Dual-arm mango-harvesting robot	Coordinated target allocation and execution	Natural-orchard validation: mango orchard	Success rate: 73.92%; 8.93 s/fruit	Harvesting time reduced by 48.38% compared with single-arm operation	[239]
Dual-arm apple-harvesting robot	Feedback-based alternating operation	Natural-orchard validation: commercial apple orchard	Success rate: 79.7–80.7%; 5.97 s/cycle	Single-arm baseline: 8.29 s/cycle; time reduced by 28%	[241]
Four-arm apple-harvesting robot	Workspace partitioning and asynchronous operation	Natural-orchard validation: high-density dwarf apple orchard	4.87 s/fruit; visible-fruit success rate: 82%	Single-arm: 9.59 s/fruit; approximately 1.96-fold efficiency improvement	[242]
Aerial dual-arm avocado-harvesting robot	Pedicel stabilization and fruit separation	Controlled validation: avocado grasping tests	Dual-arm grasping success rate: 80% (4/5)	Single-arm grasping success rate: 0% (0/3)	[251]

Note: Reported performance values and efficiency improvement ratios were obtained under different crop types, robot configurations, coordination modes, validation settings, and single-arm baselines. They therefore reflect system-specific performance rather than directly comparable measures across studies.

5.2. Multi-Robot Harvesting Perception and Task Scheduling

Unlike multi-arm systems, where collaboration mainly occurs within the same platform, multi-robot harvesting requires information sharing and task coordination across multiple independent platforms over a larger workspace. The key issue is no longer only whether a single robot can recognize and harvest targets, but whether multiple robots can maintain consistent environmental states and stable task allocation under distributed observations, dynamic target sets, and resource competition. Studies on agricultural multi-robot systems generally indicate that collaborative operation depends on integrating distributed observations into shared environmental and task states, while reducing repeated operations and resource conflicts through dynamic allocation and updating mechanisms [270,271].

As a harvesting-specific example, Lytridis et al. [272] demonstrated cooperative grape harvesting with two heterogeneous autonomous robots in field experiments. An expert robot performed grape harvesting, while a helper robot received and transported the harvested grapes. Their cooperation was achieved through complementary role allocation and coordinated navigation, with the helper robot estimating the relative pose of the expert robot through visual detection of an ArUco marker. Beyond this harvesting-specific

system, general multi-robot SLAM frameworks offer transferable mechanisms for shared environmental representation and inter-robot state synchronization. Representative examples include COVINS and Kimera-Multi, which were developed for general multi-robot operation rather than harvesting-specific applications. Nevertheless, their shared-mapping and cross-robot synchronization capabilities can inform the development of multi-robot harvesting systems. According to their information-fusion and coordination mechanisms, multi-robot SLAM frameworks can be categorized into centralized, distributed, and hybrid architectures. These architectures maintain consistent poses and maps through cross-robot information exchange and provide a state basis for task scheduling [273].

In practical orchard deployment, the timeliness and reliability of this information exchange directly affect the consistency of the shared states on which subsequent task scheduling relies. Centralized frameworks are convenient for building unified reference coordinates and global maps. For example, COVINS enables joint localization and map sharing among multiple robots [274]. However, centralized state fusion depends on communication conditions with sufficiently low latency and high bandwidth. When communication delays occur or the number of robots increases, updates of the global map and task states may lag behind local observations. Distributed frameworks are more suitable for communication-limited or large-scale operations. Kimera-Multi improves the consistency of shared environmental representations through inter-robot map consensus [275]. This architecture reduces dependence on a single communication node, but intermittent communication or data loss may still slow state convergence or cause discrepancies among local maps. For heterogeneous platforms, probabilistic 3D map fusion can integrate local maps generated by different sensors and improve cross-platform observation consistency [276]. Hybrid architectures can retain local observations when full synchronization is temporarily unavailable, thereby balancing global consistency and local autonomy. In orchard scenarios, complementary observations from UGVs and UAVs can also be used to construct three-dimensional environmental representations, providing agricultural examples for sharing canopy structure and workspace information across platforms [277].

At the task-allocation and scheduling level, multi-robot systems must balance target positions, robot states, operation time, energy consumption, and workload [278]. When target sets are relatively stable, combinatorial optimization can improve allocation quality. In the two-robot, five-task instances reported by Guo et al. [279], cooperative evolutionary allocation reduced the maximum task-completion time by 18.04% and 5.50% compared with random and geographically nearest allocation, respectively. For changing target or robot states, one-time allocation is insufficient, and dynamic methods such as improved NSGA-II can incorporate state-triggered updates into multi-objective multiple traveling salesman formulations. Delayed or incomplete state updates, however, may still cause duplicate assignments, missed tasks, and unresolved resource conflicts.

The number of execution units also affects scheduling gains. Lu et al. [280] reported that the total travel distance decreased from 22,100.46 with one robot to 19,262.47 with three robots, a reduction of 12.84%, but increased to 22,240.73 with five robots. Thus, increasing the number of robots does not necessarily improve efficiency, because the benefits of parallel execution also depend on task distribution, path organization, and load balancing.

Engineering-oriented scheduling must further account for operation time, energy consumption, and workload balance. Chen et al. [281] addressed these objectives using a hierarchical reconstruction strategy for electric harvesting robots. Across 45 test instances, the method obtained lower IGD+ values in 71.1% of the cases and higher hypervolume values in 93.3% of the cases. As the number of robots increases, however, the task-allocation scale and state-update communication load also increase, potentially reducing the responsiveness of centralized scheduling and limiting system-level efficiency gains.

Multi-robot harvesting is shifting from optimizing a single task allocation algorithm toward closed-loop coupling among perception sharing, state synchronization, and online scheduling. The performance of this closed-loop process depends not only on the task-allocation algorithm itself, but also on whether shared target and robot states can be synchronized with appropriate timeliness and reliability under changing communication conditions and system scales. Compared with multi-arm systems, multi-robot collaboration needs to balance global-state consistency, communication overhead, and local decision-making autonomy in distributed workspaces. Therefore, future systems should combine lightweight shared-state maintenance, event-triggered information updates, dynamic target-set updates, online reallocation, and local task-execution capability during temporary communication interruptions. This will help multi-robot systems reduce duplicate operations and resource conflicts in large-scale orchard harvesting while maintaining scheduling responsiveness and overall efficiency.

With respect to the available validation evidence, multi-robot studies mainly evaluate completion time, path cost, and scheduling solution quality through task-allocation tests, scenario data, simulations, and benchmark instances. These results can reflect the roles of task-allocation and rescheduling mechanisms. In contrast, validation of a complete harvesting system in natural orchards also needs to assess perception stability, execution success rate, system throughput, and communication reliability.

6. Future Challenges and Prospects

Multi-target harvesting tasks are usually accompanied by dense targets, frequent occlusion, and complex constraints. The system needs to ensure perceptual stability, real-time decision-making, and execution safety during continuous operation. As harvesting robots evolve from single-arm operation to multi-arm parallel harvesting and multi-robot collaboration, operational efficiency and spatial coverage are improved, but cross-agent information sharing, task allocation, and conflict coordination become more challenging. Therefore, future research should further focus on the following directions.

(1) Spatiotemporal consistent perception under occlusion and dense distribution.

In real orchards, branch occlusion, self-occlusion, and dense fruit distribution may cause short-term target loss, identity drift, and attribute-estimation fluctuation during continuous robot motion. Future studies can introduce spatiotemporal attention mechanisms into target detection and multi-target tracking to model cross-frame information jointly [282]. Meanwhile, implicit neural representations and related three-dimensional reconstruction techniques can help complete fruit spatial shapes under limited viewpoints, improving state stability in occluded scenes [283]. However, these methods are computationally expensive, and lightweight deployment on mobile harvesting platforms remains a key issue.

(2) State-representation gap between perception outputs and scheduling inputs.

Existing systems often process perception and scheduling as separate modules. The categories, positions, and keypoints output by perception modules cannot directly satisfy scheduling requirements for priority, reachability, and execution cost. Future studies can introduce structured representations such as dynamic scene graphs to organize target attributes, spatial relationships, occlusion relationships, and reachability constraints in a unified way [284]. Such representations should also retain task progress, execution-unit availability, and state-update time. This would allow local changes, such as target appearance, target disappearance, harvesting failures, or the temporary unavailability of an execution unit, to be incorporated in a timely manner without reconstructing the entire task

state. In this way, perception results can be transformed from discrete detection outputs into state representations directly usable for task allocation and harvesting-sequence planning.

(3) Decision adaptability and solution efficiency under complex constraints.

Multi-target harvesting decisions need to consider reachability, collision avoidance, action timing, and resource allocation simultaneously. Heuristic methods offer good real-time performance but limited adaptability. Optimization-based methods can explicitly represent constraints but incur higher computational costs. Learning-based methods can provide stronger adaptability, but their practical reliability depends on the adequacy of training-data coverage, their ability to generalize across different orchard conditions, and whether safety constraints can be enforced during execution. Future research may develop data-efficient and safety-aware learning-assisted optimization frameworks, in which learning models rapidly generate candidate priority sequences and constrained optimization then verifies feasibility and safety [189]. Combining simulation-based pretraining with targeted data adaptation for real orchards may further reduce data-collection costs and mitigate performance degradation during sim-to-real transfer. Further integration with neuro-symbolic computing may also enhance decision boundaries and interpretability while maintaining environmental adaptability [285].

(4) Task recovery and low-cost reallocation under dynamic disturbances.

During harvesting, changes in leaf and branch occlusion relationships, wind-induced motion of fruits or branches, target-state updates, and individual grasping failures can all alter local task conditions and may render previously generated local trajectories infeasible. Future research may extend model predictive control and event-triggered mechanisms to the task-allocation layer. When conditions permit, local task recovery and reallocation should be prioritized. Global replanning should be triggered only when disturbances substantially affect the remaining task set or resource availability [286]. This can reduce replanning cost while maintaining task continuity.

(5) Collaborative efficiency and collision safety in shared workspaces.

When multiple arms or robots operate in dense regions, path crossing, local waiting, and shared-space conflicts can easily occur. Existing methods based on partitioning, yielding, or velocity adjustment can reduce collision risk, but may still sacrifice parallel efficiency when targets are highly clustered. Future studies can explore multi-agent reinforcement learning to enable multiple execution units to learn more flexible spatiotemporal coordination strategies through interaction [287]. However, such methods still need to be combined with safety constraints to avoid reducing execution reliability in pursuit of efficiency.

(6) State consistency and communication-bandwidth constraints in distributed collaboration.

Continuous multi-target harvesting requires visual inference, three-dimensional state construction, online scheduling, motion control, and safety monitoring to be completed within their respective task-update periods. Therefore, real-time deployment should not be evaluated solely on the basis of algorithmic accuracy, but should also consider the cumulative computational load, memory requirements, energy consumption, and communication conditions of the complete perception-to-decision chain. Future systems may adopt cloud–edge–device collaborative architectures and allocate computational tasks according to latency and reliability requirements. On-board devices should retain time-sensitive functions, including local perception updates, safety responses, and basic task execution. When communication conditions permit, edge nodes can undertake computationally intensive tasks, such as multi-view fusion, three-dimensional reconstruction, and online coordination. Cloud resources can be used for offline model training, model updating, and long-term global analysis. Semantic communication can further reduce transmission demands by

prioritizing task intent, target states, and key features rather than complete raw data [288]. Digital-twin environments can also support low-cost validation of multi-robot scheduling algorithms and facilitate the establishment of unified testing benchmarks and engineering evaluation systems.

7. Conclusions

This paper reviews multi-target harvesting tasks in complex orchard environments from the perspectives of task constraints, visual perception, decision-making methods, and multi-agent collaboration. The main conclusions are as follows.

(1) The core difficulty of multi-target harvesting lies in the coupling between unstructured environments and multiple constraints. Dense fruit distribution, branch and leaf occlusion, dynamic illumination, and fruit attribute variation increase perceptual uncertainty, while manipulator reachability, collision avoidance, and damage control further limit execution feasibility. Therefore, multi-target harvesting is not simply a problem of detecting multiple fruits, but a complex task jointly constrained by perception, decision-making, and execution.

(2) Visual perception and multi-target decision-making have made important progress and have formed several core technical routes, but their system-level integration remains incomplete. Existing perception studies cover sensor configuration, target recognition, attribute identification, spatial localization, environmental modeling, and spatiotemporal consistency maintenance. Decision-making methods have also formed three main routes: heuristic, optimization-based, and learning-based methods. Heuristic methods have strong real-time performance, optimization-based methods are suitable for global constraint modeling, and learning-based methods show potential in complex-environment adaptation. However, a high FPS of an individual visual model does not guarantee end-to-end real-time performance, as the system response is also affected by the computational overhead of spatial-state construction, target-state maintenance, and online scheduling. Practical systems often need to integrate these methods according to task scale and environmental complexity.

(3) Multi-arm and multi-robot systems provide new engineering paths for improving harvesting efficiency. Multi-arm harvesting relies on collaborative perception, region partitioning, action organization, and inter-arm collision avoidance to improve parallel efficiency. Multi-robot harvesting places more emphasis on cross-platform perception sharing, state synchronization, and online task scheduling. Overall, multi-target harvesting has shown potential to improve orchard operation efficiency, but long-term stable application still depends on breakthroughs in continuous perception, unified state representation, real-time decision-making under strong constraints, and multi-agent collaborative scheduling.

Although this paper reviews the perception and decision-making research in multi-target harvesting, certain limitations remain.

(1) Existing studies focus more on perception accuracy and decision algorithms, while research on how perception outputs are organized, filtered, and transformed into scheduling-oriented state representations remains scattered. Therefore, this review cannot fully cover all engineering implementation details of the state transformation process among perception, decision-making, and scheduling.

(2) Multi-arm and multi-robot harvesting systems differ greatly in platform structure, crop type, sensor configuration, and operation workflow and their collaborative mechanisms are often strongly related to specific system designs. Therefore, the summary of general design principles for multi-agent harvesting remains limited.

(3) Public studies provide limited information on communication constraints, long-term operational stability, and recovery under abnormal conditions, leaving gaps in the discussion of system-level robustness in multi-agent collaborative harvesting. Future research should further strengthen unified benchmarks, long-term field validation, and integrated perception–decision–scheduling system design, so as to promote multi-target harvesting robots from experimental verification to stable application.

(4) The literature search in this review was conducted using the Web of Science platform. This platform provides broad cross-disciplinary coverage and supports standardized subject-area filtering. However, databases may differ in source coverage, coverage of conference proceedings, and indexing updates. Therefore, a single-platform search may affect the completeness of coverage for some interdisciplinary studies. Future reviews may integrate other bibliographic databases under consistent screening rules and perform deduplication and cross-checking to broaden the literature coverage of interdisciplinary studies.

Author Contributions: Conceptualization, W.J.; methodology, W.J. and Z.S.; software, Z.S.; validation, M.W.; formal analysis, M.W.; investigation, B.X.; resources, H.W.; data curation, M.W.; writing—original draft preparation, M.W.; writing—review and editing, M.W. and W.J.; visualization, D.W.; supervision, W.J.; project administration, W.J.; funding acquisition, W.J. and B.X. 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 (No. 61973141) and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (No. PAPD).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT 5.6 to assist in improving the clarity and readability of the writing. The authors have reviewed and edited the content and take full responsibility for the content of the publication.

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

References

1. Jiang, L.; Xu, B.; Husnain, N.; Wang, Q. Overview of agricultural machinery automation technology for sustainable agriculture. *Agronomy* **2025**, *15*, 1471. [CrossRef]
2. 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]
3. Zhang, X.; He, L.; Zhang, J.; Whiting, M.D.; Karkee, M.; Zhang, Q. Determination of key canopy parameters for mass mechanical apple harvesting using supervised machine learning and principal component analysis (PCA). *Biosyst. Eng.* **2020**, *193*, 247–263. [CrossRef]
4. 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]
5. 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]
6. Zhu, Y.; Zhang, S.; Tang, S.; Gao, Q. Research progress and applications of artificial intelligence in agricultural equipment. *Agriculture* **2025**, *15*, 1703. [CrossRef]
7. 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. [CrossRef]
8. Zhang, Q.; Zhao, Z.; Gao, G. Fuzzy comprehensive evaluation for grasping prioritization of stacked fruit clusters based on relative hierarchy factor set. *Agronomy* **2022**, *12*, 663. [CrossRef]
9. Yang, Y.; Han, Y.; Li, S.; Yang, Y.; Zhang, M.; Li, H. Vision based fruit recognition and positioning technology for harvesting robots. *Comput. Electron. Agric.* **2023**, *213*, 108258. [CrossRef]

10. Park, Y.; Seol, J.; Pak, J.; Jo, Y.; Kim, C.; Son, H.I. Human-centered approach for an efficient cucumber harvesting robot system: Harvest ordering, visual servoing, and end-effector. *Comput. Electron. Agric.* **2023**, *212*, 108116. [CrossRef]
11. Dai, N.; Fang, J.; Yuan, J.; Liu, X. 3MSP2: Sequential picking planning for multi-fruit congregated tomato harvesting in multi-clusters environment based on multi-views. *Comput. Electron. Agric.* **2024**, *225*, 109303. [CrossRef]
12. Liu, X.; Song, Z.; Tan, Y.; Yang, S.; Ma, Y. Conflict avoidance task planning strategy for dual-arm cooperative strawberry harvesting robots. *Comput. Electron. Agric.* **2026**, *241*, 111067. [CrossRef]
13. 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]
14. Gao, W.; Liu, J.; Deng, J.; Jiang, Y.; Jin, Y. Research status and trends in universal robotic picking end-effectors for various fruits. *Agronomy* **2025**, *15*, 2283. [CrossRef]
15. Shen, Y.; Shen, Y.; Zhang, Y.; Huo, C.; Shen, Z.; Su, W.; Liu, H. Research progress on path planning and tracking control methods for orchard mobile robots in complex scenarios. *Agriculture* **2025**, *15*, 1917. [CrossRef]
16. Montoya-Cavero, L.-E.; Díaz de León Torres, R.; Gómez-Espinosa, A.; Escobedo Cabello, J.A. Vision systems for harvesting robots: Produce detection and localization. *Comput. Electron. Agric.* **2022**, *192*, 106562. [CrossRef]
17. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
18. Ao, J.; Ji, W.; Yu, X.; Ruan, C.; Xu, B. End-effectors for fruit and vegetable harvesting robots: A review of key technologies, challenges, and future prospects. *Agronomy* **2025**, *15*, 2650. [CrossRef]
19. Zhang, H.; Ji, W.; Xu, B.; Yu, X. Optimizing contact force on an apple picking robot end-effector. *Agriculture* **2024**, *14*, 996. [CrossRef]
20. 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]
21. 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]
22. Wang, T.; Du, W.; Zeng, L.; Su, L.; Zhao, Y.; Gu, F.; Liu, L.; Chi, Q. Design and testing of an end-effector for tomato picking. *Agronomy* **2023**, *13*, 947. [CrossRef]
23. Hu, X.; Zhang, Q.; Hu, C. Simulation and experiment for retractable four-point flexible gripper for grape picking end-effector. *Agronomy* **2025**, *15*, 2813. [CrossRef]
24. Ranjan, A.; Machavaram, R.; Patidar, P. Design and development of a peduncle-holding end effector for robotic harvesting of mango. *Cogent Eng.* **2024**, *11*, 2403706. [CrossRef]
25. Navas, E.; Blanco, K.; Rodríguez-Nieto, D.; Fernández, R. A modular soft gripper with embedded force sensing and an iris-type cutting mechanism for harvesting medium-sized crops. *Actuators* **2025**, *14*, 432. [CrossRef]
26. Ai, X.; Zhang, T.; Yuan, T.; Zheng, X.; Xiong, Z.; Yuan, J. YOLOv8np-RCW: A multi-task deep learning model for comprehensive visual information in tomato harvesting robot. *Int. J. Agric. Biol. Eng.* **2025**, *18*, 246–258. [CrossRef]
27. Zhang, F.; Chen, Z.; Wang, Y.; Bao, R.; Chen, X.; Fu, S.; Tian, M.; Zhang, Y. Research on flexible end-effectors with humanoid grasp function for small spherical fruit picking. *Agriculture* **2023**, *13*, 123. [CrossRef]
28. Kim, J.; Pyo, H.; Jang, I.; Kang, J.; Ju, B.; Ko, K. Tomato harvesting robotic system based on Deep-ToMaToS: Deep learning network using transformation loss for 6D pose estimation of maturity classified tomatoes with side-stem. *Comput. Electron. Agric.* **2022**, *201*, 107300. [CrossRef]
29. Han, D.; Zhang, H.; Li, G.; Wang, G.; Wang, X.; Chen, Y.; Chen, X.; Wen, X.; Yang, Q.; Zhao, R. Development of a bionic picking device for high harvest and low loss rate pod pepper harvesting and related working parameter optimization details. *Agriculture* **2024**, *14*, 859. [CrossRef]
30. Du, C.; Fang, W.; Han, D.; Chen, X.; Wang, X. Design and experimental study of a biomimetic pod-pepper-picking drum based on multi-finger collaboration. *Agriculture* **2024**, *14*, 314. [CrossRef]
31. Zhu, X.; Xu, Y.; Han, C.; You, J.; Zhang, X.; Mao, H.; Ma, X. Design and experiment of in-situ bionic harvesting device for edible sunflower. *Agriculture* **2024**, *14*, 1169. [CrossRef]
32. Yu, X.; Ji, W.; Zhang, H.; Ruan, C.; Xu, B.; Wu, K. Grasping force optimization and DDPG impedance control for apple picking robot end-effector. *Agriculture* **2025**, *15*, 1018. [CrossRef]
33. Pi, J.; Liu, J.; Zhou, K.; Qian, M. An octopus-inspired bionic flexible gripper for apple grasping. *Agriculture* **2021**, *11*, 1014. [CrossRef]
34. 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]
35. 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]

36. Lin, J.; Hu, Q.; Xia, J.; Zhao, L.; Du, X.; Li, S.; Chen, Y.; Wang, X. Non-destructive fruit firmness evaluation using a soft gripper and vision-based tactile sensing. *Comput. Electron. Agric.* **2023**, *214*, 108256. [CrossRef]
37. Ji, W.; Zhang, J.; Xu, B.; Tang, C.; Zhao, D. Grasping mode analysis and adaptive impedance control for apple harvesting robotic grippers. *Comput. Electron. Agric.* **2021**, *186*, 106210. [CrossRef]
38. Zhang, Y.; Li, N.; Zhang, L.; Lin, J.; Gao, X.; Chen, G. A review on the recent developments in vision-based apple-harvesting robots for recognizing fruit and picking pose. *Comput. Electron. Agric.* **2025**, *231*, 109968. [CrossRef]
39. Zhong, W.; Yang, W.; Zhu, J.; Jia, W.; Dong, X.; Ou, M. An improved UNet-based path recognition method in low-light environments. *Agriculture* **2024**, *14*, 1987. [CrossRef]
40. Guan, X.; Shi, L.; Yang, W.; Ge, H.; Wei, X.; Ding, Y. Multi-feature fusion recognition and localization method for unmanned harvesting of aquatic vegetables. *Agriculture* **2024**, *14*, 971. [CrossRef]
41. Villacrés, J.; Vougioukas, S. Assessing a multi-camera system to enhance fruit visibility for robotic harvesting in a V-trellised apple orchard. *Comput. Electron. Agric.* **2024**, *224*, 109164. [CrossRef]
42. Luo, L.; Liu, B.; Chen, M.; Wang, J.; Wei, H.; Lu, Q.; Luo, S. DRL-enhanced 3D detection of occluded stems for robotic grape harvesting. *Comput. Electron. Agric.* **2025**, *229*, 109736. [CrossRef]
43. He, Z.; Liu, Z.; Zhou, Z.; Karkee, M.; Zhang, Q. Improving picking efficiency under occlusion: Design, development, and field evaluation of an innovative robotic strawberry harvester. *Comput. Electron. Agric.* **2025**, *237*, 110684. [CrossRef]
44. Yao, S.; Pan, S.; Bennewitz, M.; Hauser, K. Safe leaf manipulation for accurate shape and pose estimation of occluded fruits. In *Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA), Atlanta, GA, USA, 19–23 May 2025*; IEEE: New York, NY, USA, 2025; pp. 16795–16802. [CrossRef]
45. Liu, X.; Zhang, Z.; Li, Z.; Li, M.; Zhang, M.; Jiang, S. Apple Occlusion and Clustering State Detection in Natural Environments Using Improved YOLOv5s. *IFAC-PapersOnLine* **2025**, *59*, 245–250. [CrossRef]
46. Chu, P.; Li, Z.; Zhang, K.; Chen, D.; Lammers, K.; Lu, R. O2RNet: Occluder-occludee relational network for robust apple detection in clustered orchard environments. *Smart Agric. Technol.* **2023**, *5*, 100284. [CrossRef]
47. Yuan, Y.; Liu, H.; Yang, Z.; Zheng, J.; Li, J.; Zhao, L. A detection method for occluded and overlapped apples under close-range targets. *Pattern Anal. Appl.* **2024**, *27*, 12. [CrossRef]
48. Khan, Z.; Shen, Y.; Liu, H. ObjectDetection in agriculture: A comprehensive review of methods, applications, challenges, and future directions. *Agriculture* **2025**, *15*, 1351. [CrossRef]
49. Chen, J.; Wu, J.; Wang, Z.; Qiang, H.; Cai, G.; Tan, C.; Zhao, C. Detecting ripe fruits under natural occlusion and illumination conditions. *Comput. Electron. Agric.* **2021**, *190*, 106450. [CrossRef]
50. Zuo, Z.; Gao, S.; Peng, H.; Xue, Y.; Han, L.; Ma, G.; Mao, H. Lightweight detection of broccoli heads in complex field environments based on LBDC-YOLO. *Agronomy* **2024**, *14*, 2359. [CrossRef]
51. Zhuang, J.J.; Luo, S.M.; Hou, C.J.; Tang, Y.; He, Y.; Xue, X.Y. Detection of orchard citrus fruits using a monocular machine vision-based method for automatic fruit picking applications. *Comput. Electron. Agric.* **2018**, *152*, 64–73. [CrossRef]
52. 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]
53. Zeeshan, S.; Aized, T.; Riaz, F. In-depth evaluation of automated fruit harvesting in unstructured environment for improved robot design. *Machines* **2024**, *12*, 151. [CrossRef]
54. 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]
55. Liu, Y.; Han, X.; Ren, L.; Ma, W.; Liu, B.; Sheng, C.; Song, Y.; Li, Q. Surface defect and malformation characteristics detection for fresh sweet cherries based on YOLOv8-DCPF method. *Agronomy* **2025**, *15*, 1234. [CrossRef]
56. Li, R.; Ji, Z.; Hu, S.; Huang, X.; Yang, J.; Li, W. Tomato maturity recognition model based on improved YOLOv5 in greenhouse. *Agronomy* **2023**, *13*, 603. [CrossRef]
57. Viveros Escamilla, L.D.; Gómez-Espinosa, A.; Escobedo Cabello, J.A.; Cantoral-Ceballos, J.A. Maturity recognition and fruit counting for sweet peppers in greenhouses using deep learning neural networks. *Agriculture* **2024**, *14*, 331. [CrossRef]
58. Lu, Y.; Gong, M.; Li, J.; Ma, J. Strawberry defect identification using deep learning infrared–visible image fusion. *Agronomy* **2023**, *13*, 2217. [CrossRef]
59. Lu, J.; Chen, W.; Lan, Y.; Qiu, X.; Huang, J.; Luo, H. Design of citrus peel defect and fruit morphology detection method based on machine vision. *Comput. Electron. Agric.* **2024**, *219*, 108721. [CrossRef]
60. Raufer, L.; Wiedey, J.; Mueller, M.; Penava, P.; Buettner, R. A deep learning-based approach for the detection of cucumber diseases. *PLoS ONE* **2025**, *20*, e0320764. [CrossRef] [PubMed]
61. Pinheiro, I.; Moreira, G.; Queirós da Silva, D.; Magalhães, S.; Valente, A.; Moura Oliveira, P.; Cunha, M.; Santos, F. Deep learning YOLO-Based solution for grape bunch detection and assessment of biophysical lesions. *Agronomy* **2023**, *13*, 1120. [CrossRef]
62. Arellano, C.; Hofmann, N.; Sagredo, K.; Muñoz, C.; Govan, J. Deep learning models to detect wax bloom on blueberry fruits from images. *Chil. J. Agric. Res.* **2025**, *85*, 601–610. [CrossRef]

63. Feng, W.; Cheng, X.; Sun, J.; Xiong, Z.; Zhai, Z. Specular highlight removal and depth estimation based on polarization characteristics of light field. *Opt. Commun.* **2023**, *537*, 129467. [CrossRef]
64. Bai, Y.; Guo, Y.; Zhang, Q.; Cao, B.; Zhang, B. Multi-network fusion algorithm with transfer learning for green cucumber segmentation and recognition under complex natural environment. *Comput. Electron. Agric.* **2022**, *194*, 106789. [CrossRef]
65. Xie, L.; Ji, W.; Xu, B.; Wu, D.; Ao, J. Research and implementation of peach fruit detection and growth posture recognition algorithms. *Agriculture* **2026**, *16*, 193. [CrossRef]
66. Muñoz-Postigo, J.; Valero, E.M.; Martínez-Domingo, M.A.; Lara, F.J.; Nieves, J.L.; Romero, J.; Hernández-Andrés, J. Band selection pipeline for maturity stage classification in bell peppers: From full spectrum to simulated camera data. *J. Food Eng.* **2024**, *365*, 111824. [CrossRef]
67. Dai, C.; Sun, J.; Huang, X.; Zhang, X.; Tian, X.; Wang, W.; Sun, J.; Luan, Y. Application of hyperspectral imaging as a nondestructive technology for identifying tomato maturity and quantitatively predicting lycopene content. *Foods* **2023**, *12*, 2957. [CrossRef] [PubMed]
68. Sun, J.; He, X.; Ge, X.; Wu, X.; Shen, J.; Song, Y. Detection of key organs in tomato based on deep migration learning in a complex background. *Agriculture* **2018**, *8*, 196. [CrossRef]
69. Ismail, N.; Malik, O.A. Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Inf. Process. Agric.* **2022**, *9*, 24–37. [CrossRef]
70. Jiang, L.; Wang, Y.; Yan, H.; Yin, Y.; Wu, C. Strawberry fruit deformity detection and symmetry quantification using deep learning and geometric feature analysis. *Horticulturae* **2025**, *11*, 652. [CrossRef]
71. 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]
72. Rong, J.; Liu, Y.; Li, X.; Gao, C.; Wang, P.; Yuan, T.; Li, W. Decoupled motion planning method for 7-DOF manipulator and lifting joint in automated tomato harvesting. *Comput. Electron. Agric.* **2025**, *237*, 110693. [CrossRef]
73. Ma, P.; Zhu, A.; Chen, Y.; Tu, Y.; Mao, H.; Song, J.; Wang, X.; Su, S.; Li, D.; Dong, X. Multi objective motion planning of fruit harvesting manipulator based on improved BIT* algorithm. *Comput. Electron. Agric.* **2024**, *227*, 109567. [CrossRef]
74. Shi, M.; Cheng, Y.; Rong, B.; Zhao, W.; Yao, Z.; Yu, C. Research on vibration suppression and trajectory tracking control strategy of a flexible link manipulator. *Appl. Math. Model.* **2022**, *110*, 78–98. [CrossRef]
75. Yang, H.; Chen, L.; Ma, Z.; Chen, M.; Zhong, Y.; Deng, F.; Li, M. Computer vision-based high-quality tea automatic plucking robot using Delta parallel manipulator. *Comput. Electron. Agric.* **2021**, *181*, 105946. [CrossRef]
76. Fuentes, C.; Sanhueza, P.; Pereira, G.; Arancibia, R.; Pacheco, G.; Godoy, R. Design of Low-Cost Mature-Tomato Automatic Harvesting Robots. In *Recent Advances and Emerging Challenges in STEM*; Springer Nature: Cham, Switzerland, 2024; pp. 499–508. [CrossRef]
77. Wang, S.; Feng, W.; Wang, W.; Liu, J.; Yang, D.; Li, T. Hyper-redundant hybrid manipulator modeling method. *Int. J. Mech. Sci.* **2025**, *301*, 110421. [CrossRef]
78. Elias, A.J.; Wen, J.T. Redundancy parameterization and inverse kinematics of 7-DOF revolute manipulators. *Mech. Mach. Theory* **2024**, *204*, 105824. [CrossRef]
79. Yu, P.; Tan, N.; Zhong, Z.; Liao, S. Model-free kinematic control of redundant manipulators with simultaneous joint-physical-limit and joint-angular-drift handling. *ISA Trans.* **2023**, *139*, 635–649. [CrossRef] [PubMed]
80. 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]
81. 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] [PubMed]
82. Yuan, Y.; Ji, W.; Fang, H.; Yuan, H.; Ding, S.; Jiang, Y.; Zha, Y. Time-optimal trajectory planning for tool-changing robots based on the hybrid search algorithm of sparrow-osprey. *Int. J. Robot. Autom.* **2025**, *40*, 415–424. [CrossRef]
83. Burusa, A.K.; Scholten, J.; Wang, X.; Rapado-Rincón, D.; van Henten, E.J.; Kootstra, G. Semantics-aware next-best-view planning for efficient search and detection of task-relevant plant parts. *Biosyst. Eng.* **2024**, *248*, 1–14. [CrossRef]
84. Sun, T.; Zhang, W.; Gao, X.; Zhang, W.; Li, N.; Miao, Z. Efficient occlusion avoidance based on active deep sensing for harvesting robots. *Comput. Electron. Agric.* **2024**, *225*, 109360. [CrossRef]
85. 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]
86. Bajcsy, R.; Aloimonos, Y.; Tsotsos, J.K. Revisiting active perception. *Auton. Robot.* **2018**, *42*, 177–196. [CrossRef] [PubMed]
87. Zhu, C.; Hao, S.; Liu, C.; Wang, Y.; Jia, X.; Xu, J.; Guo, S.; Huo, J.; Wang, W. An efficient computer vision-based dual-face target precision variable spraying robotic system for foliar fertilisers. *Agronomy* **2024**, *14*, 2770. [CrossRef]
88. 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]

89. Tao, T.; Wei, X. STBNA-YOLOv5: An improved YOLOv5 network for weed detection in rapeseed field. *Agriculture* **2025**, *15*, 22. [CrossRef]
90. 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]
91. Tang, S.; Xia, Z.; Gu, J.; Wang, W.; Huang, Z.; Zhang, W. High-precision apple recognition and localization method based on RGB-D and improved SOLOv2 instance segmentation. *Front. Sustain. Food Syst.* **2024**, *8*, 1403872. [CrossRef]
92. Ma, Z.; Yang, S.; Li, J.; Qi, J. Research on SLAM localization algorithm for orchard dynamic vision based on YOLOD-SLAM2. *Agriculture* **2024**, *14*, 1622. [CrossRef]
93. Qiu, Z.; Martínez-Sánchez, J.; Arias, P. Evaluation and comparison of different time of flight cameras for outdoor applications. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *X-1/W1-2023*, 17–23. [CrossRef]
94. Yang, R.; Miao, W.; Zhang, Z.; Liu, Z.; Li, M.; Lin, B. SA-MVSNet: Self-attention-based multi-view stereo network for 3D reconstruction of images with weak texture. *Eng. Appl. Artif. Intell.* **2024**, *131*, 107800. [CrossRef]
95. Xie, B.; Jin, Y.; Faheem, M.; Gao, W.; Liu, J.; Jiang, H.; Cai, L.; Li, Y. Research progress of autonomous navigation technology for multi-agricultural scenes. *Comput. Electron. Agric.* **2023**, *211*, 107963. [CrossRef]
96. Jin, Y.; Liu, J.; Wang, J.; Xu, Z.; Yuan, Y. Far-near combined positioning of picking-point based on depth data features for horizontal-trellis cultivated grape. *Comput. Electron. Agric.* **2022**, *194*, 106791. [CrossRef]
97. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]
98. Mach, J.; Svatý, Z.; Šoupa, O.; Nouzovský, L.; Halecký, M. Implementation of an SfM-MVS-based photogrammetry approach for detailed 3D reconstruction of plants. *Plant Methods* **2025**, *21*, 127. [CrossRef] [PubMed]
99. Schönberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016; pp. 4104–4113. [CrossRef]
100. Qin, X.; Zhu, Y.; Wang, L.; Zhang, X.; He, C.; Dong, Q. Self-supervised monocular depth learning from unknown cameras: Leveraging the power of raw data. *Image Vis. Comput.* **2025**, *157*, 105505. [CrossRef]
101. Shiu, Y.C.; Ahmad, S. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB. *IEEE Trans. Robot. Autom.* **1989**, *5*, 16–29. [CrossRef]
102. Zhang, X.; Yao, M.; Cheng, Q.; Liang, G.; Fan, F. A novel hand-eye calibration method of picking robot based on TOF camera. *Front. Plant Sci.* **2023**, *13*, 1099033. [CrossRef] [PubMed]
103. Jiang, Z.; Zhou, J.; Han, H.; Wang, Y. An apple fruit localization system based on accurate and flexible hand-eye pose acquisition for robotic harvesting. *Comput. Electron. Agric.* **2025**, *229*, 109705. [CrossRef]
104. Yan, T.; Li, P.; Liu, Y.; Jia, T.; Yu, H.; Chen, G. Research on hand–eye calibration accuracy improvement method based on iterative closest point algorithm. *Agriculture* **2023**, *13*, 2026. [CrossRef]
105. Zhu, Y.Y.; Cui, B.B.; Yu, Z.L.; Gao, Y.Y.; Wei, X.H. Tillage depth detection and control based on attitude estimation and online calibration of model parameters. *Agriculture* **2024**, *14*, 2130. [CrossRef]
106. Li, T.; Qiu, Q.; Zhao, C. Hybrid visual servoing tracking control of uncalibrated robotic systems for dynamic dwarf culture orchards harvest. In *Proceedings of the 2021 IEEE International Conference on Development and Learning (ICDL), Beijing, China, 23–26 August 2021*; IEEE: New York, NY, USA, 2021; pp. 1–6. [CrossRef]
107. Li, T.; Yu, J.; Qiu, Q.; Zhao, C. Hybrid uncalibrated visual servoing control of harvesting robots with RGB-D cameras. *IEEE Trans. Ind. Electron.* **2023**, *70*, 2729–2738. [CrossRef]
108. Sun, Q.; Chai, X.; Zeng, Z.; Zhou, G.; Sun, T. Noise-tolerant RGB-D feature fusion network for outdoor fruit detection. *Comput. Electron. Agric.* **2022**, *198*, 107034. [CrossRef]
109. 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]
110. Jiang, H.; Liu, J.; Lei, X.; Xu, B.; Jin, Y. Multi-stage fusion of dual attention mask R-CNN and geometric filtering for fast and accurate localization of occluded apples. *Artif. Intell. Agric.* **2026**, *16*, 187–205. [CrossRef]
111. 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]
112. Ma, J.; Li, M.; Fan, W.; Liu, J. State-of-the-art techniques for fruit maturity detection. *Agronomy* **2024**, *14*, 2783. [CrossRef]
113. Peng, Y.; Zhao, S.; Liu, J. Fused deep features-based grape varieties identification using support vector machine. *Agriculture* **2021**, *11*, 869. [CrossRef]
114. Peng, H.; Chen, H.; Zhang, X.; Liu, H.; Chen, K.; Xiong, J. Retinanet_G2S: A multi-scale feature fusion-based network for fruit detection of punna navel oranges in complex field environments. *Precis. Agric.* **2024**, *25*, 889–913. [CrossRef]
115. Lu, S.; Chen, W.; Zhang, X.; Karkee, M. Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. *Comput. Electron. Agric.* **2022**, *193*, 106696. [CrossRef]
116. Zhang, D.; Hou, R.; Guo, J.; Zhao, M.; Wang, Q.; Luo, Z.; Xu, K. A conditional segmentation-guided network for pomegranate image completion under occlusion. *Plant Methods* **2025**, *21*, 153. [CrossRef] [PubMed]

117. Chen, J.; Ma, A.; Huang, L.; Li, H.; Zhang, H.; Huang, Y.; Zhu, T. Efficient and lightweight grape and picking point synchronous detection model based on key point detection. *Comput. Electron. Agric.* **2024**, *217*, 108612. [CrossRef]
118. Ni, J.; Zhu, L.; Dong, L.; Wang, R.; Chen, K.; Gao, J.; Wang, W.; Zhou, L.; Zhao, B.; Rong, J.; et al. TomatoPoseNet: An efficient keypoint-based 6D pose estimation model for non-destructive tomato harvesting. *Agronomy* **2024**, *14*, 3027. [CrossRef]
119. Ma, Z.; Dong, N.; Gu, J.; Cheng, H.; Meng, Z.; Du, X. STRAW-YOLO: A detection method for strawberry fruits targets and key points. *Comput. Electron. Agric.* **2025**, *230*, 109853. [CrossRef]
120. Zu, L.; Zhao, Y.; Liu, J.; Su, F.; Zhang, Y.; Liu, P. Detection and segmentation of mature green tomatoes based on Mask R-CNN with automatic image acquisition approach. *Sensors* **2021**, *21*, 7842. [CrossRef] [PubMed]
121. Crespo, A.; Moncada, C.; Crespo, F.; Morocho-Cayamcela, M.E. An efficient strawberry segmentation model based on Mask R-CNN and TensorRT. *Artif. Intell. Agric.* **2025**, *15*, 327–337. [CrossRef]
122. Baek, N.R.; Lee, Y.; Noh, D.H.; Lee, H.M.; Cho, S.W. AS-YOLO: Enhanced YOLO using ghost bottleneck and global attention mechanism for apple stem segmentation. *Sensors* **2025**, *25*, 1422. [CrossRef] [PubMed]
123. Wang, D.; Wang, X.; Chen, Y.; Wu, Y.; Zhang, X. Strawberry ripeness classification method in facility environment based on red color ratio of fruit rind. *Comput. Electron. Agric.* **2023**, *214*, 108313. [CrossRef]
124. Wang, C.; Han, Q.; Li, C.; Zou, T.; Zou, X. Fusion of fruit image processing and deep learning: A study on identification of citrus ripeness based on R-LBP algorithm and YOLO-CIT model. *Front. Plant Sci.* **2024**, *15*, 1397816. [CrossRef] [PubMed]
125. Liu, G.; Mao, S.; Kim, J.H. A mature-tomato detection algorithm using machine learning and color analysis. *Sensors* **2019**, *19*, 2023. [CrossRef] [PubMed]
126. Xu, L.; Chen, Y.; Wang, X.; Chen, H.; Tang, Z.; Shi, X.; Chen, X.; Wang, Y.; Kang, Z.; Zou, Z.; et al. Non-destructive detection of kiwifruit soluble solid content based on hyperspectral and fluorescence spectral imaging. *Front. Plant Sci.* **2023**, *13*, 1075929. [CrossRef] [PubMed]
127. Tian, Y.; Sun, J.; Zhou, X.; Cong, S.; Dai, C.; Shi, L. Nondestructive detection of soluble solids content in apples based on multi-attention convolutional neural network and hyperspectral imaging technology. *Foods* **2025**, *14*, 3832. [CrossRef] [PubMed]
128. Guo, Z.; Chen, X.; Zhang, Y.; Sun, C.; Jayan, H.; Majeed, U.; Watson, N.J.; Zou, X. Dynamic nondestructive detection models of apple quality in critical harvest period based on near-infrared spectroscopy and intelligent algorithms. *Foods* **2024**, *13*, 1698. [CrossRef] [PubMed]
129. Lauretti, C.; Tamantini, C.; Zompanti, A.; Cimini, S.; De Gara, L.; Santonico, M.; Zollo, L. An Intelligent robotic platform for fruit selective harvesting. In *European Robotics Forum 2024*; Springer Nature: Cham, Switzerland, 2024; pp. 252–257. [CrossRef]
130. Bertoglio, R.; Piliego, M.; Guadagna, P.; Gatti, M.; Poni, S.; Matteucci, M. On-the-go table grape ripeness estimation via proximal snapshot hyperspectral imaging. *Comput. Electron. Agric.* **2024**, *226*, 109354. [CrossRef]
131. Meng, Q.; Tan, T.; Feng, S.; Wen, Q.; Shang, J. Prediction and visualization map for physicochemical indices of kiwifruits by hyperspectral imaging. *Front. Nutr.* **2024**, *11*, 1364274. [CrossRef] [PubMed]
132. Kim, D.G.; Burks, T.F.; Qin, J.; Bulanon, D.M. Classification of grapefruit peel diseases using color texture feature analysis. *Int. J. Agric. Biol. Eng.* **2009**, *2*, 41–50. [CrossRef]
133. Blasco, J.; Aleixos, N.; Moltó, E. Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm. *J. Food Eng.* **2007**, *81*, 535–543. [CrossRef]
134. Liang, X.; Jia, X.; Huang, W.; He, X.; Li, L.; Fan, S.; Li, J.; Zhao, C.; Zhang, C. Real-Time grading of defect apples using semantic segmentation combination with a pruned YOLO V4 network. *Foods* **2022**, *11*, 3150. [CrossRef] [PubMed]
135. Miao, H.; Wang, Z.; Zhang, Y.; Lu, X.; Lou, Y.; Li, X.; Ma, J.; Wang, Y.; Huang, S.; Xu, S.; et al. New kiwifruit quality grading methods based on image multi-feature fusion and ResTNet model. *Expert Syst. Appl.* **2026**, *297*, 129507. [CrossRef]
136. Yang, W.; He, Z.; Song, Y.; Ma, Y. 3D point cloud denoising method based on global feature guidance. *Vis. Comput.* **2024**, *40*, 6137–6153. [CrossRef]
137. Ng, H.T.; Ibrahim, H.; Rajendran, P. Statistical-based methods to improve precision of DBSCAN clustering algorithm for obstacle detection application in autonomous vehicles. *Multimed. Tools Appl.* **2025**, *84*, 47101–47125. [CrossRef]
138. Palos, M.; Cortés, I.; Madridano, Á.; Navas, F.; Barbero, C.; Milanés, V.; García, F. LiDAR-based perception system for logistics in industrial environments. *Appl. Intell.* **2025**, *55*, 696. [CrossRef]
139. Liu, E.; Lou, J.; Tang, W.; Wu, J.; Liang, X. A Trunk Map Construction Method for Long-Term Localization and Navigation for Orchard Robots. In *Intelligent Robotics and Applications*; Springer Nature: Singapore, 2023; pp. 442–455. [CrossRef]
140. Ye, L.; Ma, J.; Lv, Y.; Guo, Z.; Lai, Z.; Ou, C.; Li, J.; Wu, F. The YOLO-OBB-Based approach for citrus fruit stem pose estimation and robot picking. *Agriculture* **2025**, *15*, 2330. [CrossRef]
141. Qi, Y.; Dong, X.; Chen, P.; Lee, K.-H.; Lan, Y.; Lu, X.; Jia, R.; Deng, J.; Zhang, Y. Canopy Volume Extraction of Citrus reticulate Blanco cv. Shatangju Trees Using UAV Image-Based Point Cloud Deep Learning. *Remote Sens.* **2021**, *13*, 3437. [CrossRef]
142. Walker, V.; Vanegas, F.; Gonzalez, F. NanoMap: A GPU-Accelerated OpenVDB-Based Mapping and Simulation Package for Robotic Agents. *Remote Sens.* **2022**, *14*, 5463. [CrossRef]

143. Xiong, J.; Liang, J.; Zhuang, Y.; Hong, D.; Zheng, Z.; Liao, S.; Hu, W.; Yang, Z. Real-time localization and 3D semantic map reconstruction for unstructured citrus orchards. *Comput. Electron. Agric.* **2023**, *213*, 108217. [CrossRef]
144. Burusa, A.K.; van Henten, E.J.; Kootstra, G. Attention-driven next-best-view planning for efficient reconstruction of plants and targeted plant parts. *Biosyst. Eng.* **2024**, *246*, 248–262. [CrossRef]
145. Niu, J.; Zhang, L.; Zhang, T.; Guan, J.; Shi, S. Orchard robot navigation via an improved RTAB-Map algorithm. *Appl. Sci.* **2025**, *15*, 11673. [CrossRef]
146. Bu, X.; Liu, C.; Liu, H.; Yang, G.; Shen, Y.; Xu, J. DFSNet: A 3D point cloud segmentation network toward trees detection in an orchard scene. *Sensors* **2024**, *24*, 2244. [CrossRef] [PubMed]
147. Liu, P.; Chai, X.Y.; Cui, Z.H.; Zhu, Z.H.; Hu, J.P.; Yang, W.P.; Xu, L.Z. StarNet-RiceSeg: An efficient high-dimensional feature mapping network with spatial attention for real-time rice lodging detection. *Agriculture* **2026**, *16*, 775. [CrossRef]
148. Pan, Y.; Hu, K.; Cao, H.; Kang, H.; Wang, X. A novel perception and semantic mapping method for robot autonomy in orchards. *Comput. Electron. Agric.* **2024**, *219*, 108769. [CrossRef]
149. Xie, B.; Liu, J.; Jiang, H.; Cai, L.; Liu, L.; Li, Y. Development and experimental analysis of a small integrated edge navigation sensor based on principle of circular arc array. *Front. Plant Sci.* **2022**, *13*, 892388. [CrossRef] [PubMed]
150. Zhao, G.; Dong, S.; Wen, J.; Ban, Y.; Zhang, X. Selective fruit harvesting prediction and 6D pose estimation based on YOLOv7 multi-parameter recognition. *Comput. Electron. Agric.* **2025**, *229*, 109815. [CrossRef]
151. Gao, R.; Zhou, Q.; Cao, S.; Jiang, Q. An algorithm for calculating apple picking direction based on 3D vision. *Agriculture* **2022**, *12*, 1170. [CrossRef]
152. Niu, J.; Bi, M.; Yu, Q. Apple pose estimation based on SCH-YOLO11s segmentation. *Agronomy* **2025**, *15*, 900. [CrossRef]
153. Kang, H.; Zhou, H.; Chen, C. Visual perception and modeling for autonomous apple harvesting. *IEEE Access* **2020**, *8*, 62151–62163. [CrossRef]
154. Lehnert, C.; Sa, I.; McCool, C.; Upcroft, B.; Perez, T. Sweet pepper pose detection and grasping for automated crop harvesting. In *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016*; IEEE: New York, NY, USA, 2016; pp. 2428–2434. [CrossRef]
155. Luo, L.; Yin, W.; Ning, Z.; Wang, J.; Wei, H.; Chen, W.; Lu, Q. In-field pose estimation of grape clusters with combined point cloud segmentation and geometric analysis. *Comput. Electron. Agric.* **2022**, *200*, 107197. [CrossRef]
156. Zhang, Q.; Pang, Y.; Li, B. Visual positioning and picking pose estimation of tomato clusters based on instance segmentation. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 205–215. [CrossRef]
157. Wang, X.; Kang, H.; Zhou, H.; Au, W.; Chen, C. Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. *Comput. Electron. Agric.* **2022**, *193*, 106716. [CrossRef]
158. Ci, J.; Wang, X.; Rapado-Rincón, D.; Burusa, A.K.; Kootstra, G. 3D pose estimation of tomato peduncle nodes using deep keypoint detection and point cloud. *Biosyst. Eng.* **2024**, *243*, 57–69. [CrossRef]
159. Park, Y.; Pak, J.; Kim, C.; Son, H.I. 3D point cloud-based 6D pose estimation using the pedicel morphological features of fruits and vegetables. *Robot. Auton. Syst.* **2025**, *194*, 105151. [CrossRef]
160. He, Z.; Yuan, F.; Zhou, Y.; Cui, B.; He, Y.; Liu, Y. Stereo vision based broccoli recognition and attitude estimation method for field harvesting. *Artif. Intell. Agric.* **2025**, *15*, 526–536. [CrossRef]
161. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016*; IEEE: New York, NY, USA, 2016; pp. 3464–3468. [CrossRef]
162. Wang, Z.; Walsh, K.; Koirala, A. Mango fruit load estimation using a video based mangoYOLO—Kalman filter—Hungarian algorithm method. *Sensors* **2019**, *19*, 2742. [CrossRef] [PubMed]
163. Shen, L.; Su, J.; He, R.; Song, L.; Huang, R.; Fang, Y.; Song, Y.; Su, B. Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLOv5s. *Comput. Electron. Agric.* **2023**, *206*, 107662. [CrossRef]
164. Rong, J.; Zhou, H.; Zhang, F.; Yuan, T.; Wang, P. Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion. *Comput. Electron. Agric.* **2023**, *207*, 107741. [CrossRef]
165. Zhang, W.; Wang, J.; Liu, Y.; Chen, K.; Li, H.; Duan, Y.; Wu, W.; Shi, Y.; Guo, W. Deep-learning-based in-field citrus fruit detection and tracking. *Hortic. Res.* **2022**, *9*, uhac003. [CrossRef] [PubMed]
166. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017*; IEEE: New York, NY, USA, 2017; pp. 3645–3649. [CrossRef]
167. Lin, Y.; Hu, W.; Zheng, Z.; Xiong, J. Citrus identification and counting algorithm based on improved YOLOv5s and DeepSort. *Agronomy* **2023**, *13*, 1674. [CrossRef]
168. Zhou, L.; Yang, Z.; Fu, L.; Duan, J. Yield estimation in banana orchards based on DeepSORT and RGB-Depth images. *Agronomy* **2025**, *15*, 1119. [CrossRef]

169. Ge, Y.; Lin, S.; Zhang, Y.; Li, Z.; Cheng, H.; Dong, J.; Shao, S.; Zhang, J.; Qi, X.; Wu, Z. Tracking and counting of tomato at different growth period using an improving YOLO-Deepsort network for inspection robot. *Machines* **2022**, *10*, 489. [CrossRef]
170. An, Q.; Cui, Y.; Tong, W.; Liu, Y.; Zhao, B.; Wei, L. StraTracker: A dynamic counting method for growing strawberries based on multi-target tracking. *Comput. Electron. Agric.* **2024**, *227*, 109564. [CrossRef]
171. Du, P.; Chen, S.; Li, X.; Hu, W.; Lan, N.; Lei, X.; Xiang, Y. Green pepper fruits counting based on improved DeepSort and optimized Yolov5s. *Front. Plant Sci.* **2024**, *15*, 1417682. [CrossRef] [PubMed]
172. Mehta, S.S.; Ton, C.; Asundi, S.; Burks, T.F. Multiple camera fruit localization using a particle filter. *Comput. Electron. Agric.* **2017**, *142*, 139–154. [CrossRef]
173. Zhai, Y.; Zhang, L.; Hu, X.; Yang, F.; Huang, Y. A dynamic kalman filtering method for multi-object fruit tracking and counting in complex orchards. *Sensors* **2025**, *25*, 4138. [CrossRef] [PubMed]
174. Ge, Y.; Xiong, Y.; From, P.J. Three-dimensional location methods for the vision system of strawberry-harvesting robots: Development and comparison. *Precis. Agric.* **2023**, *24*, 764–782. [CrossRef]
175. Neupane, C.; Koirala, A.; Wang, Z.; Walsh, K.B. Evaluation of depth cameras for use in fruit localization and sizing: Finding a successor to Kinect v2. *Agronomy* **2021**, *11*, 1780. [CrossRef]
176. Gao, C.; Jiang, H.; Liu, X.; Li, H.; Wu, Z.; Sun, X.; He, L.; Mao, W.; Majeed, Y.; Li, R.; et al. Improved binocular localization of kiwifruit in orchard based on fruit and calyx detection using YOLOv5x for robotic picking. *Comput. Electron. Agric.* **2024**, *217*, 108621. [CrossRef]
177. Hu, S.; Sun, M.; Zhao, C.; Xiong, Y. Calibration-enhanced multi-view RGB-D vision for robust recognition and 3D localization of strawberries under occlusions. *Comput. Electron. Agric.* **2026**, *241*, 111221. [CrossRef]
178. Jing, X.; Jiang, H.; Niu, S.; Zhang, H.; Gilbert Murengami, B.; Wu, Z.; Li, R.; Zhou, C.; Ye, H.; Chen, J.; et al. End-to-end stereo matching network with two-stage partition filtering for full-resolution depth estimation and precise localization of kiwifruit for robotic harvesting. *Comput. Electron. Agric.* **2024**, *225*, 109333. [CrossRef]
179. Shen, Q.; Xu, D.; Guo, T.; Mao, X.; Xia, F. Two–stage multimodal 3D point localization framework for automatic grape harvesting. *Smart Agric. Technol.* **2025**, *12*, 101062. [CrossRef]
180. Zhang, K.; Chu, P.; Lammers, K.; Li, Z.; Lu, R. Active laser-camera scanning for high-precision fruit localization in robotic harvesting: System design and calibration. *Horticulturae* **2024**, *10*, 40. [CrossRef]
181. Kootstra, G.; Wang, X.; Blok, P.M.; Hemming, J.; van Henten, E. Selective harvesting robotics: Current research, trends, and future directions. *Curr. Robot. Rep.* **2021**, *2*, 95–104. [CrossRef]
182. Bechar, A.; Vigneault, C. Agricultural robots for field operations: Concepts and components. *Biosyst. Eng.* **2016**, *149*, 94–111. [CrossRef]
183. Kang, H.; Zhou, H.; Wang, X.; Chen, C. Real-time fruit recognition and grasping estimation for robotic apple harvesting. *Sensors* **2020**, *20*, 5670. [CrossRef] [PubMed]
184. Bhattacharya, S.; Arunachalam, C.; Zhang, K.; Li, J.; Lu, R.; Li, Z. Detect, segment, cluster: Apple localization for robotic harvesting in complex orchards. *Smart Agric. Technol.* **2025**, *12*, 101642. [CrossRef]
185. Zhu, K.; Li, J.; Zhang, K.; Arunachalam, C.; Bhattacharya, S.; Lu, R.; Li, Z. Foundation model-based apple ripeness and size estimation for selective harvesting. *Comput. Electron. Agric.* **2025**, *236*, 110407. [CrossRef]
186. Xie, H.H.; Zhang, Z.J.; Zhang, K.L.; Yang, L.; Zhang, D.X.; Yu, Y. Research on the visual location method for strawberry picking points under complex conditions based on composite models. *J. Sci. Food. Agr.* **2024**, *104*, 8566–8579. [CrossRef] [PubMed]
187. Han, J.-W.; Cho, J.-H.; Kim, Y.-T. Design of fruit harvesting robot system based on a reachability and inverse reachability map. *AgriEngineering* **2025**, *7*, 417. [CrossRef]
188. Cho, Y.; Choi, D.; Ko, K.; Park, J.H. Position selection and collision-free path planning for fruit picking robots. *Appl. Sci.* **2025**, *15*, 4419. [CrossRef]
189. 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]
190. Droukas, L.; Doulgeri, Z.; Tsakiridis, N.L.; Triantafyllou, D.; Kleitsiotis, I.; Mariolis, I.; Giakoumis, D.; Tzovaras, D.; Kateris, D.; Bochtis, D. A Survey of Robotic Harvesting Systems and Enabling Technologies. *J. Intell. Robot. Syst.* **2023**, *107*, 21. [CrossRef] [PubMed]
191. Zhu, Y.; Vougioukas, S.G. Fast scheduling and trajectory planning for robotic fruit harvesters with multiple cartesian arms. *IEEE Robot. Autom. Lett.* **2026**, *11*, 1562–1569. [CrossRef]
192. Rong, J.; Wang, P.; Wang, T.; Hu, L.; Yuan, T. Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots. *Comput. Electron. Agric.* **2022**, *202*, 107430. [CrossRef]
193. Hou, G.; Chen, H.; Niu, R.; Li, T.; Ma, Y.; Zhang, Y. Research on multi-layer model attitude recognition and picking strategy of small tomato picking robot. *Comput. Electron. Agric.* **2025**, *232*, 110125. [CrossRef]
194. Zhang, Y.; Xie, H.; Zhang, D.; Yang, L.; Lu, Z.; Cui, T.; He, X.; Zhang, K. Design and experimental study of ridge-grown strawberry automatic harvesting robot. *Comput. Electron. Agric.* **2026**, *240*, 111112. [CrossRef]

195. He, Z.; Karkee, M.; Zhang, Q. Enhanced machine vision system for field-based detection of pickable strawberries: Integrating an advanced two-step deep learning model merging improved YOLOv8 and YOLOv5-cls. *Comput. Electron. Agric.* **2025**, *234*, 110173. [CrossRef]
196. Wang, K.; Yang, S.; Liu, X.; Song, J.; Li, Y.; Xie, F. Research on target recognition, localization, and picking sequence planning for tomato-picking robots in greenhouse environments. *Smart Agric. Technol.* **2025**, *12*, 101628. [CrossRef]
197. Ning, Z.; Luo, L.; Ding, X.; Dong, Z.; Yang, B.; Cai, J.; Chen, W.; Lu, Q. Recognition of sweet peppers and planning the robotic picking sequence in high-density orchards. *Comput. Electron. Agric.* **2022**, *196*, 106878. [CrossRef]
198. van Henten, E.J.; Hemming, J.; van Tuijl, B.A.J.; Kornet, J.G.; Meuleman, J.; Bontsema, J.; Van Os, E.A. An Autonomous Robot for Harvesting Cucumbers in Greenhouses. *Auton. Robot.* **2002**, *13*, 241–258. [CrossRef]
199. Bac, C.W.; van Henten, E.J.; Hemming, J.; Edan, Y. Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *J. Field Robot.* **2014**, *31*, 888–911. [CrossRef]
200. Silwal, A.; Davidson, J.R.; Karkee, M.; Mo, C.; Zhang, Q.; Lewis, K. Design, integration, and field evaluation of a robotic apple harvester. *J. Field Robot.* **2017**, *34*, 1140–1159. [CrossRef]
201. Ji, W.; Zhang, T.; Xu, B.; He, G. Apple recognition and picking sequence planning for harvesting robot in a complex environment. *J. Agric. Eng.* **2023**, *55*, 1549. [CrossRef]
202. Lehnert, C.; English, A.; McCool, C.; Tow, A.W.; Perez, T. Autonomous sweet pepper harvesting for protected cropping systems. *IEEE Robot. Autom. Lett.* **2017**, *2*, 872–879. [CrossRef]
203. Zeeshan, S.; Aized, T. Performance analysis of path planning algorithms for fruit harvesting robot. *J. Biosyst. Eng.* **2023**, *48*, 178–197. [CrossRef]
204. Kurtser, P.; Edan, Y. Planning the sequence of tasks for harvesting robots. *Robot. Auton. Syst.* **2020**, *131*, 103591. [CrossRef]
205. Chakraa, H.; Guérin, F.; Leclercq, E.; Lefebvre, D. Optimization techniques for multi-robot task allocation problems: Review on the state-of-the-art. *Robot. Auton. Syst.* **2023**, *168*, 104492. [CrossRef]
206. Rema, C.; Costa, P.; Silva, M.; Pires, E.J.S. Task scheduling with mobile robots—A systematic literature review. *Robotics* **2025**, *14*, 75. [CrossRef]
207. 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]
208. Yang, T.; Du, X.T.; Zhang, B.; Wang, X.; Zhang, Z.P.; Wu, C.D. Coverage path planning based on region segmentation and path orientation optimization. *Agriculture* **2025**, *15*, 1479. [CrossRef]
209. Zhang, L.; He, Z.; Zhu, H.; Wei, Z.; Lu, J.; He, X. Adaptive SOM-GA hybrid algorithm for grasping sequence optimization in apple harvesting robots: Enhancing efficiency in open-field orchards. *Agronomy* **2025**, *15*, 1230. [CrossRef]
210. Li, C.; Ma, W.; Liu, F.; Fang, B.; Lu, H.; Sun, Y. Recognition of citrus fruit and planning the robotic picking sequence in orchards. *Signal Image Video Process.* **2023**, *17*, 4425–4434. [CrossRef]
211. Sun, X.; Zhai, X. Electrical control system of harvesting robotic arm based on PLC and particle swarm Algorithm. *J. Electron. Syst.* **2024**, *14*, 117–123. [CrossRef]
212. Xiong, Z.; Hong, G. Optimizing agricultural harvesting with a PLC-based robotic arm. *J. Netw. Technol.* **2025**, *16*, 163–171.
213. Ma, W.; Yang, Z.; Qi, X.; Xu, Y.; Liu, D.; Tan, H.; Li, Y.; Yang, X. Study on the fragrant pear-picking sequences based on the multiple weighting method. *Agriculture* **2023**, *13*, 1923. [CrossRef]
214. Bwambale, E.; Wanyama, J.; Adongo, T.A.; Umukiza, E.; Ntole, R.; Chikavumbwa, S.R.; Sibale, D.; Jeremaih, Z. A review of model predictive control in precision agriculture. *Smart Agric. Technol.* **2025**, *10*, 100716. [CrossRef]
215. Zhao, J.; Fan, S.; Zhang, B.; Wang, A.; Zhang, L.; Zhu, Q. Research status and development trends of deep reinforcement learning in the intelligent transformation of agricultural machinery. *Agriculture* **2025**, *15*, 1223. [CrossRef]
216. Zhou, L.; Hu, A.; Cheng, Y.; Zhang, W.; Zhang, B.; Lu, X.; Wu, Q.; Ren, N. Barrier-free tomato fruit selection and location based on optimized semantic segmentation and obstacle perception algorithm. *Front. Plant Sci.* **2024**, *15*, 1460060. [CrossRef] [PubMed]
217. van de Ven, R.; Leylavi Shoushtari, A.; Nieuwenhuizen, A.; Kootstra, G.; van Henten, E.J. Using learning from demonstration (LfD) to perform the complete apple harvesting task. *Comput. Electron. Agric.* **2024**, *224*, 109195. [CrossRef]
218. van de Ven, R.; Nieuwenhuizen, A.; van Henten, E.J.; Kootstra, G. DualLQR: Efficient grasping of oscillating apples using task parameterized learning from demonstration. *arXiv* **2024**, arXiv:2409.16957. [CrossRef]
219. Li, L.; Kasaei, H. Enhanced View Planning for Robotic Harvesting: Tackling Occlusions with Imitation Learning. In *Proceedings of the IEEE International Conference on Robotics and Automation, Atlanta, GA, USA, 19–23 May 2025*; IEEE: New York, NY, USA, 2025; pp. 13146–13152. [CrossRef]
220. Kim, C.H.; Silwal, A.; Kantor, G. Autonomous robotic pepper harvesting: Imitation learning in unstructured agricultural environments. *IEEE Robot. Autom. Lett.* **2025**, *10*, 3406–3413. [CrossRef]
221. Chen, Y.; Yu, Z.; Han, Z.X.; Sun, W.H.; He, L. A decision-making system for cotton irrigation based on reinforcement learning strategy. *Agronomy* **2024**, *14*, 11. [CrossRef]

222. Yi, T.; Zhang, D.; Luo, L.; Wang, Y.; Liu, B. View planning for grape harvesting based on self-supervised deep reinforcement learning under occlusion. *Comput. Electron. Agric.* **2025**, *239*, 110913. [CrossRef]
223. Alaaudeen, K.M.; Selvarajan, S.; Manoharan, H.; Jhaveri, R.H. Intelligent robotics harvesting system process for fruits grasping prediction. *Sci. Rep.* **2024**, *14*, 2820. [CrossRef] [PubMed]
224. Li, X.; Ma, N.; Han, Y.; Yang, S.; Zheng, S. AHPPEBot: Autonomous robot for tomato harvesting based on phenotyping and pose estimation. In *Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024*; IEEE: New York, NY, USA, 2024; pp. 18150–18156. [CrossRef]
225. Fei, Z.; Lu, W.; Hou, L.; Peng, C. Learning to pick: A visuomotor policy for clustered strawberry picking. *arXiv* **2025**, arXiv:2509.14530. [CrossRef]
226. Porichis, A.; Inglezou, M.; Kegkeroglou, N.; Mohan, V.; Chatzakos, P. Imitation learning from a single demonstration leveraging vector quantization for robotic harvesting. *Robotics* **2024**, *13*, 98. [CrossRef]
227. Pagliarani, N.; Tzafestas, C.; Papadopoulos, E.; Maragos, P.; Mastrogeorgiou, A.; Porichis, A.; Grogan, H.; Ehrhardt, R.; Cianchetti, M. SoftGrip: Towards a soft robotic platform for automatized mushroom harvesting. In *European Robotics Forum 2024*; Springer Nature: Cham, Switzerland, 2024; pp. 48–52. [CrossRef]
228. Mahmoudi, S.; Davar, A.; Sohrabipour, P.; Bist, R.B.; Tao, Y.; Wang, D. Leveraging imitation learning in agricultural robotics: A comprehensive survey and comparative analysis. *Front. Robot. AI* **2024**, *11*, 1441312. [CrossRef] [PubMed]
229. Williams, E.; Polydoros, A. Zero-shot sim-to-real reinforcement learning for fruit harvesting. In *Proceedings of the 2025 IEEE 21st International Conference on Automation Science and Engineering (CASE), Los Angeles, CA, USA, 17–21 August 2025*; IEEE: New York, NY, USA, 2025; pp. 1423–1428. [CrossRef]
230. Zheng, A.; Fang, Z.; Li, Z.; Dong, H.; Li, K. DRFW-TQC: Reinforcement learning for robotic strawberry picking with dynamic regularization and feature weighting. *AgriEngineering* **2025**, *7*, 208. [CrossRef]
231. Patil, S.A.; Sukte, C.D.; Patil, J.R.; Mhaske, C.R.; Dakhorkar, M.; Kodmelwar, M.K. Enhancing IOT-based fruit picking with reinforcement learning, transfer learning and neuroevolution. In *ICT Systems and Sustainability*; Springer Nature: Cham, Switzerland, 2026; pp. 124–134. [CrossRef]
232. Zhang, Y.; Li, Y.; Feng, Q.; Sun, J.; Peng, C.; Gao, L.; Chen, L. Compliant motion planning integrating human skill for robotic arm collecting tomato bunch based on improved DDPG. *Plants* **2025**, *14*, 634. [CrossRef] [PubMed]
233. Xie, H.; Zhang, D.; Yang, L.; Cui, T.; He, X.; Zhang, K.; Zhang, Z. Development, integration, and field evaluation of a dual-arm ridge cultivation strawberry autonomous harvesting robot. *J. Field Robot.* **2025**, *42*, 1783–1798. [CrossRef]
234. Jiang, Y.; Liu, J.; Wang, J.; Li, W.; Peng, Y.; Shan, H. Development of a dual-arm rapid grape-harvesting robot for horizontal trellis cultivation. *Front. Plant Sci.* **2022**, *13*, 881904. [CrossRef] [PubMed]
235. Lammers, K.; Zhang, K.; Zhu, K.; Chu, P.; Li, Z.; Lu, R. Development and evaluation of a dual-arm robotic apple harvesting system. *Comput. Electron. Agric.* **2024**, *227*, 109586. [CrossRef]
236. Yan, B.; Li, X. An intelligent operation area allocation and automatic sequential grasping algorithm for dual-arm horticultural smart harvesting robot. *Horticulturae* **2025**, *11*, 740. [CrossRef]
237. Chen, B.; Gong, L.; Xie, S.; Zhao, X.; Gao, P.; Luo, H.; Luo, C.; Li, Y.; Liu, C. Dual-arm coordination of a tomato harvesting robot with subtask decoupling and synthesizing. *Agriculture* **2026**, *16*, 267. [CrossRef]
238. Sarabu, H.; Ahlin, K.; Hu, A.P. Graph-based cooperative robot path planning in agricultural environments. In Proceedings of the 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Hong Kong, China, 8–12 July 2019; pp. 519–525. [CrossRef]
239. Yin, C.; Huang, J.; Xia, Y.; Zheng, H.; Fu, W.; Zhang, B. Design, development, integration and field evaluation of a dual robotic arm mango harvesting robot. *J. Field Robot.* **2025**, *42*, 3705–3725. [CrossRef]
240. Huang, W.; Miao, Z.; Wu, T.; Guo, Z.; Han, W.; Li, T. Design of and experiment with a dual-arm apple harvesting robot system. *Horticulturae* **2024**, *10*, 1268. [CrossRef]
241. Zhu, K.; Lammers, K.; Zhang, K.; Arunachalam, C.; Bhattacharya, S.; Li, J.; Lu, R.; Li, Z. Advancement and field evaluation of a dual-arm apple harvesting robot. *Smart Agric. Technol.* **2025**, *12*, 101343. [CrossRef]
242. Li, T.; Xie, F.; Zhao, Z.; Zhao, H.; Guo, X.; Feng, Q. A multi-arm robot system for efficient apple harvesting: Perception, task plan and control. *Comput. Electron. Agric.* **2023**, *211*, 107979. [CrossRef]
243. Yoshida, T.; Onishi, Y.; Kawahara, T.; Fukao, T. Automated harvesting by a dual-arm fruit harvesting robot. *ROBOMECH J.* **2022**, *9*, 19. [CrossRef]
244. Arikapudi, R.; Vougioukas, S.G. Robotic tree-fruit harvesting with arrays of cartesian arms: A study of fruit pick cycle times. *Comput. Electron. Agric.* **2023**, *211*, 108023. [CrossRef]
245. Lei, X.; Liu, J.; Jiang, H.; Xu, B.; Jin, Y.; Gao, J. Design and testing of a four-arm multi-joint apple harvesting robot based on singularity analysis. *Agronomy* **2025**, *15*, 1446. [CrossRef]
246. Yan, B.; Li, X. Task-parceling and synchronous retrieval scheme for twin-arm orchard apple tree automaton. *Plants* **2025**, *14*, 2798. [CrossRef] [PubMed]

247. Patra, K.; Sinha, A.; Guha, A. Online capability based task allocation of cooperative manipulators. *J. Intell. Robot. Syst.* **2024**, *110*, 23. [CrossRef]
248. Niu, J.; Yu, Q.; Bi, M.; Zhao, J.; Zhang, T. Deep reinforcement learning-based cooperative harvesting strategy for dual-arm robots in apple picking. *Agronomy* **2025**, *15*, 2565. [CrossRef]
249. Guo, Z.; Fu, H.; Wu, J.; Han, W.; Huang, W.; Zheng, W.; Li, T. Dynamic task planning for multi-arm apple-harvesting robots using LSTM-PPO reinforcement learning algorithm. *Agriculture* **2025**, *15*, 588. [CrossRef]
250. Sepúlveda, D.; Fernández, R.; Navas, E.; Armada, M.; González-De-Santos, P. Robotic aubergine harvesting using dual-arm manipulation. *IEEE Access* **2020**, *8*, 121889–121904. [CrossRef]
251. Liu, Z.; Zhou, J.; Mucchiani, C.; Karydis, K. Vision-assisted avocado harvesting with aerial bimanual manipulation. *Adv. Robot. Res.* **2026**, *2*, e202500003. [CrossRef]
252. Liu, Z.; Zhou, J.; Karydis, K. Hierarchical tri-manual planning for vision-assisted fruit harvesting with quadrupedal robots. In *Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA), Atlanta, GA, USA, 19–23 May 2025*; IEEE: New York, NY, USA, 2025; pp. 7153–7159. [CrossRef]
253. Stavridis, S.; Droukas, L.; Doulgeri, Z. Bimanual grape manipulation for human-inspired robotic harvesting. *IEEE/ASME Trans. Mechatron.* **2024**, *30*, 2722–2732. [CrossRef]
254. Gursoy, E.; Navarro, B.; Cosgun, A.; Kulić, D.; Cherubini, A. Towards vision-based dual arm robotic fruit harvesting. In *Proceedings of the 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), Auckland, New Zealand, 26–30 August 2023*; IEEE: New York, NY, USA, 2023; pp. 1–6. [CrossRef]
255. Lei, M.; Wang, T.; Yao, C.; Liu, H.; Wang, Z.; Deng, Y. Real-time kinematics-based self-collision avoidance algorithm for dual-arm robots. *Appl. Sci.* **2020**, *10*, 5893. [CrossRef]
256. Wang, Y.; Wang, L. Whole-body collision avoidance control design using quadratic programming with strict and soft task priorities. *Robot. Comput.-Integr. Manuf.* **2020**, *62*, 101882. [CrossRef]
257. Luo, X.; Zhang, R.; Yang, S.; Sun, Z.; Zhang, R.; Wang, J. Reactive self-collision avoidance for dual-arm robots using a temporal feature modeling and fusion network. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 22625–22637. [CrossRef]
258. Liu, Y.; Yu, C.; Sheng, J.; Zhang, T. Self-collision avoidance trajectory planning and robust control of a dual-arm space robot. *Int. J. Control Autom. Syst.* **2018**, *16*, 2896–2905. [CrossRef]
259. Wu, Y.; Jia, X.; Li, T.; Liu, J. A real-time collision avoidance method for redundant dual-arm robots in an open operational environment. *Robot. Comput.-Integr. Manuf.* **2025**, *92*, 102894. [CrossRef]
260. Liu, J.; Xu, C.; Jia, X.; Wu, Y.; Li, T. Cooperative collision avoidance control with relative velocity information for redundant dual-arm robotic manipulators. *J. Bionic Eng.* **2025**, *22*, 1111–1125. [CrossRef]
261. Surya Prakash, S.K.; Prajapati, D.; Narula, B.; Shukla, A. IAPF: An improved artificial potential field framework for asymmetric dual-arm manipulation with real-time inter-arm collision avoidance. *Front. Robot. AI* **2025**, *12*, 1604506. [CrossRef] [PubMed]
262. Wang, J.; Xiong, G.; Dang, B.; Chen, J.; Zhang, J.; Xie, H. Improved RRT-based obstacle-avoidance path planning for dual-arm robots in complex environments. *Machines* **2025**, *13*, 621. [CrossRef]
263. Yang, D.; Dong, L.; Dai, J.K. Collision avoidance trajectory planning for a dual-robot system: Using a modified APF method. *Robotica* **2024**, *42*, 846–863. [CrossRef]
264. Wei, L.; Wang, Y.; Hu, Y.; Lam, T.L.; Wei, Y. Online dual robot–human collaboration trajectory generation by convex optimization. *Robot. Comput.-Integr. Manuf.* **2025**, *91*, 102850. [CrossRef]
265. Gautam, S.; Li, M.; Roozbahani, D.; Alizadeh, M.; Handroos, H. Design and development of an intelligent collision avoidance platform for a teleoperated dual robotic arm. *J. Field Robot.* **2025**, *42*, 3047–3062. [CrossRef]
266. Wang, P.; Ghergherehchi, M.; Kim, J.; Zhang, M.; Song, J. Transformer-based path planning for single-arm and dual-arm robots in dynamic environments. *Int. J. Adv. Manuf. Technol.* **2025**, *139*, 3801–3819. [CrossRef]
267. Ying, K.-C.; Pourhejazy, P.; Cheng, C.-Y.; Cai, Z.-Y. Deep learning-based optimization for motion planning of dual-arm assembly robots. *Comput. Ind. Eng.* **2021**, *160*, 107603. [CrossRef]
268. Dong, Z.; Zhong, B.; He, J.; Gao, Z. Dual-arm obstacle avoidance motion planning based on improved RRT algorithm. *Machines* **2024**, *12*, 472. [CrossRef]
269. Luo, Y.; Zhang, M.; Liu, Y.; Lin, J.; Zhang, Z. Dynamic neural learning for obstacle avoidance of humanoid robot performing cooperative tasks. *Neurocomputing* **2025**, *633*, 129727. [CrossRef]
270. Ju, C.; Kim, J.; Seol, J.; Son, H.I. A review on multirobot systems in agriculture. *Comput. Electron. Agric.* **2022**, *202*, 107336. [CrossRef]
271. Mao, W.; Liu, Z.; Liu, H.; Yang, F.; Wang, M. Research progress on synergistic technologies of agricultural multi-robots. *Appl. Sci.* **2021**, *11*, 1448. [CrossRef]
272. Lytridis, C.; Bazinas, C.; Kalathas, I.; Siavalas, G.; Tsakmakis, C.; Spirantis, T.; Badeka, E.; Pachidis, T.; Kaburlasos, V.G. Cooperative grape harvesting using heterogeneous autonomous robots. *Robotics* **2023**, *12*, 147. [CrossRef]

273. Ma, N.; Cao, S.; Bai, T.; Kong, F.; Sun, W. Research progress and prospect of multi-robot collaborative SLAM in complex agricultural scenarios. *Smart Agric.* **2024**, *6*, 23–43. [CrossRef]
274. Schmuck, P.; Ziegler, T.; Karrer, M.; Perraudin, J.; Chli, M. COVINS: Visual-inertial SLAM for centralized collaboration. In *Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Bari, Italy, 4–8 October 2021*; IEEE: New York, NY, USA, 2021; pp. 171–176. [CrossRef]
275. Tian, Y.; Chang, Y.; Arias, F.H.; Nieto-Granda, C.; How, J.P.; Carlone, L. Kimera-multi: Robust, distributed, dense metric-semantic SLAM for multi-robot systems. *IEEE Trans. Robot.* **2022**, *38*, 2022–2038. [CrossRef]
276. Yue, Y.; Yang, C.; Wang, Y.; Zhang, J.; Wen, M.; Tang, X.; Zhang, H.; Wang, D. Multi-robot map fusion framework using heterogeneous sensors. In *Proceedings of the 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), Bangkok, Thailand, 18–20 November 2019*; IEEE: New York, NY, USA, 2019; pp. 536–541. [CrossRef]
277. Tagarakis, A.C.; Filippou, E.; Kalaitzidis, D.; Benos, L.; Busato, P.; Bochtis, D. Proposing UGV and UAV systems for 3D mapping of orchard environments. *Sensors* **2022**, *22*, 1571. [CrossRef] [PubMed]
278. Lee, H.; Bae, J.; Patil, A.; Park, M.; Nguyen, V. Heuristic approaches for coordinating collaborative heterogeneous robotic systems in harvesting automation with size constraints. *Sensors* **2025**, *25*, 6443. [CrossRef] [PubMed]
279. Guo, H.; Miao, Z.; Ji, J.C.; Pan, Q. An effective collaboration evolutionary algorithm for multi-robot task allocation and scheduling in a smart farm. *Knowl.-Based Syst.* **2024**, *289*, 111474. [CrossRef]
280. Lu, Z.; Zhao, Z.; Long, L.; Ma, Y.; LeiLi, L.; Liu, Z.; Dai, F.; Zhang, Y.; Li, J. Multi-robot task allocation in agriculture scenarios based on the improved NSGA-II algorithm. In *Proceedings of the 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), Hong Kong, China, 10–13 October 2023*; IEEE: New York, NY, USA, 2023; pp. 1–6. [CrossRef]
281. Chen, P.; Liang, J.; Song, H.; Qiao, K.J.; Yue, C.T.; Yu, K.J.; Suganthan, P.N.; Pedrycz, W. Multi-objective task allocation for electric harvesting robots: A hierarchical route reconstruction approach. *IEEE Trans. Cybern.* **2025**, *56*, 1583–1596. [CrossRef] [PubMed]
282. Wang, Y.; Wang, Y.; Cao, Y.; Sartoretti, G. Spatio-temporal attention network for persistent monitoring of multiple mobile targets. In *Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023*; IEEE: New York, NY, USA, 2023; pp. 3903–3910. [CrossRef]
283. Hu, K.; Ying, W.; Pan, Y.; Kang, H.; Chen, C. High-fidelity 3D reconstruction of plants using neural radiance fields. *Comput. Electron. Agric.* **2024**, *220*, 108848. [CrossRef]
284. Olivastri, E.; Francis, J.; Pretto, A.; Sünderhauf, N.; Rana, K. Multi-modal 3D scene graph updater for shared and dynamic environments. *arXiv* **2024**, arXiv:2411.02938. [CrossRef]
285. Gu, Q.; Kuwajerwala, A.; Morin, S.; Jatavallabhula, K.M.; Sen, B.; Agarwal, A.; Rivera, C.; Paul, W.; Ellis, K.; Chellappa, R.; et al. Conceptgraphs: Open-vocabulary 3d scene graphs for perception and planning. In *Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024*; IEEE: New York, NY, USA, 2024; pp. 5021–5028. [CrossRef]
286. Zhang, Y.; Yao, J.; Qian, C. Event-triggered nonlinear visual predictive control strategy for robots. *J. Intell. Robot. Syst.* **2025**, *111*, 80. [CrossRef]
287. Bassiouny, A.; Elsayed, A.H.; Falomir, Z.; del Pobil, A.P. UJI-Butler: A Symbolic/non-symbolic robotic system that learns through multi-modal interaction. *Int. J. Soc. Robot.* **2025**, *17*, 2883–2903. [CrossRef]
288. Zhang, M.; Abdi, M.; Dasari, V.R.; Restuccia, F. Semantic edge computing and semantic communications in 6G networks: A unifying survey and research challenges. *Comput. Netw.* **2025**, *270*, 111531. [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.