



# Fruit recognition, task plan, and control for apple harvesting robots<sup>1</sup>

## Reconhecimento de frutas, plano de tarefas e controle para robôs de colheita de maçãs

Huawei Yang<sup>2,3,4</sup>, Jie Wu<sup>3</sup>, Aifeng Liang<sup>5</sup>, Shaowei Wang<sup>2</sup>, Yinfa Yan<sup>3</sup>, Hongjian Zhang<sup>3</sup>,  
Ning Li<sup>2</sup>, Yinzeng Liu<sup>2</sup>, Jinxing Wang<sup>3</sup> & Jianfeng Qiu<sup>4\*</sup>

<sup>1</sup> Research developed at Shandong Academy of Agricultural Machinery Sciences, Jinan Shandong, China

<sup>2</sup> Shandong Academy of Agricultural Machinery Sciences, Jinan Shandong, China

<sup>3</sup> Shandong Agricultural University/College of Mechanical and Electronic Engineering, Taian Shandong, China

<sup>4</sup> Shandong First Medical University & Shandong Academy of Medical Sciences/School of Radiology, Taian Shandong, China

<sup>5</sup> Tsinghua University/Department of Physics, Beijing, China

### HIGHLIGHTS:

*The labour-intensive harvesting process leads to high production costs of apples.*

*Automatic mechanized harvesting reduces labour costs and enhances competitiveness.*

*Deep learning and collision-free path planning to achieve lossless harvesting of apples can replace manual harvesting.*

**ABSTRACT:** Intelligent apple-harvesting robots use a staggered distribution of branches and leaves during operation, causing problems such as slow motion planning, low operational efficiency, and high path cost for multi-degrees-of-freedom (DOF) harvesting manipulators. This study presents an autonomous apple-harvesting robotic arm-hand composite system that aims to improve the operational efficiency of intelligent harvesting in dwarf anvil-planted apple orchards. The machine vision system for fruit detection uses the deep learning convolutional neural network (CNN) YOLOv7 and RGB-D camera online detection coupling technology to rapidly recognise apples. The spatial depth information of the fruit area was then extracted from the aligned depth image for precise positioning. Coordinate transformation was used to obtain the coordinates of the fruit under the coordinate system of the manipulator. Based on the informed rapid-exploration random tree (Informed-RRT\*) algorithm and path-planning model, the identified target apples were harvested without collision path planning. In an apple-harvesting test, the recognition accuracy of the visual system was 89.4%, and the average time to harvest a single apple was 9.69 s, which was 4.8% faster than the mainstream general harvesting technology. Moreover, the harvesting time for a single apple was reduced by 1.7%. Thus, the proposed system enabled accurate and efficient fruit harvesting.

**Key words:** apple harvesting robotic arm-hand composite, manipulator, deep learning, path planning, harvesting sequence planning

**RESUMO:** Os robôs de colheita inteligente de maçãs têm uma distribuição escalonada de galhos e folhas durante a operação, o que causa problemas como o planejamento de movimentos lentos, a baixa eficiência operacional e o alto custo de trajetória dos manipuladores de colheita com vários graus de liberdade (DOF). Este artigo apresenta um sistema composto de braço-mão robótico autônomo para colheita de maçãs que visa melhorar a eficiência operacional da colheita inteligente em pomares de maçãs plantadas com anéis anões. O sistema de visão mecânica para detecção de frutas usa uma rede neural convolucional de aprendizagem profunda (CNN) YOLOv7 e uma tecnologia de acoplamento de detecção on-line de câmera RGB-D para reconhecer rapidamente as maçãs. Em seguida, as informações de profundidade espacial da área da fruta são extraídas da imagem de profundidade alinhada para um posicionamento preciso. A transformação de coordenadas é usada para obter as coordenadas da fruta no sistema de coordenadas do manipulador. Com base no algoritmo de árvore aleatória de exploração rápida informada (Informed-RRT\*) e no modelo de planejamento de caminho, as maçãs-alvo identificadas são colhidas sem planejamento de caminho de colisão. Em um teste de colheita de maçãs, a precisão do reconhecimento do sistema visual foi de 89,4%, e o tempo médio de colheita de uma única maçã foi de 9,69 s, 4,8% mais rápido do que a tecnologia de colheita geral convencional. Além disso, o tempo de colheita de uma única maçã foi reduzido em 1,7%. Assim, o sistema proposto permite uma colheita de frutas precisa e eficiente.

**Palavras-chave:** composição robótica de braço e mão para colheita de maçã, manipulador, aprendizagem profunda, planejamento de caminho; planejamento de sequência de colheita



## INTRODUCTION

Apples are the most abundant fruit tree cultivars in northern China. According to statistics, the national apple planting area reached 2.08808 million hectares in 2021, and the national apple production reached 45.9734 million tons (Yang et al., 2020). In apple production operations, fruit harvesting is an important link (Legun et al., 2021) and is part of seasonal, labour-intensive operations. However, with the rapid advancement of urbanisation in China, the rural labour force has decreased and aged, and labour costs have increased significantly (Zhong et al., 2019). Thus, intelligent apple-harvesting technology can help alleviate labour shortages, reduce labour intensity, improve apple-harvesting quality and efficiency, and improve planting efficiency (Tian et al., 2019).

In recent years, dense dwarf anvil planting and the cultivation of apple orchards have become mainstream in the combined development of orchard agronomy (Reig et al., 2019), and supporting orchard-harvesting robot technology has developed rapidly. In addition, significant progress has been achieved in fruit identification and positioning, motion simulation, trajectory planning, and flexible grasping.

At present, the success rate of apple-harvesting robots remains mediocre, and failures are mostly due to the obstruction of fruits by apple clusters, branches, leaves, and other obstacles, as well as calibration and spatial positioning errors caused by the harvesting system and the swinging of branches and fruits (Zhang et al., 2016). To solve the above problems, it was proposed an apple-harvesting robotic arm-hand composite system that integrates a vision system, flexible end effector, and manipulator, and features an intelligent algorithm. In the vision system, the binocular visual camera system is used as the main visual sensor, and a fruit deep learning model is constructed to identify and locate the target apple. Subsequently, the operation path of the manipulator with the end effector was planned to complete the grasping action and realise accurate and collision-free fruit harvesting. This study presents an autonomous apple-harvesting robotic arm-hand composite system that aims to improve the operational efficiency of intelligent harvesting in dwarf anvil-planted apple orchards.

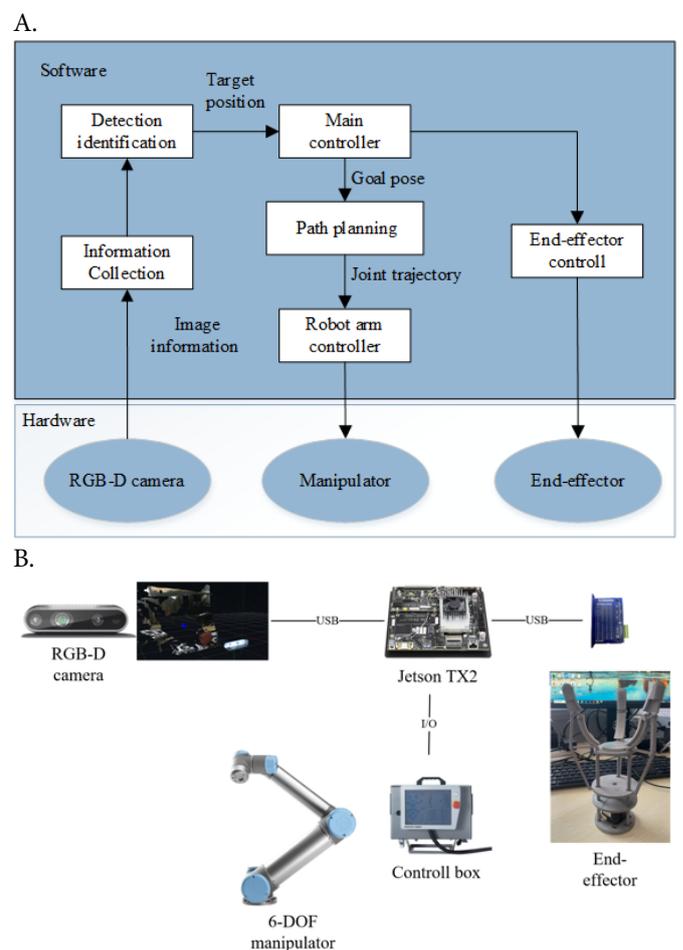
## MATERIAL AND METHODS

In the orchard scenario under the dense cultivation method of a dwarf anvil, the high-density cluster distribution of apples, staggered distribution of branches and leaves, and presence of other obstacles pose challenges to the identification and avoidance of harvesting (He et al., 2017). In this paper, it is presented a harvesting system designed to realise precise positioning and non-destructive grasping of apples in the environment of a dense dwarf anvil-planted orchard. It also presents an intelligent algorithm model, improves the adaptability and robustness of the system, and demonstrates the efficient harvesting operation of the harvesting manipulator-hand complex.

The harvesting manipulator-hand complex hardware system has a modular design that is mainly composed of a

six-DOF manipulator, a RGB-D binocular stereo camera, an end effector, and an upper computer. The manipulator was selected from the six-DOF manipulator UR5 of the Danish UR Company as a carrier for the movement of the end effector. The end effector adopts a three-finger flexible adaptive bionic mechanical easy claw to achieve damage-free fruit grasping. Based on YOLOV7 deep convolutional neural network (CNN) model recognition, the binocular stereo camera selects realsense435i, which is responsible for capturing images online and calculating depth, in addition to combining the camera internal reference matrix to obtain the 3D spatial position of the fruit and complete fruit recognition and positioning. The NVIDIA Jetson TX2 image edge computing unit running Ubuntu 18.04 as the main controller and the Python programming language were used to control operational behaviours such as manipulator movement, end effector closure, fruit detection, and data logging in the robot operating system (ROS). The harvesting system software and hardware are shown in Figure 1.

When harvesting apples from orchards, a multi-DOF manipulator is used as the harvesting device, which is combined with visual-based fruit recognition and trajectory planning algorithms to perform automatic fruit harvesting. The orchard harvesting scenario differs from the structured working environment of a traditional industrial scenario. In other words, the working environment of a harvesting robot is a complex, changeable, and unstructured natural environment,



**Figure 1.** Harvesting robot software system (A) and hardware system (B)

and obstacles such as naturally growing branches and immature fruits create difficulties in harvesting by the manipulator. The agronomy of different orchard cultivation types also affects the harvesting operations. Orchard harvesting requires a manipulator to comprehensively consider the cultivation method of the fruit so that the harvested crop is within its working space, based on an auxiliary sensor and dynamic planning of the harvesting trajectory (Zhang et al., 2020). The manipulator can then avoid obstacles (leaves and stems) and accurately grasp the target fruit without injuring the plant.

The harvesting operation of the apple-harvesting robotic arm-hand complex can be divided into three stages. (1) During the visual inspection stage of the harvesting robot, the camera captures an image in real time, and the main control machine determines the presence of harvestable apples through the visual model. (2) In the positioning stage after detecting a mature, harvestable apple by the vision system, the camera acquires images after the manipulator becomes relatively stationary, thereby reducing the relative spatial position error of the target apple. The system then identifies the target apple using a CNN to complete the spatial positioning. (3) In the execution stage, the harvesting sequence of the target apples is planned by the end effector such that the robotic arm-hand complex physically interacts with the fruits. All ripe fruits in the field of view can then be harvested and placed in harvesting order. After one fruit is harvested, the complex returns to its initial position and enters the thread of the next harvest, thereby repeating the harvesting process. If the operation fails, the fruit is not harvested and the robotic arm hand returns to the starting point.

Owing to the excellent feature learning and feature extraction abilities of a CNN (Schertz & Brown, 1968), combined with the stable depth extraction ability of a RGB-D binocular vision sensor, a coupled model designed to detect and localise target fruits was designed. The canopy of the standard orchard with dense planting of a dwarf anvil is completely narrow, the fruit distribution is similar to that of the “fruit wall”, the topology of the fruit is simple, and the obstacle avoidance planning is easy to achieve.

The primary condition for the automatic harvesting of apples is the detection and positioning of an apple on a tree (Bloch et al., 2018). Apple detection aims to separate apples from a complex orchard background (i.e., mixed scenes of leaves, branches, and other fruits). Fruit localisation attempts to calculate the 3D space for apple detection more accurately than a camera coordinate system. This system uses an Intel RealSense435i RGB-D double-sided visual sensor to collect environmental information. The visual system has the advantages of structural compactness, small volume proportion, and high accuracy (Hayashi et al., 2010).

In the complex, unstructured environment of orchards, the detection and positioning of apples can be subject to interference, such as leaf occlusion and mutual occlusion between fruits (Xiong et al., 2020). To solve these problems and improve the robustness and stability of the apple model recognition, the YOLOV7 deep neural network detection algorithm was used. To train the network, sample datasets were collected from an apple demonstration base in

Shangluojia Village, in the southern mountainous area of Licheng District, Jinan City, Shandong Province, and from an apple demonstration base in Lanting New Village, Longkou City, Yantai City, Shandong Province. The apple trees at the demonstration base were planted densely using dwarf anvils, and the cultivars used were Yantai red Fuji apples. The position and height of the end effector equipped with the camera were simulated during collection. The sample image had 2,350 spokes with a resolution of  $2532 \times 1170$  pixels, and the acquisition device was an Apple iPhone 12. The height of the apple trees was 2.5–3 m, the diameter was 2.4 m. Labelme labelling software was used to label the images and generate the annotation files required for training.

YOLOV7 is currently the mainstream detection model, and the detection accuracy and speed are higher than other single-stage detection models in 5–160 FPS. The network structure is illustrated in Figure 2. At the input end of the model, the input pixels were first converted into an apple orchard image of  $640 \times 640 \times 3$  channels and the original sample was adaptively filled during the process of scaling and filling. During training, the network obtains the prediction box from the initial anchor box calculation, calculates the real box through the adaptive anchor box, and updates it in reverse.

Backbone: Through the feature extraction network, a four-layer CBS module was first used for feature extraction. The feature map contains  $160 \times 160 \times 128$  channels, and through multilayer efficient layer aggregation network (ELAN) and maximum pooling (MP) modules, in which the ELAN controls the shortest and longest gradient paths, the feature extraction ability of the network can be enhanced with stronger robustness. The role of MP is downsampling; that is, it outputs feature maps C3, C4, and C5 as  $80 \times 80 \times 128$ ,  $40 \times 40 \times 256$ , and  $20 \times 20 \times 1024$  channels, respectively. The head had a PA Net structure. First, the channel is reduced by SPPCSP for C5. Then, upsampling is performed from top to bottom, and C3 and C4 are fused to obtain feature maps P3, P4, and P5. Then, upsampling is performed from the bottom up, and P4 and P5 are fused. For the feature map output by the neck, the  $1 \times 1$  convolution was used for prediction after adjusting the channel using Rep Conv.

For the apples detected in the bounding box, the location information was calculated by combining the depth information from the RealSense RGB-D camera. Combined with the depth and anchor frame information, the Cartesian position of the apples was determined by back-projection, and iterative calculations were performed on each apple to obtain the position of apple detection under 2-bit pixels.

When the vision system recognises multiple apples to be picked and outputs the spatial coordinates, the order in which the outputs are formed is random. If the manipulator picks randomly in this order, it can easily cause the gripper to collide with the apples around the target apple (Ouf, 2023), causing the apple to be picked to shake and the real position to shift from the initial position. Therefore, a reasonable picking order can overcome problems such as interference errors caused by end effectors during the picking process.

In a group of apples, the end actuator preferentially picks the smallest apple near the apple (De-An et al., 2011), and the



Figure 2. Structure diagram of the YOLOV7 visual recognition network

area near the apple will be empty, which will reduce collisions with the target fruit during the operation of the end actuator, thereby reducing the target fruit, thereby reducing, the error caused by the target apple offset. Therefore, it is necessary to calculate the number of apples around the apples. The YOLOV7 detector predicts the pixel position of the fruit, which is marked with a boundary frame. The boundary box generally contains coordinates in the upper-left and lower-right corners. As shown in Eqs. 1 and 2, the coordinates of the centre of the circle are determined by the bounding box. The radius of the fitted circle was calculated using Eq. 3. The quality of the fruit can be estimated according to Types 1 and 2. The relative distance between the apples is calculated using Eq. 4. Then, Eq. 5 is used to determine whether the two apples obscure each other or not, if the distance between the centroids of the fruits is less than the sum of the radii of the fitted circles of the two apples, then the two apples are overlapping. Finally, the two apples can be adjacent to each other.

$$x_c^i = \frac{x_{ul}^i + x_{dr}^i}{2} \tag{1}$$

$$y_c^i = \frac{y_{ul}^i + y_{dr}^i}{2} \tag{2}$$

$$r_i = \frac{(x_{ul}^i - x_{dr}^i) + (y_{dr}^i + y_{ul}^i)}{4} \tag{3}$$

$$L_i = \sqrt{(x_c^i - x_c^j)^2 + (y_c^i - y_c^j)^2} \tag{4}$$

$$0 < L_i \leq (r_i + r_j) \tag{5}$$

where:

$(x_c^i, y_c^i)$  - coordinates of the geometric center of the apple marked as  $i$ ;

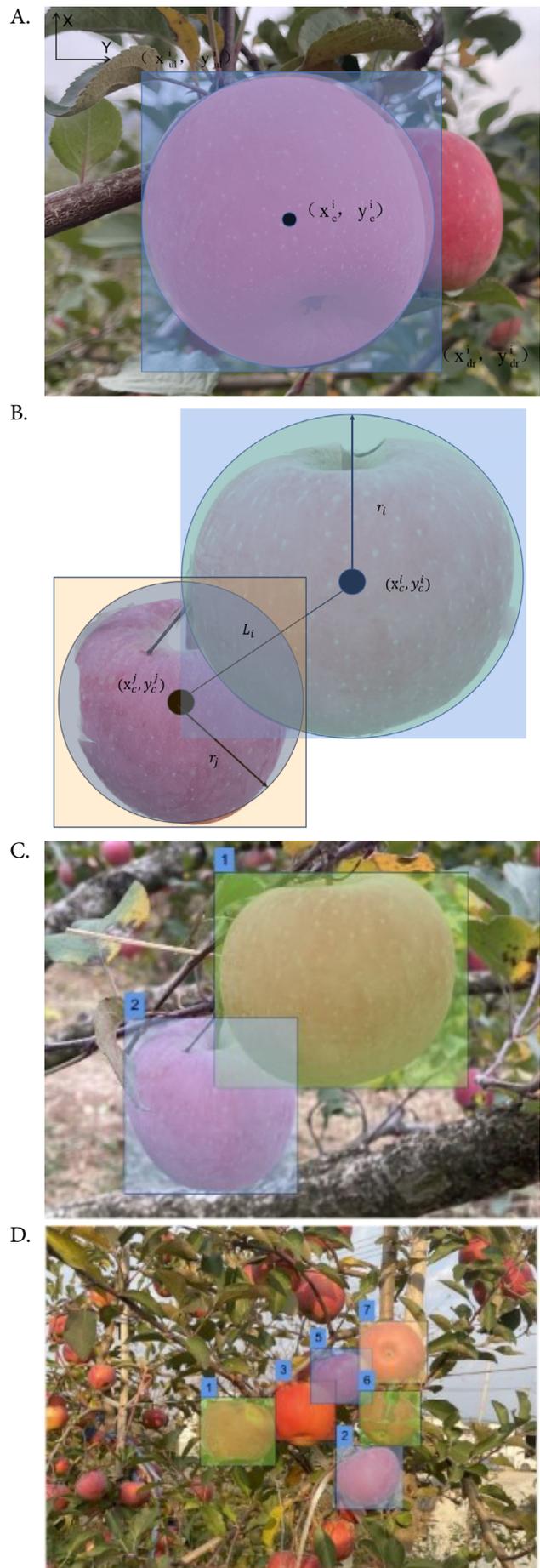
$(x_{ul}^i, y_{ul}^i)$  - coordinates of the upper left corner of the bounding box;  
 $(x_{dr}^i, y_{dr}^i)$  - coordinates of the lower right corner of the bounding box;  
 $r_i$  - radius of the apple fitting circle; and,  
 $L_i$  - distance between the centres of the geometric circles of the two apples.

Next, the apple to be picked with the least number of adjacent apples is listed as the first harvestable object, and after the picking is complete, its information is removed from the sequence, and the next target to be harvested is iteratively calculated until all the apples in the field of view are harvested. In addition, if the same number of apples is adjacent to the harvestable apples, based on the experience of manually picking and selecting ripe apples with the shortest relative distance in the field of view, the depth parameter outputs from the depth sensor are compared, and apples with the shortest distance from the vision sensor are selected for priority harvesting. When harvesting using this model, it can be ensured that each apple to be picked has the least number of apples in its vicinity, thereby reducing the error caused by end-effector interference. The centre of the apple represents the distance between the two apples, as shown in Figures 3A and B. After sorting, picking robots can harvest apples in a reasonable order. As shown in Figure 3D, six apples were selected. Among them, the robot prioritises picking the top, which is marked as one apple (the minimum number near the apple). Multiple apples is near an apple are shown in Figure 3C.

After hand-eye calibration, the rotation and translation matrix from the camera coordinate system to the manipulator coordinate system is obtained. Therefore, only the camera coordinates output by the vision system must be converted into manipulator coordinates through coordinate transformation.

Obstacle-avoidance path planning is a key technology in fruit and vegetable harvesting. This refers to the path from the starting point to the target point given the position of the obstacles and the position of the start and target (Fu et al., 2020). However, the fruit harvest in orchards currently has a long harvest cycle, short window of freshness, fragile appearance, random growth location, and many other challenges, such as branches and leaves, a complicated working environment, and other factors. To solve the problem of high-dimensional planning, researchers have proposed sampling-based motion planning algorithms based on fast-search random trees.

For example, RRT have excellent characteristics and do not require modelling obstacles or exploring high-dimensional spaces. A Cartesian space obstacle is mapped onto the joint space. Subsequently, the free motion space of the manipulator is obtained, and the A\* heuristic search algorithm is used to search the path in the free space of the manipulator. The square with the lowest current "cost" is then selected according to the extension node for the next search until the end point has been searched to plan the path with the lowest cost. This algorithm adopts an ant colony algorithm that combines the principle of positive feedback of information with a heuristic algorithm (Nguyen et al., 2013). Alternatively, based on deep reinforcement learning (DRL), a reward mechanism for



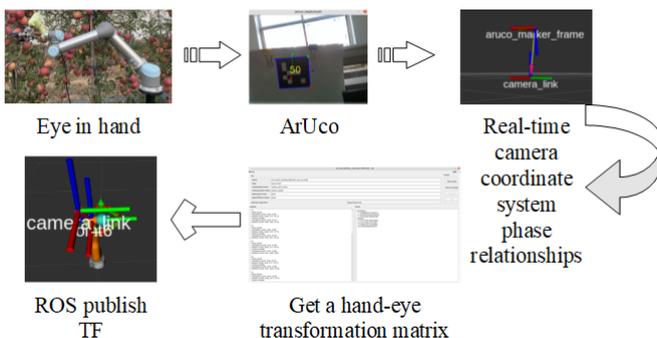
**Figure 3.** Apple image centre coordinate positioning (A), apple centre coordinate distance (B), harvesting sequence example 1 (C), harvesting sequence example 2 (D)

training can be established that rewards the desired path, and plan a suitable obstacle avoidance path (Xu et al., 2023).

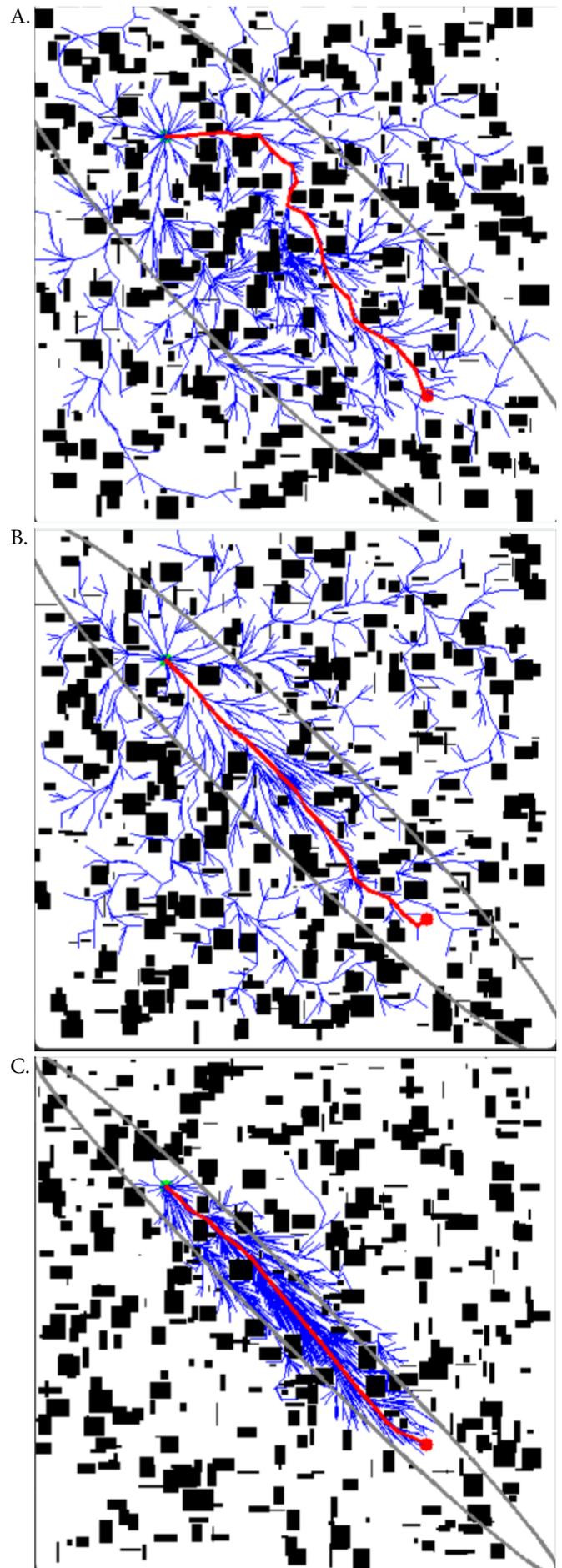
Although the orchard under the dwarf-anvil-dense-planting standard is more standardised, the staggered distribution of its fruit growth branches gives the orchard the characteristics of a typical unstructured complex scene, which requires the path planned by the end path of the manipulator to avoid obstacles. Simultaneously, to improve the movement efficiency and reduce unnecessary repeated path points to reduce the probability of collision, the path length should be short, the coverage area should be small, and unnecessary collisions should be reduced (Wang & He, 2021). When a fruit is detected, the manipulator moves the end effector to a position where it can be harvested without obstacles. The manipulator moves by changing the angles of the six motors on the arm to control the joints. Solving the joint angle of a manipulator at a given target position is called an inverse kinematics problem, and multiple solutions typically exist. According to the Unified Robot Description Format (URDF) used to describe the harvesting system model, the motion control of the manipulator is realised using a special motion planning software, ROS MoveIt, which provides several motion planners and inverse kinematics solvers. For obstacle avoidance, it is necessary to choose IK-FAST and the Open Motion Planning Library (OMPL package with Informed-RRT\*) as kinematics solvers. The hand-eye calibration process is shown in Figure 4.

The optimal programming problem is defined as finding a collision-free ( $X_{obs}$  obstacle space) path from the initial position  $X_{start}$  to the target location  $X_{goal}$  in state space  $X_{free} = X - X_{obs}$ ,  $X_{start}, X_{goal}, X_{obs} \in X, \sigma[0,1] \rightarrow X_{free}$ . For the best path cost, the current solution state subset  $X_f \subseteq X, f(x)$  is the shortest path cost from  $X_{start}$  to  $X_{goal}$ .

The RRT is suitable for solving the motion-planning problem of a high-dimensional space; therefore, it is widely used in motion planning. The RRT is a random expansion method in a global environment, with a constant step size and randomly extended tree growth, and is used to generate a collision-free path from the starting point to the target point. Because the expansion of the RRT is random, it can easily cause redundant operations in a global environment. The asymptotically optimal RRT (RRT\*) achieves asymptotic optimisation by rewiring the nodes and then obtaining the optimal path through continuous iteration (Ma et al., 2023). RRT\* obtains the optimal solution but also entails a huge amount of computation. The informed-RRT\* ellipsoid subset convergence is illustrated in Figure 5. Informed-RRT\*



**Figure 4.** Hand-eye calibration is completed using Easyhand to obtain a hand-eye transformation matrix



**Figure 5.** Example of Informed-RRT\* ellipsoid subset convergence (A, B, and C)

performs the centralised optimal planning of paths using ellipsoid subset sampling. However, because the sum of the distances between the points in the ellipse and the two focal points is less than the sum of the distances between the points on the ellipse and the two focal points, the newly generated path length in the ellipse is shorter than the length of the path outside the ellipse. Thus, as the number of iterations increases, the elliptical space continues to shrink, and the length of the path decreases. Finally, an optimal solution is obtained.

The heuristic domain is an ellipse focusing on the start  $X_{start}$  and target points, and the eccentricity of the ellipse  $C_{min}/C_{best}$  is the ratio of the theoretical minimum cost of the initial state  $C_{min}$  and the target state to the cost of finding the best solution thus far,  $C_{best}$  (Figure 6).

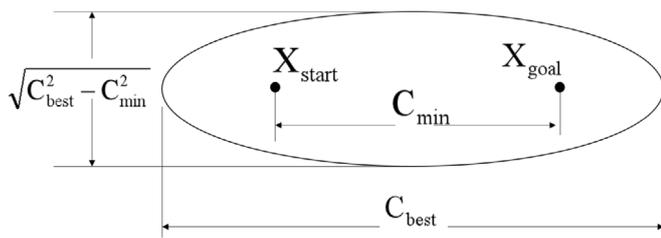


Figure 6. A brief illustration of the ellipsoid subset

## RESULTS AND DISCUSSION

An experimental design was used to evaluate the training recognition advantages and disadvantages of different algorithms using labelled apple datasets. The datasets were divided into training and validation sets in a 7:3 ratio. The model training adopted a deep learning workstation. The CPU was an Intel Core i7-10875H, the graphics card was RTX2060 6 g, the operating system was Ubuntu 18.04, the programming language was Python, and the libraries included pytorch1.11 and opencv4.1. The vision system was deployed in the GPU, CUDA was used for acceleration, and detection was completed in the CPU. To ensure that the anchor box values had a minimal impact on the trained model during training, this experiment adapted the anchor box values to the datasets. Before model training, the k-means++ clustering algorithm was used for the homemade apple datasets to obtain the corresponding prediction box (Table 1). Three detection algorithms (YOLOV4, YOLOV5s, and YOLOV7) were evaluated. All training sets used 300 epochs with a batch size of eight. The initial and minimum learning rates of the optimiser were set to 0.01 and 0.001, respectively, which are the results of experience.

Three performance metrics (precision, recall, and mAP@0.5:0.95) were used to evaluate the detector. Precision is the ratio of the correctly predicted target to the total number of predicted targets. Recall represents the ratio of correctly predicted targets to all labelled targets. The metric mAP0.5:0.95 represents the average mAP for different cross-joint thresholds

Table 1. kmeans++Anchor box value

Feature map scale/pixels	Receptive field	Anchor size		
20*20	Large	18 , 24	32 , 41	77 , 43
40*40	Medium	52 , 80	76 , 109	106 , 160
80*80	Small	144 , 220	186 , 282	265 , 404

(IoU) from 0.5 to 0.95. After training 300 batches, the indices converged and stabilised. As shown in Table 2, the YOLOV7 detector outperformed the other object detection algorithms in the apple dataset, with an accuracy of 0.894, recall of 0.79, and mAP0.5:0.95 of 0.915. Figure 7 compares the recognition effect

Table 2. Comparison of performance parameters of vision model detection algorithms

Model	Precise	Recall	mAP@0.5:0.95
YOLOV4	0.846	0.78	0.668
YOLOV5s	0.874	0.85	0.757
YOLOV7	0.894	0.79	0.915

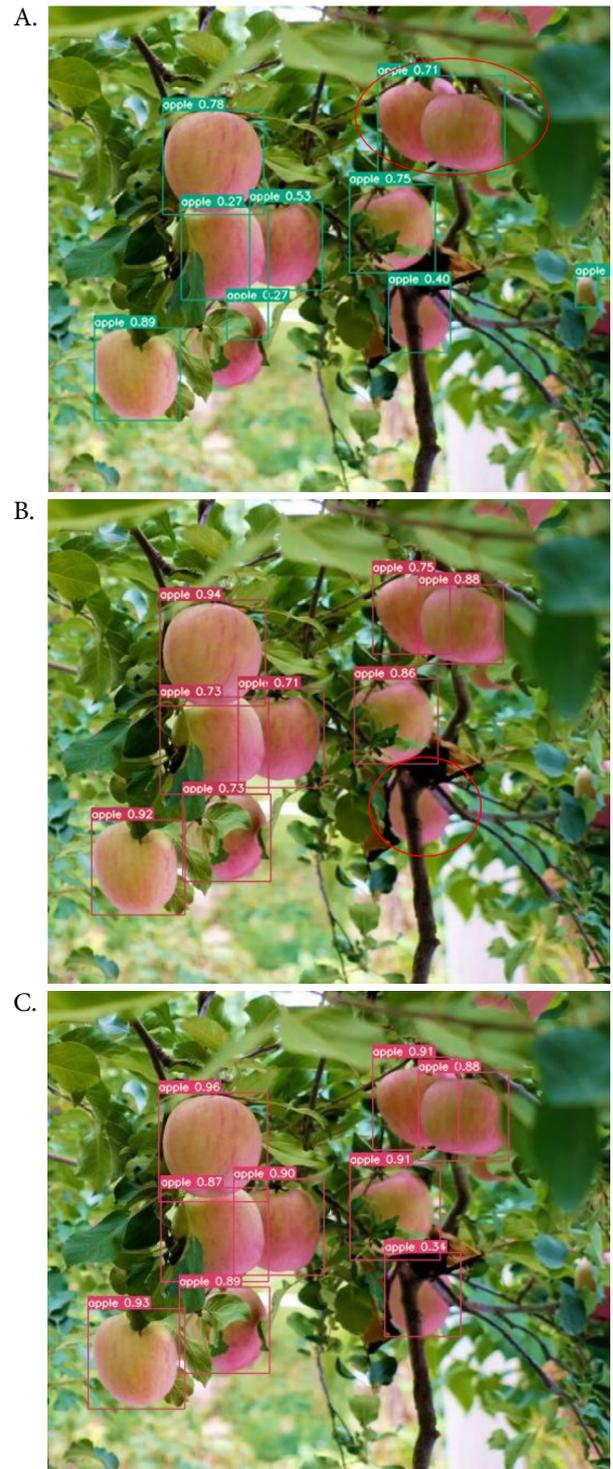
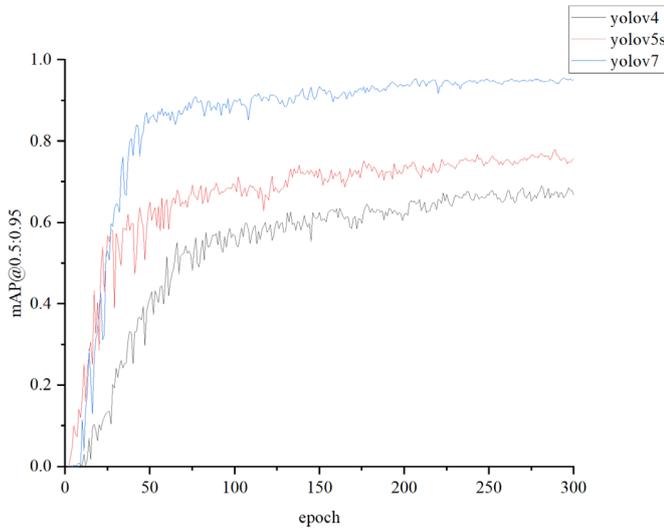


Figure 7. Comparison of detection performance of YOLOV4 (A), YOLOV5s (B), and YOLOV7 (C)

of the three algorithms: YOLOV4 and YOLOV5s algorithms both have leakage detection (the leakage apples are marked with a red round box); the YOLOV7 algorithm is better than the other algorithms in terms of the detection effect. According to Figure 8, the YOLOV7 algorithm is in the training stage of 300 epochs. The speed of convergence in mAP is greater than that of the other algorithms, and the accuracy is also better than that of the other two algorithms.

This experiment compared the results of the RRT, RRT\*, and Informed-RRT\* for path planning through 2D visualisation. As shown in Table 3, the informed-RRT\* sampling nodes were better than those of the RRT and weaker than those of the RRT\*, lagging behind the other two algorithms in terms of sampling time. However, as shown in Figure 9, under the same obstacle space, Informed-RRT\* has relevant paths that are better than those of the RRT and RRT\* in terms of smoothness and path length\*. Moreover, in the unstructured scenario of orchards, redundant repeated paths and unsmooth paths increase the probability of collision between the manipulator and obstacles, such as fruit trees (Nguyen et al., 2013). Therefore, collision-free harvesting of the orchard would be affected, and the time difference used in the planning would be within the acceptable range of harvesting (Zhang et al., 2016). Accordingly, Informed-RRT\* algorithm was selected as the final path-planning algorithm for the manipulator.

To verify the effectiveness of the harvesting system, a harvesting planning test based on the steps shown in Figure 10 was conducted using a harvesting robotic arm-hand complex.



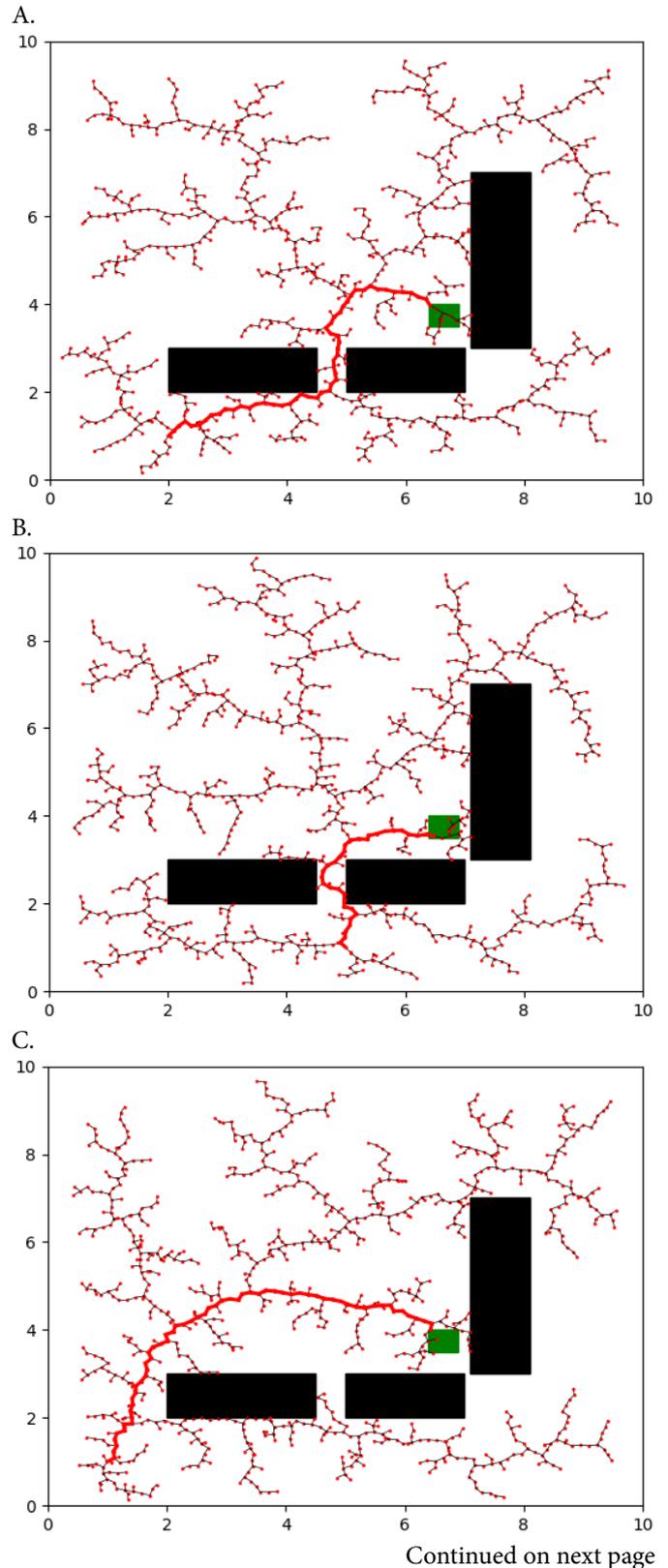
**Figure 8.** mAP0.5: 0.95 comparison of convergence line charts of each algorithm

**Table 3.** Detection results of RRT, RRT\*, and Informed-RRT\* in terms of the number of sampling nodes and planning time

Algorithm	Sampling nodes	Planning time (s)
RRT(1)	894	0.87
RRT(2)	614	1.75
RRT(3)	873	0.73
RRT*(1)	853	4.11
RRT*(2)	913	3.98
RRT*(3)	928	3.92
Informed-RRT*(1)	726	6.80
Informed-RRT*(2)	759	7.34
Informed-RRT*(3)	758	7.20

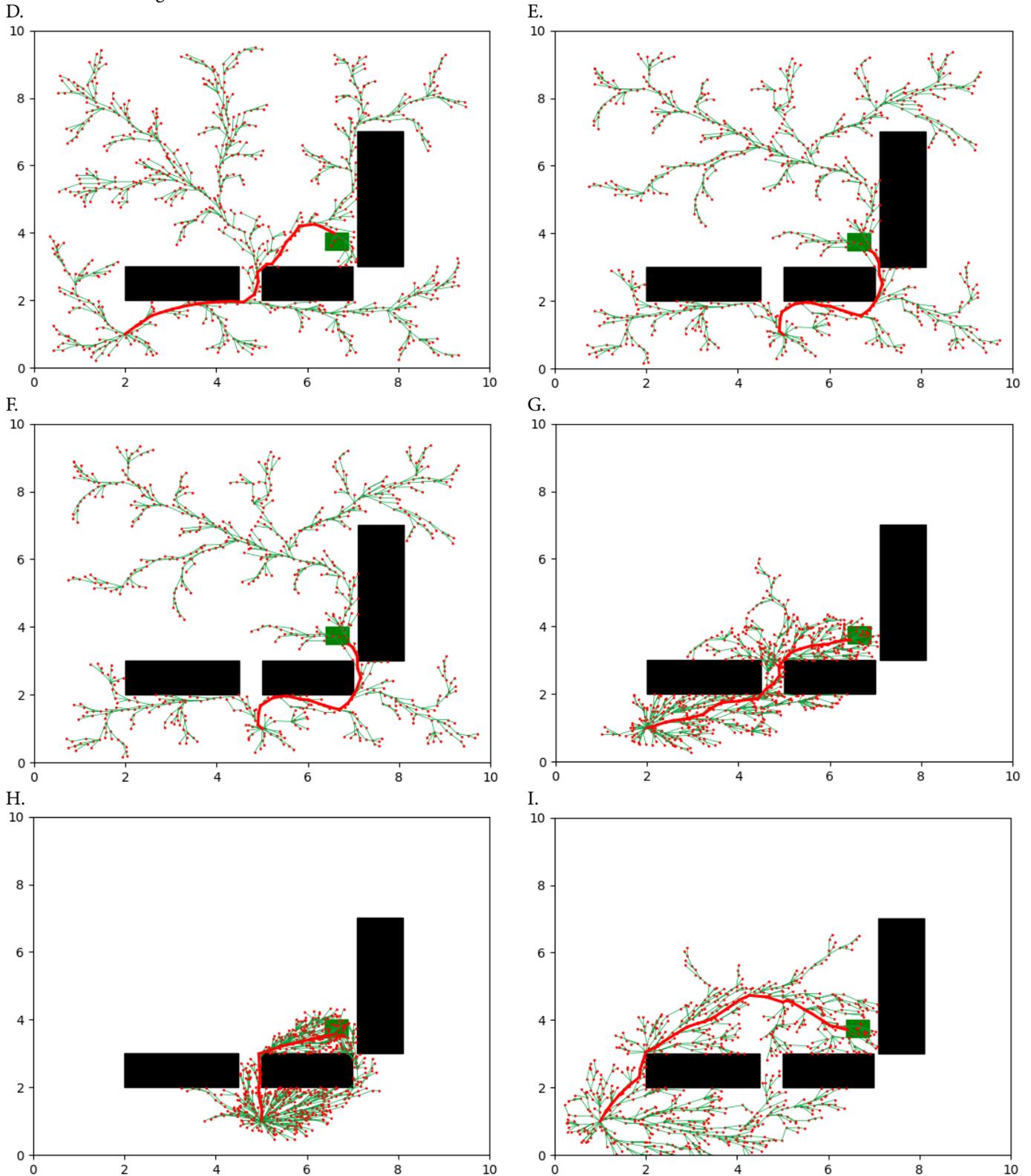
The harvesting rate and harvesting time were used as evaluation indicators, where harvesting rate refers to the ratio of the number of apples harvested to the total number of apples that can be harvested. The harvesting time refers to the time required to harvest a single apple, including the time spent on each step.

Six sets of experiments were set up for the RRT, RRT\*, informed-RRT\*, and each of the three algorithms to fuse the picking and harvesting sequences. Each group comprised 40



Continued on next page

Continued from Figure 9

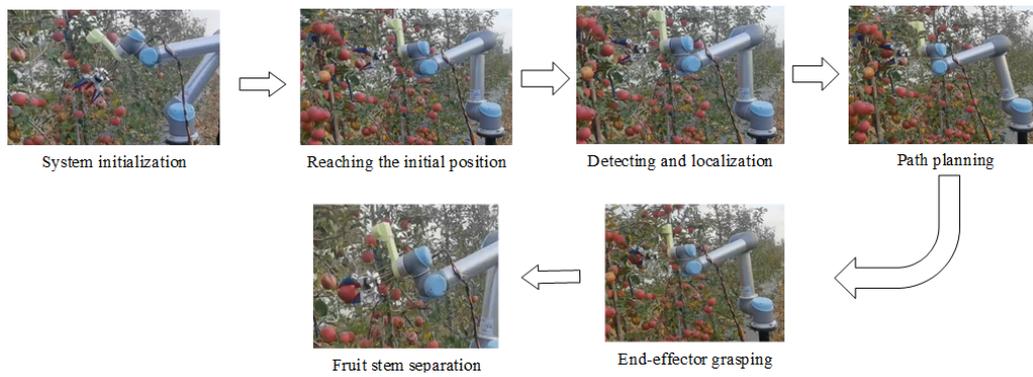


**Figure 9.** Demonstration of three path planning algorithms RRT (A, B, and C), RRT\* (D, E, and F) and Informed-RRT\* (G, H, and I) in three different 2D scenarios.

effective picking tests. The six effective picking groups were 28, 28, 30, 25, 26, and 29. After the harvested apples had rested for 48 hours at 24 °C, no visible abrasions or injuries were found.

From the results of the harvesting test in Table 4, it can be seen that the actual number of harvests and harvesting rate using the harvesting planning strategy were slightly higher than those without the strategy. This is because the harvest-

planning strategy can reduce the effect of dense fruits on the accurate grasping of the manipulator. However, the harvesting rate using the RRT path planning algorithm was slightly higher than that using the Informed-RRT\* algorithm. Nevertheless, in terms of the average harvesting time, the efficiency of the Informed-RRT\* algorithm was higher, and the movement path of the RRT algorithm was longer than that of Informed-



**Figure 10.** Fruit harvesting equipment completing the process of one harvest

**Table 4.** Results of harvesting experiments

Experiment serial number	Algorithm	Harvest planning strategy	Total number of fruits	Average harvesting time (s)	Number of successfully harvested fruits	Harvesting rate (%)
1	RRT*	×	40	12.65	28	70
2	RRT*	√	40	12.70	28	70
3	RRT	√	40	12.89	30	75
4	RRT	×	40	12.65	25	62.5
5	Informed-RRT*	×	40	12.64	26	65
6	Informed-RRT*	√	40	12.69	29	72.5

RRT\* in actual harvesting; therefore, various performance indicators were comprehensive. After the harvesting planning strategy was adopted, the harvesting rate of the RRT (75%) was better than that of the Informed-RRT\* (72.5%); however, the average harvesting time of the former (12.89 s) was longer than that of the latter (12.69 s). Because the informed-RRT\* algorithm ensured optimal paths, which reduced the possibility of redundant movements of the manipulator, and because the RRT path was not asymptotically optimal, the fast harvesting of fruits was also one of the performance indicators considered during the harvesting (Xiong et al., 2020); therefore, considering the two indicators of harvesting average time and harvesting rate, the use of Informed-RRT\* algorithm and harvesting strategy had a certain optimization effect on harvesting.

**CONCLUSIONS**

1. An apple harvesting device adapted to the background of dwarf anvil dense planting and cultivation method was designed.
2. The visual perception system and mechanical system coordination design were integrated into hardware.
3. Visual neural network model recognition and manipulator path-planning obstacle avoidance problems were integrated in the software.
4. These steps aimed to solve the problem of apple cluster distribution recognition in unstructured scenes and obstacle avoidance of fruit branches, leaves, and other obstacles around the target fruit. In this study, based on the accurate detection of a visual neural network and sequence planning of fruits to be harvested, a reasonable harvesting order for fruits under high-density distribution was obtained.
5. Furthermore, a 3D model of an orchard was established based on an RGB-D binocular vision sensor, and collision-free path planning of the manipulator was completed using the informed-RRT\* algorithm.

6. The latter algorithm solves the inverse kinematics problem and introduces ellipsoid subsets to obtain the optimal solution to complete the harvesting of a single apple from one thread. Thus, the proposed system enabled accurate and efficient fruit harvesting. These results can help improve the technical capabilities of future intelligent apple-harvesting robots.

**Contribution of authors:** Huawei Yang: Writing-Original Draft; Methodology, Jie Wu: Data Curation and graph; Shaowei Wang: Writing-Review & Editing; Jinxing Wang: Formal analysis, Supervision, Aifeng liang, Funding acquisition; Jianfeng Qiu: Project administration; Ning Li and Yinzeng Liu: Model training; Yinfa Yan and Hongjian Zhang: Methodology.

**Supplementary documents:** There are no supplementary sources.

**Conflict of interest:** The authors declare no conflict of interest.

**Financing statement:** This research was supported by the earmarked fund for CARS, great number CARS 27; This research was funded by Shandong Province Key Research and Development Plan ( major scientific and technological innovation project ) Project, Agricultural Manipulator Motion Planning and Intelligent Drive Technology Research and Development, grant number 2022CXGC020701 ; This research was funded by Shandong Province Science and Technology Small and Medium-sized Enterprise Innovation Ability Promotion Project, New Orchard Pomegranate Intelligent Picking Manipulator Research and Development and Application, grant number 2022TSGC2253 ; This research was funded by 2022 Science and Technology Think Tank Youth Talent Plan Project, Analysis of The Current Situation, Problems and Countermeasures of Modern Orchard Technology in Shandong Province, grant number 20220615ZZ07110137.

**Acknowledgments:** The author thanks the tutor and all the staff in the team for their guidance and help, and also thanks

the Shandong provincial government for the financial support of the project. Finally, I am really grateful to all those who devote much time to reading this thesis and give me much advice, which will benefit me in my later study.

### LITERATURE CITED

- Bloch, V.; Degani, A.; Bechar, A. A methodology of orchard architecture design for an optimal harvesting robot. *Biosystems Engineering*, v.166, p.126-137, 2018. <https://doi.org/10.1016/j.biosystemseng.2017.11.006>
- De-An, Z.; Jidong, L.; Wei, J.; Ying, Z.; Yu, C. Design and control of an apple harvesting robot. *Biosystems Engineering*, v.110, p.112-122, 2011. <https://doi.org/10.1016/j.biosystemseng.2011.07.005>
- Fu, L.; Majeed, Y.; Zhang, X.; Karkee, M.; Zhang, Q. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting. *Biosystems Engineering*, v.197, p.245-256, 2020. <https://doi.org/10.1016/j.biosystemseng.2020.07.007>
- Hayashi, S.; Shigematsu, K.; Yamamoto, S.; Kobayashi, K.; Kohno, Y.; Kamata, J.; Kurita, M. Evaluation of a strawberry-harvesting robot in a field test. *Biosystems Engineering*, v.105, p.160-171, 2010. <https://doi.org/10.1016/j.biosystemseng.2009.09.011>
- He, L.; Fu, H.; Karkee, M.; Zhang, O. Effect of fruit location on apple detachment with mechanical shaking. *Biosystems Engineering*, v.157, p.63-71, 2017. <https://doi.org/10.1016/j.biosystemseng.2017.02.009>
- Legun, K.; Burch, K. Robot-ready: How apple producers are assembling in anticipation of new AI robotics. *Journal of Rural Studies*, v.82, p.380-390, 2021. <https://doi.org/10.1016/j.jrurstud.2021.01.032>
- Ma, G.; Duan, Y.; Li, M.; Xie, Z.; Zhu, J. A probability smoothing Bi-RRT path planning algorithm for indoor robot. *Future Generation Computer Systems*, v.143, p.349-360, 2023. <http://doi.org/10.1016/j.future.2023.02.004>
- Nguyen, T. T.; Kayacan, E.; De Baedemaeker, J.; Saeys, W. Task and motion planning for apple harvesting robot. *IFAC Proceedings Volumes*, v.46, p.247-252, 2013. <https://doi.org/10.3182/20130828-2-SF-3019.00063>
- Ouf, N. S. Leguminous seeds detection based on convolutional neural networks: Comparison of faster R-CNN and YOLOv4 on a small custom dataset. *Artificial Intelligence in Agriculture*, v.8, p.30-45, 2023. <https://doi.org/10.1016/j.aiaa.2023.03.002>
- Reig, G.; Lordan, J.; Sazo, M. M.; Hoying, S.; Fargione, M.; Reginato, G.; Donahue, D. J.; Francescato, P.; Fazio, G.; Robinsin, T. Long-term performance of 'Gala', 'Fuji' and 'Honeycrisp' apple trees grafted on Geneva® rootstocks and trained to four production systems under New York State climatic conditions. *Scientia Horticulturae*, v.244, p.277-293, 2019. <https://doi.org/10.1016/j.scienta.2018.09.025>
- Schertz, C. E.; Brown, G. K. Basic considerations in mechanizing citrus harvest. *Transactions of the ASAE*, v.11, p.343-346, 1968. <https://doi.org/10.13031/2013.39405>
- Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, v.157, p.417-426, 2019. <https://doi.org/10.1016/j.compag.2019.01.012>
- Wang, D.; He, D. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, v.210, p.271-281, 2021. <https://doi.org/10.1016/j.biosystemseng.2021.08.015>
- Xiong, Y.; Ge, Y.; Grimstad, L.; From, P. J. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. *Journal of Field Robotics*, v.37, p.202-224, 2020. <https://doi.org/10.1002/rob.21889>
- Xu, J.; Yao, J.; Zhai, H.; Li, Q.; Xu, Q.; Xiang, Y.; Liu, Y.; Liu, T. Ma, H.; Mao, Y.; Wu, F.; Wang, Q.; Feng, X.; Mu, J.; Lu, Y. Trichome YOLO: A Neural Network for Automatic Maize Trichome Counting. *Plant Phenomics*, v.5, p.24-35, 2023. <https://doi.org/10.34133/plantphenomics.0024>
- Yang, Q.; Chen, C.; Dai, J.; Xun, Y.; Bao, G. Tracking and recognition algorithm for a robot harvesting oscillating apples. *International Journal of Agricultural and Biological Engineering*, v.13, p.163-170, 2020. <https://doi.org/10.25165/ij.ijabe.20201305.5520>
- 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). *Biosystems Engineering*, v.193, p.247-263, 2020. <https://doi.org/10.1016/j.biosystemseng.2020.03.006>
- Zhang, Z.; Heinemann, P. H.; Liu, J.; Baugher, T. A.; Schupp, J. R. The development of mechanical apple harvesting technology: A review. *Transactions of the ASABE*, v.59, p.1165-1180, 2016. <https://doi.org/10.13031/trans.59.11737>
- Zhong, Y.; Fei, L.; Li, Y.; Zeng, J.; Dai, Z. Response of fruit yield, fruit quality, and water use efficiency to water deficits for apple trees under surge-root irrigation in the Loess Plateau of China. *Agricultural Water Management*, v.222, p.221-230, 2019. <https://doi.org/10.1016/j.agwat.2019.05.035>