Intelligent Robotic Arm Grasping System for the Agricultural Sector

Wencheng Yue*

School of Intelligent Science and Engineering, Yanshan University, Hebei, 066004, China

*Corresponding author: yuewencheng@stumail.ysu.edu.cn

Abstract. In order to promote the intelligent transformation of China's agricultural picking process and solve the development problem of insufficient mechanization level in the apple industry, this article takes the agricultural intelligent robotic arm grasping system as the research object and systematically elaborates on its key technical system. The article analyzes the application value of technology based on the actual needs of the industry, and at the same time, identifies the shortcomings of existing research, providing direction for optimizing intelligent picking equipment. In terms of core technology, the improved YOLOv7 algorithm has improved the accuracy of apple recognition, and binocular vision technology has achieved high-precision positioning. The four-finger underactuated robotic arm can adaptively grasp apples of different sizes, and the fusion system of model predictive control (MPC) and support vector classification (SVC) ensures lossless grasping efficiency. Through practical application verification, the success rate of system picking and the efficiency of single fruit operation have shown good performance. This article focuses on the collaborative analysis of technical modules, with a particular emphasis on the adaptation of low-rootstock dense planting orchard scenarios. It integrates technical indicators throughout the entire process to address research fragmentation issues, and also explores existing challenges and future development directions, such as environmental adaptability, providing a systematic reference for optimizing intelligent robotic arm grasping systems.

Keywords: Intelligent robotic arm technology, collaborative scene adaptation, picking equipment optimization.

1. Introduction

China is a major agricultural country, and the fruit and vegetable industry is an important pillar of the agricultural economy. In 2022, China's annual apple production was 45.973 million tons, accounting for 56.4% of the global total, with a planting area of 2 million hectares, firmly ranking first in the world [1]. However, the level of agricultural mechanization lags, with a comprehensive mechanization rate of only 25.88% in orchards nationwide and less than 3% in the picking process [1]. Traditional manual harvesting faces labor shortage and aging issues, with labor input and cost accounting for over 40%, and daily harvesting only 300kg [2].

Precise harvesting in complex and unstructured environments has become a bottleneck in agricultural modernization, and intelligent robotic arm grasping systems are the key to breaking through the situation. Based on this, this article focuses on perception recognition and grasping control technology, combined with the analysis of core technology innovation in low-rootstock dense planting orchards, and compensates for existing research deficiencies by strengthening technological collaboration, refining scene adaptation, and integrating full process indicators, providing a reference for the optimization of intelligent picking equipment.

2. Technical Core

2.1. Technical Roadmap

Fig. 1 shows the workflow of the agricultural intelligent robotic arm grasping system, which includes five closed-loop links: preliminary preparation, perception recognition, path planning, grasping execution, and subsequent processing. Preparation requires determining the work scenario and crops, debugging the parameters of the robotic arm, calibrating the robotic arm and manipulator, and loading algorithm programs. In the perceptual recognition process, the camera captures and preprocesses images, and after object detection algorithms (such as the improved YOLOv7) identify and screen fruits, the three-dimensional coordinates of the fruits are calculated. During path planning, obstacles are identified and modeled, and the optimal path for the robotic arm is generated through path planning algorithms such as Bi-RRT *. During the grasping execution phase, the robotic arm moves along the planned path, adjusts the posture of the robotic arm, and sensors collect data to control the grasping. The subsequent processing involves the robotic arm transferring the fruit to the collection device, recording the data, and uploading it to the cloud for analysis and optimization. It also forms a closed loop with the preliminary preparation stage, providing improvement direction for subsequent operations.

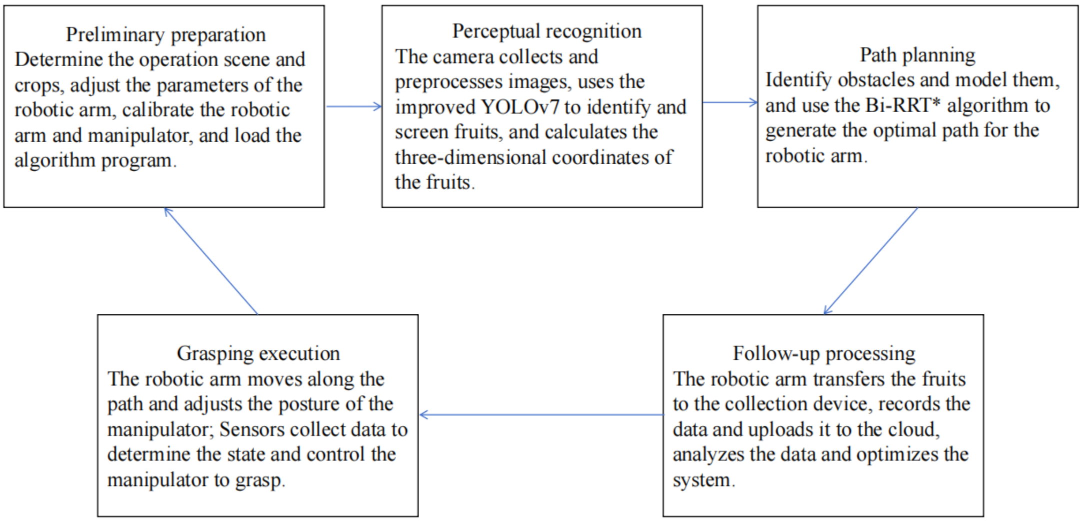

Fig. 1 Technology Roadmap (Photo/Picture credit: Original).

2.2. Perception and Recognition Technology - Object Detection and Localization in Complex Environments

The precise grasping of agricultural intelligent robotic arms begins with efficient perception and positioning of target fruits, and the unstructured environment of orchards (such as drastic changes in lighting, occlusion of branches and leaves, and overlapping fruits) is the core bottleneck. Traditional methods lack robustness, such as traditional digital image processing techniques, which have an apple recognition rate of only 18% under backlighting conditions and are difficult to handle the problem of dense overlapping fruits [3]. For example, early deep learning models such as YOLOv3 have limited ability to recognize small targets and occluded objects. The deep learning perception recognition technology for orchard fruits has achieved breakthroughs through algorithm innovation and hardware optimization, building a "data augmentation feature optimization precise positioning" technology chain.

Innovative target recognition algorithms for complex scenarios, considering the characteristics of dense fruit density and severe occlusion in low rootstock orchards, the improved YOLOv7 algorithm achieves a significant performance improvement. Introducing MobileOne module to reconstruct the backbone network, reducing model weight by 1.9MB while improving feature extraction efficiency, changing serial feature fusion to a parallel channel structure to solve the problem of small target feature loss, adding auxiliary detection heads, designing detection branches for fruits of different scales from 70-100mm, and improving the accuracy of overlapping fruit bounding box regression [2]. After improvement, the accuracy of the model increased by 6.9 percentage points, the recall rate increased by 10 percentage points, and mAP1 increased by 5 percentage points. The missed detection rate in large field of view scenes decreased by 67.74%, making up for the poor adaptability of traditional models to dense occlusion scenes [2]. The dataset is constructed and upgraded synchronously, using an overlapping image segmentation method to avoid target segmentation, combined with multi-angle rotation, brightness dynamic adjustment(±50%), and other strategies to expand the sample size from 1245 to 12450, covering extreme scenes and providing a guarantee for the model's generalization ability [2]. In summary, the innovation of target recognition algorithms and the upgrade of dataset construction have significantly enhanced the accuracy and generalization ability of fruit recognition in complex orchard environments.

The three-dimensional positioning accuracy is optimized and adapted. The traditional binocular vision positioning is affected by depth image noise and fruit oscillation, and the error often exceeds 10mm [3]. The orchard fruit multi-sensor collaborative high-precision positioning technology realizes high-precision positioning through hardware selection and algorithm optimization. The RealSense435 depth camera is used, with a depth frame rate of 90fps in the range of 0.105-9m, providing stable data for dynamic scenes. The hand-eye calibration adopts the "eye in hand" mode, and more than 10 groups of pose data are fused by the Tsai-Lenz method. The coordinate system conversion error is<2mm [2]. The depth information processing adopts the "center+neighborhood" multi-pixel depth average method, which extracts the target center point and four neighborhood pixel depth values and averages them, reducing the error by 40% compared with a single pixel [2]. The test shows that within the picking range of 400-900mm, the average deviation of the X axis is 3.08mm, the Y axis is 3.67mm, the Z axis is 4.50mm, and the maximum error is ≤ 8mm, which meets the requirements of grasping accuracy and makes up for the defects of traditional positioning technology [2]. It can be seen that the collaborative optimization of hardware selection and positioning algorithms has realized the accurate three-dimensional positioning of fruits in complex orchard scenes.

2.3. Grasping Execution and Control Technology

The grasping execution system is the core of the interaction between the manipulator and the fruit, which needs to meet the requirements of "adaptive grasping, lossless control, and efficient response". The traditional manipulator has the problems of complex structure, poor matching of grasping force, slow response, and so on. The early two-fingered manipulator has insufficient envelope to the abnormal fruit, and the damage rate is more than 20% [3].

2.3.1 Structure innovation and adaptability design of bionic manipulator

The bionic four-finger underactuated manipulator takes structural innovation and adaptive design as its core, and adopts a symmetrical four-finger two-joint structure. The curved finger surface can fit the spherical shape of the apple, and the contact area is increased by 30% compared with the flat finger, which can effectively disperse the grasping pressure [2]. Its structure is an aluminum alloy framework+flexible silica gel contact layer, which not only ensures the rigidity of anti-branch interference, but also has the performance of buffering and protecting fruits.

In the drive design, through the under drive strategy, the degree of freedom is constrained by the connecting rod drive and torsion spring, and the posture can be adjusted adaptively when the near finger contacts the fruit, so as to realize the full envelope grasp of a 70-100mm diameter apple. This design reduces the driver by 60% and the weight by 40% compared with the full drive manipulator [2]. ADAMS simulation shows that the manipulator is stable without jamming, and the grasping response time is<1.7s, which effectively solves the problems of structural redundancy and poor adaptability of traditional manipulators [2].

2.3.2 Accurate perception and data optimization of the force sensing system

Force sensing is the core of lossless grasping. Traditional sensor data is vulnerable to interference, and a single sensor is difficult to reflect the complete state [3]. The current system uses "multi-sensor fusion+Intelligent filtering", a four-finger inner integrated membrane pressure sensor, with an infrared distance sensor to detect the fruit distance, a Hall sensor to limit the opening and closing limit, forming a multi-dimensional perception network [2]. Exponential moving average (EMA) filtering (α=0.05) was used for data processing, which was 25% faster than moving average filtering;

Smote oversampling was used to expand the data set from 400 to 740 to solve the problem of sample imbalance, and the F1 score of the SVC classification model reached 0.96 [2]. Haoqian's team also pointed out in the unstructured environment research that multi-sensor fusion can improve the grasping stability [4].

2.3.3 Compliance force control strategy of MPC and SVC fusion

Traditional control is difficult to match the individual differences of fruits, and is prone to damage or falling off [3]. The current technology combines model predictive control (MPC) and support vector classification (SVC) to build a closed-loop system. SVC classifies and grabs the state in real time based on the force sensor data, outputs the compensation direction, and MPC dynamically adjusts the pulse number of the stepping motor with 15N as the target threshold [2]. The test shows that the system completes grasping within 2.3S, the contact force is stable at 15-22N, the CT scan has no internal damage, and the damage rate is reduced to 0, which solves the problem of grasping force matching [2].

2.3.4 Path planning optimization in an unstructured environment

The traditional RRT algorithm has the problems of long planning time and path redundancy. At present, the Bi RRT * algorithm, which integrates a heuristic strategy, introduces a gravity field and a repulsion field model, visualizes obstacle areas through a genetic algorithm, adaptively adjusts step size, accelerates exploration in open areas, and improves obstacle avoidance accuracy in dense areas. Compared with the original algorithm, the planning time is reduced by an average of 1.125s, the number of iterations is reduced by 107.25 rounds, and the average time from the initial pose of the manipulator to the target fruit in the dwarf rootstock orchard is 4.20-4.43s, which makes up for the shortcomings of the traditional algorithm such as low efficiency and poor obstacle avoidance [2].

Through the above innovation, the perception and recognition technology realizes the high-precision identification and positioning of apples in a complex environment, and the grasping execution and control technology achieves adaptive lossless grasping and efficient obstacle avoidance, forming a complete scheme suitable for the low rootstock planting scene, laying the foundation for the field application of intelligent picking equipment.

3. Application and Research Prospects in Agriculture

3.1. Application in Agriculture

Apple picking technology focuses on "high precision recognition+adaptive capture". Chen Qing's team improved the Mask R-CNN algorithm. The recognition accuracy of overlapping apples reached 97.31%, and combined with rgb-d camera, the positioning deviation was ≤ 8mm [2]. Yang Huawei's four-fingered underactuated manipulator uses an arc-shaped finger surface and a flexible silica gel layer, which is suitable for 70-100mm apples. With the MPC - SVC force control system, it can grasp without damage in 2.3S, and the success rate of the orchard test is 96.5% [3]. In the unstructured environment, Hao Qian's team's flexible tactile sensor improves grasping stability through silicone deformation and visual feedback. The digital twin system built by Yun Daixing and others on the basis of Unity has a joint control accuracy of 0.1 ° and can preview the path of a low rootstock dense orchard [4] [5]. An American abundant robotics vacuum adsorption arm is used to collect 15 pieces per minute in a Washington State orchard, but the coverage rate of the inner layer of the canopy is less than 50% [2].

Large and medium-sized fruits and vegetables picking relies on modular design breakthroughs. The Spanish Navas team has a modular hexagonal soft grip, with a 3D printing TPE flexible actuator and an iris cutting mechanism. The fruit stalk is cut off in 3S, and the sweet pepper picking in Madrid greenhouse is undamaged [6]. Maochengwei proposed an aluminum alloy lightweight manipulator, which can reduce the weight of the hollow structure by 40%, and the surface oxidation treatment can improve the high humidity resistance. With the flexible end effector, the tomato damage rate can be reduced to less than 2% [7].

3.2. Research Prospects

3.2.1 Existing challenges and difficulties

The landing of the agricultural intelligent manipulator grasping system technology is restricted by three aspects. First, the environmental adaptability is poor. The high temperature in the open orchard causes the deformation of PLA components. When the humidity is>60%, the airtightness of the pneumatic gripper decreases by 20%, the UV resistance linearity of carbon fiber material is weak, the service life is shortened by 30%, and the maintenance frequency increases after the end effector contacts the soil [6] [7]. Second, the cost is high. Faulhaber servo motor and other core components are imported. The localization rate in China is less than 30%, and the cost of the whole machine is more than 500000 yuan [2] [6]. The third is the lack of standardization. The planting row spacing and varieties of different orchards vary greatly. It takes 2-3 days to debug the manipulator, and the industry lacks unified planting and operation standards, so the versatility of the equipment is limited [2] [8].

Lack of environmental adaptability is the core obstacle, which is the drastic change of light (the recognition rate of green apples at night<80%) and the fruit oscillation caused by wind (amplitude 5-10mm), resulting in the positioning error of more than 8mm [2]. The Haoqian team also pointed out that the image details are lost in low-light environments, and additional enhancement processing is required to ensure the recognition accuracy [4].

In addition, the contradiction between hardware and cost is prominent. Agricultural intelligent manipulator grabbing equipment is mostly designed for a single crop, and the end effector needs to be redesigned for cross-crop adaptation (for example, an apple picking robot is replaced by a citrus picking robot), which increases the cost by 50%-80% [2]. Although 3D printing reduces the cost of end effectors (<100 yuan/piece), the core components, such as force sensors and servo motors, rely on imports, and the performance is prone to decline after domestic replacement (for example, the torque error of domestic servo motors exceeds 10%), and the compatibility of modular components is insufficient. Multi-crop adaptation still needs secondary development, which further increases the application cost [6] [7]. At the same time, the average annual operation time of the robot is<100 days, and the average operation cost per mu is more than 500 yuan, much higher than the labor cost of 300 yuan [2].

It is difficult to balance the real-time performance and accuracy of the algorithm. Although the improved yolov7 improves the apple recognition accuracy by 6.9 percentage points, its running speed on the embedded platform is only 26.3 frames/s [2]. Although the hog+svm algorithm improved by Hao Qian's team improves the accuracy of small target detection, it still needs 0.8s to process a single frame image in complex occlusion scenes [4]; the Jacobi iteration method takes>0.5s to solve the inverse kinematics, resulting in a response delay of more than 1s [2].

The conflict between losslessness and efficiency is obvious. In order to avoid fruit damage, the end effector needs to control the grasping force< 25N, resulting in an increase in the number of posture adjustments. Yanghuawei's team's experiment showed that the diameter of the fruit stalk was<5mm, and it was easy to be covered by branches and leaves, and the deviation of the cutting position would cause the fruit abscission rate to exceed 5% [3].

3.2.2 Future direction

Agricultural machinery and agronomy collaboration needs to simplify the technical difficulty at the source. In terms of planting mode, V-type and tree wall type Apple planting were promoted to increase the visibility of fruits by 40% [2]. Yun Daixing and others proposed to integrate the digital twin technology into the orchard planning stage, preset the operation path of the manipulator through virtual simulation, and guide the pruning of fruit trees to form a "machine-friendly" tree [5]. In terms of variety cultivation, breed varieties with moderate stalk toughness and uniform fruit size, and reduce the adjustment frequency of the end effector [2].

Multimodal perception and algorithm fusion can improve the robustness of a complex environment. At the perception level, Hao Qian's team combined vision touch joint perception to establish the mapping relationship between the two perception methods and improve the working efficiency of the robot in the apple picking task [4]. At the algorithmic level, the MPC - SVC fusion control (2.3S lossless capture) is combined with the ghformer net (small green fruit detection at night), and reinforcement learning is introduced to optimize the capture posture. The goal is to reduce the whole process time to less than 5S [2].

Modularization and low-cost design promote the popularization of small and medium-sized orchards. In terms of hardware, the replaceable finger module (suitable for apple, sweet pepper, and citrus) was designed based on the modular concept of the Navas team. Maochengwei proposed to use the combination of aluminum alloy and 3D printing (tpe+pla) to realize the rapid customization of the end effector, with the cost controlled within 100 yuan [7].

Multi-arm cooperation and man-machine cooperation break through the bottleneck of efficiency. Develop a 2-4-arm coordination system. The main arm is responsible for picking, and the auxiliary arm pulls out the branches and leaves, and realizes real-time communication through 5 G. The goal is to improve the picking efficiency to15-20per minute [9]. The human-machine cooperation mode is introduced, and the fruit is missed manually through the plate marking, and the robot completes the grasping, aiming to improve the success rate of picking to more than 98% [2].

5G is integrated with the Internet of Things to build a smart harvesting system. The digital twin system based on 5 G is used for cloud-side collaboration. Only sensors and actuators are reserved at the robot end. Visual recognition and path planning are completed in the cloud. The computing speed is increased by three times, and the cost of local hardware is reduced [5]. Combine the Internet of Things to build orchard digital twins, real-time monitor robot position, fruit maturity, and dynamically schedule operation tasks, and shorten the average harvest time per mu to 2 days [2].

4. Conclusion

This research focuses on the intelligent manipulator grabbing system in the field of agriculture. Aiming at the pain point that China's apple industry is the largest in the world, but the mechanization rate is less than 3% and the cost is high, this research focuses on the scene of Dwarf Rootstocks densely planted orchards, and achieves breakthroughs in many aspects of the core technology. The perceptual recognition end improves the Yolov7 algorithm, improves the recognition accuracy and recall rate, and reduces the missing detection rate of dense scenes. The innovative four-finger underactuated manipulator and MPC - SVC fusion force control system at the grasping execution end realizes lossless and fast grasping. The path planning optimization algorithm can shorten the planning time and the number of iterations. Application verification showed that the success rate of picking was 96.5%, and the single fruit took 9.27-10.20 seconds. However, the system still faces challenges such as poor environmental adaptability, high hardware cost, a difficult balance between real-time performance and accuracy of the algorithm, and lossless and efficiency conflict. In the future, it is necessary to promote the large-scale industrial application of intelligent picking equipment and help agricultural modernization through agricultural machinery and agronomy collaboration, multimodal integration, modular design, etc.

Compared with the existing research, this paper makes up for the limitations of single research fragmentation by strengthening the collaborative Association of technical modules, refining the difficulties of specific scene adaptation, and integrating the whole process indicators of "identification grab picking". In the future, it is necessary to promote the large-scale industrial application of intelligent picking equipment and help the development of agricultural modernization through the optimization of agricultural machinery and agronomy collaboration, multimodal perception fusion, modular low-cost design, and other directions.

References

[1] National Bureau of Statistics. Statistical Bulletin on National Economic and Social Development of the People's Republic of China in 2022. People's Daily, 2023.

[2] Chen Q, Yin C K, Guo Z L, et al. Research status and development trends of key technologies for apple picking robots. Journal of Agricultural Engineering, 2023, 39(4): 1-15.

[3] Yang H W. Research on key technologies for intelligent picking of apples in low-rootstock dense planting orchards. Shandong Agricultural University, 2024.

[4] Hao Q. Research on key technologies of apple picking robots in unstructured environments. North China University, 2023.

[5] Lu D X, Xue X Y, Wang Y B, et al. Design and implementation of a digital twin system for a robotic arm in automated agriculture. Ningxia Agriculture and Forestry Science and Technology, 2024, 65(9): 5-10+2.

[6] Navas E, Blanco K, Rodríguez-Nieto D, et al. A modular soft gripper with embedded force sensing and an iris-type cutting mechanism for harvesting medium-sized crops. Actuators, 2025, 14(9): 432.

[7] Mao C W. Overview of structural design of agricultural picking robot arm. China Agricultural Machinery and Equipment, 2025, (9): 72-74.

[8] 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. Journal of Intelligent and Robotic Systems, 2023, 107(2): 21.

[9] Yu Z, Lu C, Zhang Y, Jing L. Gesture-controlled robotic arm for agricultural harvesting using a data glove with bending sensor and OptiTrack systems. Micromachines (Basel), 2024, 15(7): 918.