Research on the Strawberry Picking Robot with Improved YOLO V11 and Adaptive Path Planning

Dong Yu*
College of Electronic Information, Southwest Minzu University, Chengdu, China
**Corresponding Author.*

Abstract: To address the challenges of target recognition difficulty, complex path planning, and low operational efficiency in strawberry picking, this paper designs and implements an intelligent strawberry picking robot that integrates YOLO V11 object detection and adaptive path planning technologies. The robot system consists mainly of a visual recognition module, motion control module, robotic arm picking module, and LiDAR module. The visual system is based on the YOLO V11 deep learning model, which incorporates the CA attention mechanism to achieve high-precision recognition and 3D localization of strawberry fruits. By integrating a depth camera with sensor fusion algorithms, the system effectively identifies strawberries under various stages of ripeness and occlusion. The navigation system adopts an adaptive path planning strategy based on ROS, combining global path planning with local obstacle avoidance algorithms to improve the robot's mobility efficiency and stability in the complex path conditions of greenhouse environments. The robotic arm and end-effector use coordinate transformation and posture planning to achieve flexible strawberry picking. This research provides a feasible solution for the automation and intelligence of strawberry picking operations.

Keywords: Strawberry Picking Robot; YOLO V11; Robot Operating System; Object Detection

1. Introduction

In recent years, with the rapid advancement of artificial intelligence, automatic control, and robotics, agricultural intelligence has become an important means of enhancing productivity and reducing labor costs in agricultural production [1]. In particular, within the domain of fruit and vegetable harvesting, where labor shortages and high labor intensity are common challenges, the development and application of intelligent picking robots have attracted increasing attention [2]. Among these, strawberries pose especially stringent requirements for robotic systems due to their delicate texture, dense distribution, frequent occlusion, and non-uniform ripening [3]. Traditional manual strawberry picking is not only inefficient and costly but also fails to meet the modern agricultural demand for high-efficiency and precisely controlled operations. Therefore, the development of an intelligent strawberry picking robot that integrates high-precision perception, adaptive path planning, and flexible end-effector operation holds substantial practical significance and application value [4].

At present, considerable progress has been made both domestically and internationally in the research of fruit and vegetable picking robots. Early studies abroad, particularly in countries such as Japan and the United States, have resulted in the development of robotic systems for harvesting oranges, cucumbers, pineapples, and other crops. These systems have achieved notable success in aspects such as object detection accuracy and motion coordination. For example, robotic arms equipped with visual sensors have been employed in greenhouse environments to identify and harvest fruit [5,6]. However, many of these systems rely on fixed structures or track-based mobility, making them less adaptable to the irregular planting layouts and dynamic obstacles commonly found in real agricultural settings.

In China, related research has primarily focused on the automated picking of tomatoes, apples, and citrus fruits. Some of these studies have demonstrated promising results in fruit detection using deep learning and robotic arm control. For instance, a tomato picking robot developed by Xie Xiaoxuan and colleagues from the University of Science and Technology of China adopts a ROS-based modular control architecture [7], achieving automated operation through coordinated image recognition, path planning, and execution control. Similarly, scholars such as Wang Li from Guangdong Polytechnic Institute have introduced dual-arm structures [8] and multi-sensor fusion strategies to improve picking efficiency and recognition accuracy. Nevertheless, most existing systems still suffer from limited detection robustness, weak adaptability to dynamic paths, and low operational success rates. In particular, a universally reliable solution for the complex task of strawberry harvesting has yet to be established.

In recent years, real-time object detection networks—especially the YOLO (You Only Look Once) series—have demonstrated outstanding performance in agricultural visual recognition tasks. YOLO V11, the latest iteration of the series [9], integrates attention mechanisms within its architecture [10], significantly enhancing its ability to detect small objects and maintain robustness under occlusion conditions [11,12]. Meanwhile, traditional path planning algorithms such as Dijkstra and A*, although effective in generating feasible paths, still exhibit limitations in dynamic obstacle avoidance and local path optimization. Integrating visual perception with real-time environmental feedback for adaptive path adjustment has become a key strategy to improve the autonomy and efficiency of robotic operations.

Based on the above, this study proposes the development of an intelligent strawberry picking robot that integrates YOLO V11-based visual recognition with adaptive path planning. The robot system incorporates advanced object detection algorithms, a ROS-based task scheduling and path control framework, and a flexible end-effector design, enabling accurate fruit localization, dynamic path adjustment, and low-damage picking under complex environmental conditions.

2. Structural Composition and Working Principle of the Strawberry Picking Robot

Strawberry is a typical low-growing crop, with plant heights generally ranging from 15 to 30 cm. Its fruits are scattered and grow close to the ground. Upon ripening, the fruit color gradually changes from pale green to bright red. However, in natural environments, the fruits are often occluded by stems and leaves. Additionally, strawberries have short pedicels, dense spatial distribution, varying sizes, and are highly susceptible to mechanical damage.

2.1 Structural Composition of the Strawberry Picking Robot

The growth characteristics of strawberries necessitate that the design of a strawberry picking robot possess low-position operational capability to accommodate fruits distributed close to the ground; high-precision visual recognition and localization capabilities to detect fruits that are partially occluded by foliage or presented in varying orientations; and a flexible manipulation mechanism to minimize fruit damage during the picking process. Therefore, the structural design of the picking robot must fully consider spatial parameters such as row spacing, plant height, and fruit hanging positions to ensure adequate maneuverability and adaptability of the system during operation. The strawberry picking robot designed in this study is shown in Figure 1.

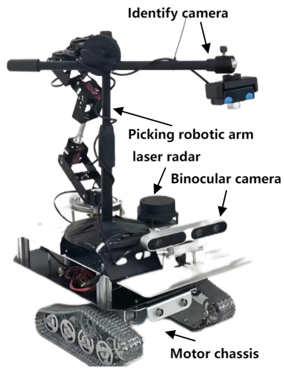

Figure 1. Physical Diagram of the Strawberry Picking Robot

To meet the increasing demand for integrated functions such as "non-destructive grasping and stable post-harvest placement" in practical applications of strawberry picking robots, this study introduces a structural innovation in the design of the end-effector. The proposed end-effector adopts a flexible, adaptive three-finger gripper, model FAE20886, which exhibits excellent compliance and grasp adaptability. This design effectively reduces the risk of compression and mechanical damage to strawberry fruits during the picking process. As shown in Figure 2, the gripper primarily consists of silicone fingers, a linkage mechanism, a push-rod drive assembly, a clamping device, and a servo actuation unit.

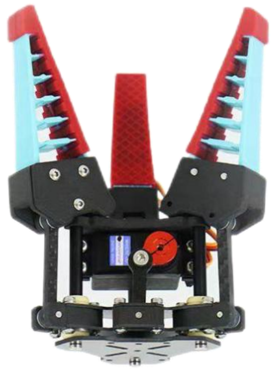

Figure 2. Flexible Adaptive Three-Finger Gripper

The three-finger structure can adaptively adjust its gripping angle and contact surface according to the shape of the fruit, enabling flexible grasping of strawberries with varying sizes and orientations when combined with a force control strategy. The introduction of this end-effector significantly enhances the system's flexibility in controlling surface pressure on the fruit, effectively addressing issues commonly associated with traditional rigid grippers—such as surface indentation and peduncle tearing when handling delicate fruits like strawberries. This, in turn, ensures better post-harvest fruit integrity and preserves the commercial value of the produce.

2.2 Working Principle of the Robot

The strawberry picking robot developed in this study features a highly integrated mechanical structure, offering strong field maneuverability and operational stability. The control system is functionally divided into three major modules: the mobile chassis system, the sensing and recognition system, and the robotic arm execution system. Together, these modules form a complete intelligent picking workflow through coordinated operation.

The mobile chassis system adopts a dual-track drive structure, providing excellent obstacle-crossing capability and ground adhesion, making it suitable for uneven terrain in protected agricultural environments. Equipped with an embedded encoder and an inertial measurement unit (IMU), the system achieves real-time posture estimation and trajectory control. Utilizing the Robot Operating System (ROS) navigation framework, the chassis performs autonomous path planning, localization, and obstacle avoidance, ensuring stable navigation to designated operational areas.

The sensor fusion and recognition system integrates a binocular RGB camera, laser rangefinders, and depth perception units. By fusing image and depth information, the system enables 3D environmental modeling, localization, and real-time detection of target strawberry fruits. During operation, the YOLO V11 model is deployed on a Jetson Nano platform to perform front-end object detection. The relative position of the fruit is calculated through depth data analysis.

The robotic arm execution system consists of a multi-degree-of-freedom robotic arm and a flexible end-effector gripper. The robotic arm achieves high-precision spatial positioning through motor-driven actuation and feedback control. An integrated lifting mechanism and rotating platform extend the arm's picking coverage area. The end-effector is responsible for gripping, picking, and placing the strawberry fruit. The system employs an eye-in-hand configuration, with the lifting guide rail, rotating components, and visual sensing unit mounted on the robotic arm, allowing flexible adjustment of operational postures in 3D space.

Once the picking operation begins, the robot autonomously navigates to the predefined working area using its path planning module. Upon arrival, the sensing and recognition system is activated to capture images and perform intelligent crop target identification. If a target is successfully identified, the system calculates its coordinates and transmits them to the robotic arm control module. The robotic arm then performs path planning and inverse kinematics computations to drive the end-effector for fruit picking. If detection fails or the target is beyond the reachable range, the system commands the chassis to make fine position adjustments or adjusts the vertical height via the lifting platform, initiating a new round of detection until the picking task is completed.

After each picking cycle, the system automatically proceeds to the next recognition and execution task, forming a continuous and efficient closed-loop workflow. This system architecture not only enhances the level of automation in strawberry harvesting but also significantly improves adaptability to complex cultivation environments. Figure 3 illustrates the overall workflow of the robot.

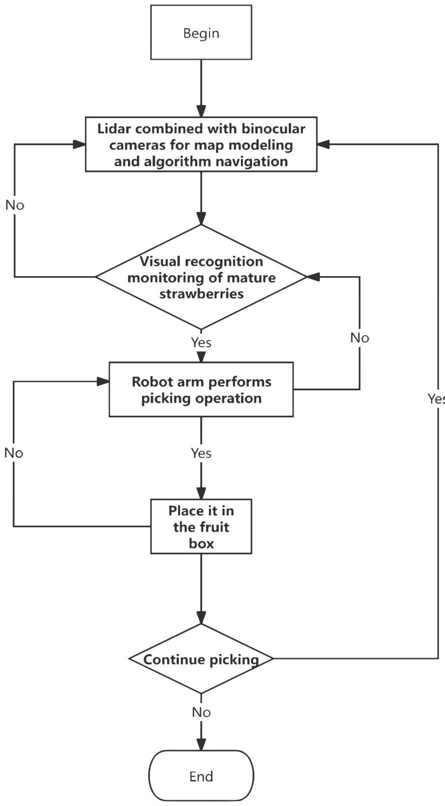

Figure 3. Robot Workflow Diagram

3. Design and Implementation of Autonomous Navigation

During the operation of the strawberry picking robot, the performance of the navigation system directly determines its mobility and operational efficiency in complex cultivation environments. To enable efficient autonomous movement within semi-structured environments such as greenhouses or plant factories and to accurately reach designated picking areas, this study designs an adaptive path planning navigation system that integrates visual perception, LiDAR, inertial measurement units (IMU), and deep learning-assisted localization. This system dynamically adjusts the travel path in response to environmental feedback, thereby improving navigation accuracy and obstacle avoidance capabilities.

The system is modularly integrated based on the Robot Operating System (ROS) platform, and a navigation subsystem with adaptive path adjustment capabilities is developed for the strawberry picking robot. By seamlessly combining multi-source environmental perception, map construction, path planning, and motion control, the navigation subsystem achieves efficient autonomous navigation in complex agricultural environments. It primarily relies on a fusion of sensors—including LiDAR, IMU, wheel odometry, and an RGB-D camera—to build an environmental perception module that continuously collects spatial structure and obstacle distribution data in real time.

The system employs Google's open-source Cartographer algorithm to construct a high-resolution 2D occupancy grid map and utilizes the Adaptive Monte Carlo Localization (AMCL) algorithm to achieve precise, real-time localization of the robot within the map. For path planning, a hybrid strategy combining global and local planning is adopted. The Dijkstra algorithm is used to generate optimal paths on the global map, while the Dynamic Window Approach (DWA) algorithm is applied to perform local path adjustments and real-time obstacle avoidance.

Motion control is managed by an STM32-based embedded controller, which receives velocity commands from the ROS planning module and dynamically adjusts wheel speeds to ensure accurate path tracking and pose correction. Under the coordinated operation of sensor fusion and a closed-loop control mechanism, the system demonstrates strong environmental adaptability and path self-adjustment capabilities. It effectively addresses challenges such as dense obstacles and dynamic path changes in plant factories or greenhouse environments, thereby ensuring the continuity and efficiency of the picking operation.

To ensure precise navigation within planting areas, the first step involves constructing a high-quality 2D occupancy grid map. In this study, the open-source Cartographer algorithm is employed for map construction, and the process is illustrated in Figure 4.

During the initial operation, the robot utilizes LiDAR to scan the surrounding environment, acquiring point cloud data and generating an initial sub-map. Subsequently, a point cloud matching algorithm is employed to estimate the robot's relative pose, which is continuously integrated with previously generated sub-maps, thereby completing the mapping of the entire operational area.

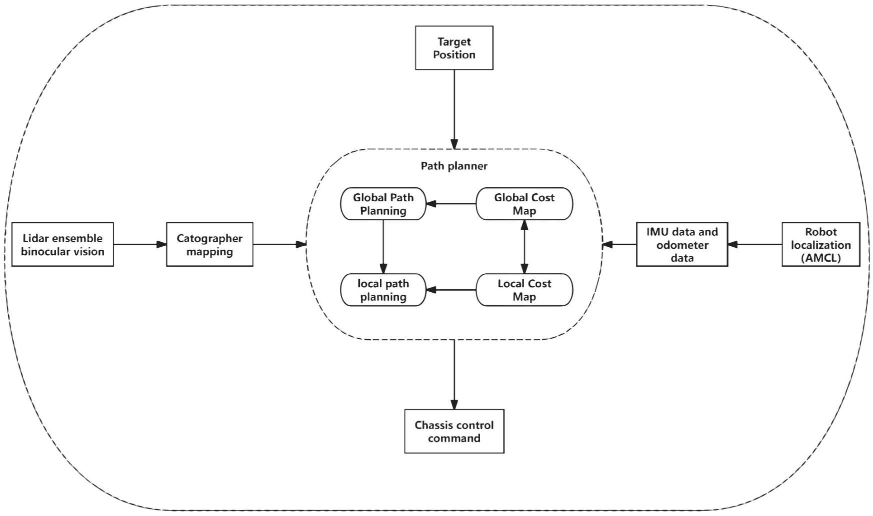

Figure 4. Framework of the Robot Modeling and Navigation System

Given the relatively structured environment and sparse visual features within plant factory cultivation zones, LiDAR-based mapping may suffer from point cloud mismatches, leading to mapping errors. To enhance map accuracy, the system incorporates an Inertial Measurement Unit (IMU) and wheel odometry as auxiliary sensors. These sensors provide additional constraints on the pose estimation process within the Cartographer algorithm, effectively mitigating drift.

In experimental validation, the robot maintained a constant speed of 0.3 m/s and successfully completed high-precision mapping along one aisle of the plant factory. As shown in Figure 5, the resulting map clearly depicts the layout of the planting racks and the width of the passageways.

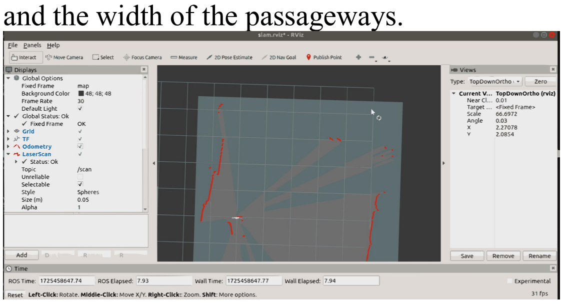

Figure 5. Mapping and Navigation Diagram

After completing the construction of the environmental map, the system utilizes the generated 2D occupancy grid map as the Global Costmap for navigation path planning and optimization. During actual operations, the robot first determines its global initial pose within the map using the Adaptive Monte Carlo Localization (AMCL) algorithm. This algorithm, based on particle filtering, integrates data from LiDAR, IMU, and the pre-built map to achieve robust pose estimation, enabling the robot to accurately localize itself even under conditions of high uncertainty.

Once localization is complete, the system assigns a target coordinate and activates the path planning module, which is managed and scheduled by the move_base package within the Robot Operating System (ROS). For path planning, a combined global and local planning strategy is employed. The global planner uses the Dijkstra algorithm to compute the optimal path from the current position to the target based on the Global Costmap. The Dijkstra algorithm offers strong path-search capabilities in complex grid topologies, effectively avoiding static obstacles and ensuring global reachability.

To address dynamic environments or unexpected obstacles, the system employs the Dynamic Window Approach (DWA) as the local path planner. DWA generates a series of feasible trajectories in the velocity space by considering the robot's current motion state, target direction, and nearby obstacle information. A cost function is then used to evaluate these trajectories, selecting the optimal one and outputting the corresponding linear and angular velocity commands. During this process, the robot continuously updates the environmental model using real-time sensor data and dynamically adjusts the Local Costmap according to predefined inflation parameters, thereby enhancing obstacle perception redundancy and improving the responsiveness of obstacle avoidance, ensuring dynamic path adjustment and stable execution.

Traditional path planning systems that rely solely on LiDAR and IMU data may exhibit delayed responses and suboptimal path selection when confronted with complex agricultural scenarios such as occluded strawberries or tilted plant growth within passageways. To address this limitation, this study integrates a front-mounted RGB-D vision module into the path planning framework, enabling visual perception to inform adaptive navigation decisions. As the robot approaches the target area, the system activates the YOLO V11 model to analyze the number and spatial density of ripe strawberries ahead. If a high concentration of targets is detected, the robot decelerates and adjusts its navigation direction toward the cluster. Using the forward-facing camera, the system assesses plant tilt and passage occlusion. If the occlusion area exceeds a defined threshold, the robot autonomously modifies its posture or selects an alternative path. Visual perception results are projected onto the Local Costmap, marking obstacle regions and triggering recalculation of the local path, thereby equipping the navigation strategy with predictive capabilities and decision-making functionality.

Finally, the target velocity commands generated by the path planning module are transmitted via serial communication in ROS to the STM32 embedded controller. Based on a differential drive model, the controller computes the actual speeds for the left and right wheels and employs a PID control algorithm to adjust wheel velocities, enabling the robot to smoothly follow the planned trajectory and achieve autonomous navigation.

4. Visual Recognition Design

In the strawberry-picking robot system, the accuracy of the visual perception module and the precision of target localization are critical to the overall operational efficiency and system stability. The strawberry cultivation environment is highly complex, with densely distributed fruits that vary in ripeness, are often partially occluded, and may grow in overlapping clusters. Traditional image recognition methods based solely on color or shape features are insufficient to meet the demands of such dynamic field conditions. To address this challenge, this study proposes a visual recognition system based on the YOLO V11 deep learning model, enhanced with a Coordinate Attention (CA) mechanism. By integrating a depth camera and multi-sensor fusion technology, the system achieves high-precision detection and three-dimensional spatial localization of strawberry fruits, thereby providing reliable data support for subsequent path planning and end-effector manipulation.

YOLO (You Only Look Once) represents a class of efficient one-stage object detection networks known for their rapid detection speed, lightweight architecture, and suitability for embedded deployment. Building upon the YOLO V11 model, this study introduces further enhancements by incorporating the Coordinate Attention (CA) mechanism. Unlike traditional channel attention mechanisms, the CA module incorporates spatial positional information into the computation of channel weights, thereby improving the model's ability to represent target regions in complex backgrounds.

Structurally, YOLO V11 employs multi-scale feature fusion and an improved neck module, which facilitates the capture of image features of strawberry fruits across varying scales and orientations. The integration of the CA module significantly enhances the model’s capacity to perceive occlusions and densely clustered targets. The updated architecture incorporating the CA attention mechanism is illustrated in Figure 6.

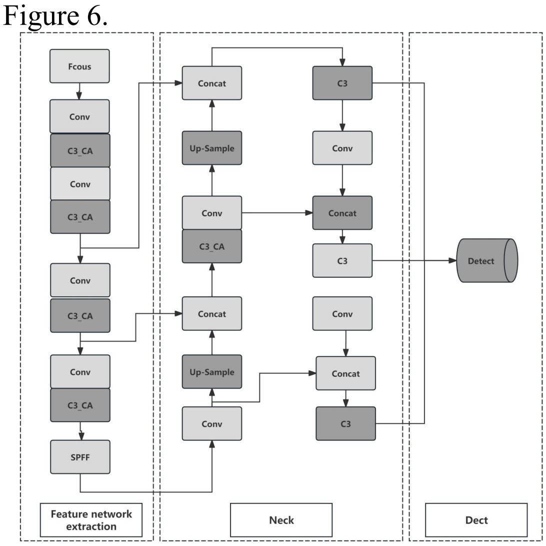

Figure 6. Algorithm Framework Incorporating the Coordinate Attention (CA) Mechanism

To further enhance the intelligence of picking decision-making, the system extends the output structure of the YOLO model to enable automatic assessment of strawberry fruit ripeness. In the annotated dataset, ripeness level labels are introduced (green/unripe, partially ripe with red and green, and fully red/ripe for picking). During training, the model learns the color, texture, and morphological features associated with different ripeness levels. In the inference phase, the model outputs not only bounding box coordinates and confidence scores but also the predicted ripeness category. This mechanism enables the robot to accurately identify harvestable targets even in scenarios involving overlapping fruits or interference from unripe specimens, thereby improving the precision and efficiency of the picking actions. Additionally, the system employs a joint decision mechanism based on confidence and ripeness thresholds to avoid false positives and mispicks.

Relying solely on 2D image recognition cannot satisfy the high-precision spatial requirements of the end-effector. To enable 3D localization, the system is equipped with a depth camera (Intel RealSense D435i) in addition to the RGB camera, allowing for real-time acquisition of RGB images alongside corresponding depth data. Through coordinate mapping, the 2D bounding boxes identified by the YOLO model are projected onto the depth map to obtain the spatial depth of the target's center point. To further improve the accuracy of 3D localization, the system integrates data from the IMU and wheel odometry, employing an Extended Kalman Filter (EKF) to achieve visual-inertial fusion localization. This enhances the robustness and temporal consistency of the depth data.

Moreover, the system incorporates a viewpoint selection mechanism: when a strawberry target is detected at an oblique angle or with blurred boundaries in the image, the system selects an optimal viewpoint for re-detection to improve localization accuracy. The final output of the 3D localization module is a spatial position vector (X, Y, Z) of the target within the robot's coordinate frame, which serves as the input for the robotic arm path planning module. Experimental results demonstrate that within a range of 1.2 meters, the localization error is controlled within ±1.5 cm, meeting the precision requirements for agricultural harvesting.

Given that strawberry fruits are often partially occluded by leaves or adjacent fruits during growth, the system incorporates an occlusion recognition and compensation strategy to improve the comprehensiveness and reliability of target detection. By introducing a visible area evaluation module into the YOLO output features, the system calculates a visibility score for each detection box. If the target's visibility is below a specified threshold but exhibits clear ripeness characteristics, it is marked as a "partially visible target" and enters the intelligent compensation process.

The compensation mechanism allows the robot to adjust its posture or reposition itself to observe the target from different angles, enabling multi-view fusion detection to reconstruct the complete contour. In parallel, based on the edge morphology of the target and the spatial distribution of nearby fruits, the system employs a lightweight point cloud completion network to reconstruct the spatial shape of the target and assist in preliminary pose estimation.

This mechanism significantly improves detection and localization of partially occluded fruits. Field tests show an occlusion recognition accuracy of 92.6%, effectively reducing target omission and maintaining harvesting efficiency. The recognition results are illustrated in Figure 7.

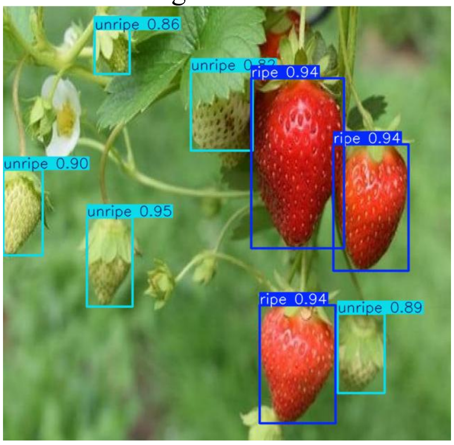

Figure 7. Strawberry Recognition Result Diagram

Table 1 presents the experimental data for visual recognition using the original YOLO V11 algorithm and the enhanced YOLO V11 algorithm integrated with the compensation mechanism and the Coordinate Attention (CA) module.

Table 1. Experimental Data Comparing the Original YOLO V11 Algorithm and the YOLO V11 Algorithm Enhanced with Coordinate Attention (CA) Mechanism and Compensation Strategy

Algorithm	Recall	Precision	mAP%
YOLO v11	0.894	0.859	88.5
YOLO v11+CA	0.937	0.948	92.6

As shown in Table 1, the introduction of the Coordinate Attention (CA) mechanism and the compensation strategy significantly improved the performance of the YOLO V11 algorithm. The recall increased from 0.894 to 0.937, the precision improved from 0.859 to 0.948, and the mean Average Precision (mAP) rose from 88.5% to 92.6%. These results indicate that the enhancements to the visual recognition system effectively improve the robustness of the picking process and enhance overall harvesting efficiency.

5. Conclusion and Outlook

This study addresses the challenges encountered in the strawberry-picking process within protected agriculture, such as complex target recognition, uncertain navigation paths, and the fragility of the fruit. An intelligent strawberry-picking robot was designed, integrating the YOLO V11 object detection algorithm with an adaptive path planning mechanism. The developed robot demonstrated strong performance in terms of fruit recognition accuracy, navigation stability, and flexible manipulation, thereby validating the effectiveness and practicality of the multi-module integrated design. This system offers valuable technical support for real-world agricultural harvesting applications.

Despite the promising experimental results, challenges such as significant morphological variability among strawberry plants and the highly dynamic nature of operational paths continue to pose uncertainties. Future research will focus on rapid reconstruction of the navigation system and the development of cooperative mechanisms to further enhance the system's generalization capability and practical value in agricultural environments.

References

[1] Mi, T. (2025). Application Research of Computer Technology in the Design, Manufacturing, and Intelligence of Agricultural Machinery. Southern Agricultural Machinery, 56(07), 188–191.

[2] Qiu, S. H. (2025). Application and Research of Intelligent Technology in Modern Agricultural Machinery. Agricultural Development and Equipment, (02), 30–32.

[3] Wang, D. Z., Wang, Y. X., Wang, X. C., et al. (2025). Management and Harvesting Technology in Agricultural Robots: Status, Challenges, and Future Trends. Journal of Intelligent Agricultural Equipment (Bilingual), 1–23.

[4] Wang, W., & Li, Y. (2024). Application Research of Intelligent Control Systems in Strawberry Picking Robots. Southern Agricultural Machinery, 55(19), 73–75.

[5] Huang, J. Y., Yin, C. H., Lou, K. R., et al. (2025). Design and Experiment of a Mango Picking Robot. Journal of Shihezi University (Natural Science Edition).

[6] Ma, W. M., Zhang, Y. A., & Chen, X. X. (2025). Design of an Intelligent Picking Robot Based on Raspberry Pi and Arduino Control. Instrument Technology, (01), 38–40+86.

[7] Xie, X. X., Kong, D. Y., Pan, B., et al. (2025). Design of a Tomato Picking Robot Control System for Plant Factories Based on ROS. Research on Agricultural Mechanization, 1–8. [Online] Available: 2025-04-11.

[8] Wang, L., Zheng, Z. P., Liu, Z. Q., et al. (2025). Research on a Four-arm Picking Robot Based on Multimodal Sensor Fusion Recognition. Agricultural Equipment and Vehicle Engineering, 63(03), 17–21.

[9] Wang, X. L., Zheng, X. L., & Lai, W. H. (2025). PCB Defect Detection and Instance Segmentation Algorithm Based on TAC-YOLOv11s. Journal of Hubei Minzu University (Natural Science Edition).

[10] Gao, Z. L. (2025). Research on Road Crack Detection System Based on Improved YOLOv11. Information and Computer, 37(06), 136–138.

[11] He, Z. X., Chen, L. L., Wang, X., et al. (2025). DMF-YOLOv11: An Object Detection Algorithm for UAV Aerial Images Based on Improved YOLOv11n. Computer Engineering and Applications, 1–14. [Online] Available: 2025-04-11.

[12]Li, X. W., Mao, X. K., Sun, B., et al. (2025). Small Object Detection Algorithm for UAV Aerial Images Based on YOLOv11n. Optical Communication Research, 1–8. [Online] Available: 2025-04-11.