OBSTACLE AVOIDANCE PLANNING OF GRAPE PICKING ROBOTS BASED ON DEEP REINFORCEMENT LEARNING
/
基于深度强化学习的葡萄采摘机器人采摘路径避障规划

Pei LIU*)
School of Economics and Management, Wuhan Railway Vocational College of Technology, Wuhan, Hubei/China
Tel: +8613871280920; E-mail: wh20090803@126.com
Corresponding author: Pei LIU
DOI: https://doi.org/10.35633/inmateh-74-70

Keywords: *deep reinforcement learning; grape; obstacle avoidance; path planning; picking robot*

ABSTRACT
Given that picking robots are faced with many picking tasks in the field operation environment and the target and obstacles are located at random and uncertain positions, an obstacle avoidance planning method for the picking path of virtual robots based on deep reinforcement learning was proposed to achieve rapid route planning of robots under a lot of uncertain tasks. Next, the random motion strategy of virtual robots was set according to the physical structure of robot bodies. By comparatively analyzing the advantages and disadvantages of the observed values input by different networks, an environmental observation set was established in combination with actual picking behaviors as the network input; then, a reward function was established by introducing the idea of target attraction and obstacle repulsion contained in the artificial potential field method, aiming to evaluate the behavior of virtual robots and increase the success rate of obstacle avoidance. The results of the simulation experiment showed that the success rate obtained by virtual robots in completing the picking task reached 95.5% under obstacles set at different positions. The coverage path length of the deep reinforcement learning algorithm was reduced by 272.79 in compared with that of genetic algorithm, with a reduction rate of 5.09%. The total time consumed by navigation was 1549.24 s, which was 83.15 s shorter than that of the traditional algorithm. The study results manifest that the system can efficiently guide virtual robots to rapidly reach the random picking points on the premise of avoiding obstacles, meet picking task requirements and provide theoretical and technical support for the picking path planning of real robots.

摘要
针对采摘机器人在野外作业环境中，面临采摘任务数量多，目标与障碍物位置具有随机性和不确定性等问题，提出一种基于深度强化学习的虚拟机器人采摘路径避障规划方法，实现机器人在大量且不确定任务情况下的快速轨迹规划。根据机器人本体物理结构设定虚拟机器人随机运动策略，通过对比分析不同网络输入观测值的优劣，结合实际采摘行为设置环境观测集合，作为网络的输入;引入人工势场法目标吸引和障碍排斥的思想建立奖惩函数，对虚拟机器人行为进行评价，提高避障成功率。仿真实验结果显示，不同位置障碍物设置情况下虚拟机器人完成采摘任务成功率达 95.5%，深度强化学习算法覆盖路径长度相比于遗传算法减少了 272.79in，缩短了 5.09%，整体导航用时 1549.24s，相比于传统算法缩短了 83.15s。研究结果表明，本系统能够高效引导虚拟机器人在避开障碍物的前提下快速到达随机采摘点，满足采摘任务要求，为真实机器人采摘路径规划提供理论与技术支撑。

INTRODUCTION
With the growth of global population and the acceleration of urbanization, there is an increasing shortage of agricultural labor force and agricultural production is faced with severe challenges. Meanwhile, consumers have increasingly higher quality and safety requirements for agricultural products, facilitating agricultural production to develop towards intelligent and automatic directions. Emerging rightly under this background, picking robots are the product of combining modern agricultural technologies and robot technologies, aiming to solve such problems as the shortage of agricultural labor force, the improvement of production efficiency and the quality guarantee of agricultural products *(Ma et al., 2024)*. Picking robots are capable of picking operations continuously and stably and immune to the influence of human fatigue and emotions, thus significantly improving production efficiency.

Compared with manual picking, picking robots can complete the picking task faster, reduce the picking cycle and improve the output of agricultural products *(Cong et al., 2024)*. With the shortage of agricultural labor force, labor costs are rising. If applied, picking robots can effectively reduce the dependence on manual labor and reduce labor costs. At the same time, picking robots can also avoid problems such as production delay and output decline caused by labor shortage *(Cao et al., 2023)*.

China is one of the important fruit producers in the world, with its main fruit varieties ranking first in the world since 2012. The labor required for fruit picking accounts for about 40% of the labor required for the whole production process, and the efficiency and quality of fruit picking greatly affect the economic benefits of fruit *(Zhao et al., 2023)*. As population aging is accelerated, the labor cost is getting higher and higher, and the economic and efficient market requirements cannot be satisfied just by relying upon manual picking. Agricultural automatic picking robots can effectively replace manual picking, and solve labor shortage to a great extent *(Xu and* Zhou*., 2023)*. In the unstructured working environment, how to make the picking robot reach the picking point accurately and quickly and complete the picking task on the premise of ensuring safe operation is a difficulty in the field of manipulator control and also the key to determining whether robots can pick fruit successfully and efficiently. Therefore, the path planning of picking robots is very important.

Literature review on the field of picking robots (especially agricultural robots) mainly includes vision system, path planning, robot control, task scheduling, crop identification, force control and mechanical design. The following are the references and reviews in some key fields, covering the technologies, challenges and latest progress related to picking robots.

In terms of robot path planning methods, such algorithms as A-Star (A*) algorithm *(Lehmann and Fendt, 2020)*, ant colony optimization algorithm *(Wang and Zhang, 2021)*, genetic algorithm (San et al., 2018) and artificial potential field method *(Liu and He, 2022)* had been widely used in vehicle navigation and path planning research of mobile robots. In the aspect of manipulators, the above algorithms had been mostly used to study the planning problem of low-degree-of-freedom (DOF) manipulators, with some defects such as slow convergence and low efficiency *(Zhou and Yu, 2023)*. In order to solve the path planning problem of multi-DOF manipulators in high-dimensional space, *Sun et al. (2020)* put forward the rapidly-exploring random trees (RRT) algorithm, which avoids space modeling of the traditional method through the collision detection of sampling points in the state space. This algorithm, which could effectively solve the path planning problem in high-dimensional space under complex constraints, had been extensively applied to mobile robot path planning and multi-DOF manipulator path planning *(Sun et al., 2020)*. Considering the disadvantages of the RRT algorithm, *Gupta and Shankar (2022)* made lots of improvements of this algorithm in optimizing the algorithm path and accelerating the search speed *(Gupta et al., 2022)*. As for the optimization of the algorithm path, the progressively optimal RRT* algorithm and its improved algorithm are the representatives, but the quality of the planned path is improved at the sacrifice of greatly increasing time consumption. In the aspect of accelerating the search speed, *Li and Chen (2022)* proposed the bidirectional random search tree (BI-RRT) algorithm. On the basis of RRT algorithm, a double-tree extension link was introduced, and two random trees were generated at the same time from the starting point and the end point to search, which greatly reduced the search time. However, this algorithm adopted the idea of random node extension of RRT algorithm, thus having the disadvantage of no goal-oriented configuration. *Nof and Sgobbi (2022)* put forward the RRT-connect algorithm, which connected two random search trees of BI-RRT algorithm through greedy search strategy, reduced the number of sampling nodes and accelerated the convergence speed of the algorithm. Wei et al. improved the RRT-connect algorithm, which further accelerated the search speed. *Chen and Zhang (2021)* introduced the concept of target gravity and the adaptive parameter adjustment method, and proposed the AtBi-RRT algorithm based on double RRT algorithms, which realized the fast collision-free motion planning of litchi picking.

In the research on agricultural robots, *Luo and Wei (2015)* introduced the background of agricultural robot technology, covering various applications from crop monitoring to automatic picking, and discussed how unmanned aerial vehicles, mobile robots and other technologies can improve efficiency in agricultural production. *Li et al. (2011)* summarized the progress of fruit picking robot technology, focusing on crop identification, manipulator design and picking strategy, especially their application in automatic picking. *Nguyen and Liu (2020)* proposed the vision-based crop identification method, which covered the application of deep learning, image processing, stereo vision and other technologies, and analyzed the advantages and disadvantages of different fruit identification methods. *Mohanan and Salgoankar (2018)* reviewed the application of machine vision in precision agriculture, especially the key technologies in crop monitoring and picking, including visual sensors, image processing technology and deep learning model.

Jiang and Zhao (2020) discussed the path planning of agricultural robots in complex farmland environment, especially the challenges in unstructured environments, such as obstacle avoidance and optimal path search. *Hu et al. (2014)* introduced the route planning method of agricultural robots in a dynamic environment, considering obstacles and the spatial limitation between different crops as well as the optimal picking path design. *Devaurs et al. (2016)* systematically reviewed the design and control strategy of the manipulator of the picking robot, covering the aspects of force control, grasping and releasing mechanism, adaptive picking of crops and so on.

Ali and Silva (2019) discussed a variety of robot mechanical designs suitable for picking tasks, especially the selection of end effectors, force control strategies and design methods suitable for different crops. *Cui et al. (2016)* discussed the potential of multi-robot cooperation in precision agriculture, especially their application in automatic picking tasks, and analyzed the challenges and opportunities faced by multi-robot cooperation.

Xie and Liu (2019) analyzed the latest research on multi-robot cooperative picking in greenhouse environment, focusing on distributed task scheduling, cooperative path planning and coordination mechanism among robots. *Bac et al. (2014)* discussed the application of deep learning in agricultural robots, with the emphasis laid on the latest research results in the fields of crop detection, automatic picking and image processing based on deep neural networks. *Luo et al. (2016)* discussed the application of artificial intelligence in agricultural picking robots, especially in crop identification, path planning and decision-making systems. Future trends include multimodal learning and adaptive control.

Hernandez and Calderon (2018) discussed the special challenges faced by picking robots for soft fruits (such as strawberries and blueberries), especially the research progress in visual recognition, tactile perception and force control. *Brogli and Rucker (2017)* explored the future development direction of agricultural robot technology, including automatic picking, cooperative robot system, intelligent crop identification and the combination of robots and the Internet of Things. These documents cover a number of key technologies in the field of picking robots, from basic visual perception to complex path planning and cooperation tasks, to the application and challenges of deep learning, and each document is helpful to understand the technological breakthroughs and development trends of agricultural robots in different application scenarios. Agricultural picking robots can not only improve production efficiency but also play an important role in coping with labor shortage and improving crop yield and quality.

Aiming at the above problems, an obstacle avoidance planning system for the picking path of virtual robots based on reinforcement learning was designed with four-DOF picking robots as the study objects. First, the random motion strategy of virtual robots was set according to the physical structure of robots, and environmental observed quantity was reasonably set as the network input by analyzing the actual picking behavior. Then, a reward function was established by introducing the idea of target attraction and obstacle repulsion of the artificial potential field method. Given that the path planning is affected by the range repulsion of the artificial potential field method, a directional penalty obstacle avoidance function was proposed through the motion collision analysis of virtual robots, and the behavior of virtual robots was evaluated, guiding virtual robots to reach the target picking point as soon as possible on the premise of evading obstacles. Finally, the interactive communication between simulation environment and reinforcement learning was established using machine learning agents (ML-Agents) module, and the virtual robots were subjected to picking training via the distributed proximal policy optimization (DPPO) algorithm, expecting to realize the intelligent obstacle avoidance path planning of picking robots.

MATERIALS AND METHODS

In the design of the grape picking robot, the operation was completed in cooperation with the mobile platform under the single-arm picking mode. The configuration design of the 4-DOF manipulator was completed based on picking operation scenes. The picking environment and overall structure of robots are exhibited in Fig. 1.

In order to meet the collection operation needs of grape picking robots, a detachable grape grid collection box was designed. The detachable structure design was convenient for the robot to replace the collection box at the collection point after full loading; the grid structure design could not only avoid the contact damage of fruit stacking in ordinary collection boxes but also provide the location information of collection points for the path planning of collection operations, so as to realize the pre-planning of collection operation paths and save the collection path planning time.

Fig. 1- Schematic diagram of robot picking environment and overall structure

Kinematic model of the manipulator

In the structural design of the picking manipulator, the coordinates of the full-rotation chain-like robot were configured, and the manipulator was completely positioned in the working space via the 4 DOF. Next, the connecting rod coordinate system of the manipulator was established through the D-H parameter method, and its D-H parameters are listed in Table 1.

Table 1

Manipulator D-H parameters

#	θ/(°)	d/mm	a/cm	ɑ/cm
0-1	θ_1	0	0	90
1-2	θ_2	0	270	0
2-3	θ_3	0	180	0
3-4	θ_4	0	180	90

In the forward kinematics solving of the manipulator, the homogeneous transformation matrix of the terminal of the picking manipulator relative to the base coordinate system is ${}^0T_H$,the formula is shown in (1).

$$ {}^0T_H = {}^0T_1 {}^0T_2 {}^0T_3 {}^0T_4 \tag{1} $$

Each row of the D-H parameter table is substituted into Equation (1) to obtain the rotational transformation matrix ${}^nT_{n+1}$ from the previous joint coordinate system $\{n\}$ to the next joint coordinate system $\{n+1\}$,the formula is shown in (2).

$$ {}^nT_{n+1} = Rot(z,\theta_{n+1}) \times Trans(z,d_{n+1}) \times Trans(x,a_{n+1}) \times Po\tau(x,\alpha_{n+1}) = \begin{bmatrix} C\theta_{n+1} & -S\theta_{n+1}C\alpha_{n+1} & S\theta_{n+1}S\alpha_{n+1} & a_{n+1}C\theta_{n+1} \\ S\theta_{n+1} & C\theta_{n+1}C\alpha_{n+1} & -C\theta_{n+1}S\alpha_{n+1} & a_{n+1}S\theta_{n+1} \\ 0 & S\alpha_{n+1} & C\alpha_{n+1} & d_{n+1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2} $$

The 4-DOF manipulator was designed based on the Pieper criterion, closed-form solutions exist in inverse kinematics, so the inverse kinematic solution of the manipulator was solved through the analytical method. The obstacle avoidance path planning algorithm of grape picking robots needs to detect the collision information between the manipulator and obstacles. The spatial position and pose of such obstacles as grape vines and leaves coming down in the orchard environment will change easily due to the disturbance of the external environment. To better encircle obstacles, the OBB bounding box with strong tightness, good real-timeliness and random direction was adopted.

Terminal operation path planning

The research on the terminal operation path planning of grape picking robots is mainly divided into the following two parts: The first is the picking operation path planning. According to the known position information, the obstacle-avoidance picking configuration of the manipulator from the initial position to the target position is calculated, and if a group of configurations can realize obstacle-avoidance picking, the obstacle-avoidance path is planned by the deep reinforcement learning algorithm to complete the picking operation. In case that there is no group of configurations that can realize obstacle avoidance picking or the obstacle avoidance path planning fails, the flexible obstacle avoidance path is planned by the flexible obstacle avoidance strategy to complete the picking operation; if the flexible obstacle avoidance path planning fails, the position of the mobile platform is adjusted and the above operations are repeated at the new position. After picking, robots will return to the original position along the planned path. The second collection operation path planning. The coordinate information of the grape collection point position and the initial position of the manipulator are known. The path of the manipulator from the initial position to the placement position of each collection point is planned in advance by the deep reinforcement learning algorithm, and the collection paths such as b1, b2 and b3 are obtained. After the collection is completed, robots will return to the original position along the planned path. Among them, obstacle avoidance path planning refers to the collision detection of the whole connecting rod of the manipulator, so as to obtain the obstacle avoidance path along which the manipulator will not collide with obstacles during the whole operation. Flexible obstacle avoidance path planning refers to the flexible obstacle avoidance path through which the connecting rod of the manipulator can push aside the flexible obstacles such as vines and leaves laterally and complete the picking operation. In the collection operation, if the traditional collection box is used, it is necessary to identify and locate the grape placement point and conduct path planning once again in each collection operation; if the grid collection box with known location information of collection points is used, time can be greatly saved by directly calling the collection path planned in advance during operation.

Guided reward mechanism design

After initializing actions and states, agents can randomly draw different action strategies according to states, but they fail to evaluate the quality of actions according to states. Designing the guided reward function can evaluate agents' behavior, increase the probability of high-scoring behavior and reduce the probability of low-scoring behavior, and then guide the agents to make correct actions in various environmental states. The reward mechanism determines the effect of training results. A reasonably designed reward function can improve the training speed, reduce the consumption of computer resources and make the training results converge faster. In most cases, continuous reward and penalty information can continuously let the agent get feedback on the action strategy adopted, which is more effective than sparse reward and penalty signals. In this study, a continuous reward function was constructed based on the deep reinforcement learning algorithm, and the agents were guided to reach the target correctly by designing the guidance function and obstacle avoidance function. At the same time, a time function is set to guide the agents to complete the task faster.

When the gravity is set using the traditional artificial potential field method, its value is determined by the current position of the object; when setting the reward function, gravity is represented by the reward signal, which should be determined by the action of the agent. When the terminal actuator cutting point is close to the target point due to the change in the agent's action, a reward will be given and is directly proportional to the approaching distance, otherwise, a penalty will be posed. Under a state st, a certain distance exists between the terminal actuator cutting point pend=(xst, yst, zst) of the virtual robot and the target picking point pgoal=(x0, y0, z0), which is expressed by the target distance Dsi. If this distance is continuously reduced during the training process, the action strategy of the robot is correct and this behavior should be rewarded. The formula for setting the guidance function is shown in (3)-(6):

$$D_{s_t} = \sqrt{(x_{s_t} - x_0)^2 + (y_{s_t} - y_0)^2 + (z_{s_t} - z_0)^2} \tag{3}$$

$$D_{\min} = \begin{cases} D_{s_0} & (i = 0) \\ \min(D_{s_i}, D_{\min}) & (i \neq 0) \end{cases} \tag{4}$$

$$R_{\text{goal}} = \begin{cases} (D_{\min} - D_{s_t})k_1 & (D_{s_t} \neq 0) \\ k_2 & (D_{s_t} = 0) \end{cases} \tag{5}$$

where $D_{\min}$ is the minimum value of the target distance in one picking round; D_{s_0} represents the initial value of $D_{\min}$ under initial state of the picking round; R_{goal} is the reward & penalty value of the target distance; k_1 and k_2 are constants.

As the agent does random actions, D_{s_t} changes, and the minimum value between the two is taken for $D_{\min}$ according to Equation (4). When the target distance is shortened, a low reward will be given according to the distance shortened, otherwise, a low penalty will be posed; when the target distance is 0, the robot correctly reaches the target picking point, a high reward will be given, and this round will be ended.

Setting of obstacle avoidance function

When the repulsive force is set by the deep reinforcement learning algorithm, as long as the object enters the potential field of the obstacle, it will be affected by the repulsive force, and the influence is sometimes unnecessary and will affect the planning of the shortest path. As shown in Figure 5, when the obstacle avoidance function is set based on the idea of the artificial potential field method, the repulsive force is represented by a penalty signal. In this case, the shortest path for the virtual robot to reach the target point is a straight line, but its path is deviated due to the influence of the potential field of the obstacle. In order to solve this problem, the spatial motion of the virtual robot was divided into 3 parts: leftward and rightward motion along x axis, back and forth motion along y axis and upward and downward motion along z axis. The range penalty of the obstacle was divided into the penalty of the above three directions of motion, and whether the movement of each rotating shaft of the virtual robot along the x, y and z axes in the current posture might collide with obstacles was judged. If it was possible, the penalty in this direction would be given, otherwise, the penalty in this direction would not be posed, so as to avoid unnecessary penalty on the virtual robot in the correct collision-free movement.

According to the physical structure of the picking robot, each rotating joint shaft of the virtual robot is regarded as a cylinder for the analysis. The random motion strategy of the virtual robot is divided into the motion along axes x, y and z. Given the similarity of motion collision analysis in the directions of the 3 axes, only the upward and downward motion of the rotating joint shaft 4 along axis z was subjected to the collision analysis, and other motion conditions could be obtained in a similar way. Assuming that the current coordinates of the virtual robot's rotating joint shaft 4 are $(x_q,\ y_q,\ z_q)$ and those of the obstacle are $(x_o,\ y_o,\ z_o)$, it is necessary to judge whether the rotating shaft 4 will collide with the obstacle during upward and downward motion so as to determine whether to give the penalty. First, the z axis of this rotating shaft experiences translation to a position consistent with the z axis of the obstacle, i.e., the coordinates of this rotating shaft are set to $(x_q,\ y_q,\ z_q)$. In this case, the two are located on $x-y$ plane, as shown in Fig. 2.

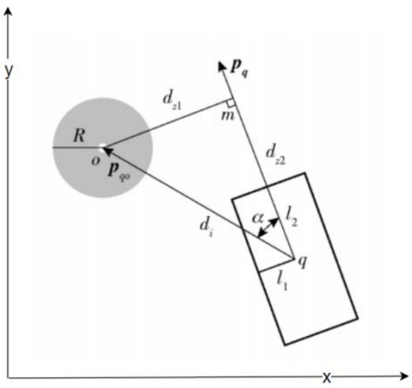

Fig. 2 - Collision analysis of upward and downward motion direction of the virtual robot's rotating shaft

In Figure 2, o represents the obstacle midpoint; R stands for the radius of the obstacle; q is the critical point of the virtual robot's rotating shaft; l_1 is the distance from q to the horizontal cylindrical surface, i.e., the radius of the cylinder; l_2 is 1/2 of the cylinder length; vector p_{qo} denotes the direction vector pointing to the obstacle midpoint o with q as the starting point; vector pq is the current forward direction vector of the virtual picking robot's rotating shaft; α is the included angle between vectors p_{qo} and pq ; d_i is the distance between points and o, which can be calculated through coordinates; d_{z1} represents the horizontal distance between points o and q ; d_{z2} is the longitudinal distance between points o and q . It can be known from Figure 6 that d_{z1} and d_{z2} can be solved based on d_i, as follows (6):

$$\begin{cases} d_{z1} = d_i \sin\alpha \\ d_{z2} = d_i \cos\alpha \quad (\alpha > \pi / 2) \\ \alpha = \pi - \alpha \end{cases} \tag{6}$$

When the horizontal distance between points o and satisfies $d_{z1} \le R + \mathrm{l}_1$ and their longitudinal distance meets $d_{z2} \le R + \mathrm{l}_2$, the rotating shaft will probably collide with the obstacle only during upward and downward motion. Hence, the obstacle avoidance coefficient k_z in z direction is set.

When the above collision conditions are met, k_z is 1, otherwise, it is 0. The obstacle avoidance function in the upward and downward motion direction of this rotating shaft is expressed as below (7):

$$R_{\text{obs}-z} = k_{zi}\frac{1}{d_{zi}} \tag{7}$$

Similarly, the obstacle avoidance coefficient and function in left-right and back-forth motion directions are set. The obstacle avoidance coefficient and the overall obstacle avoidance function can be expressed as follows (8)-(9):

$$\begin{cases} k_{xi} = \begin{cases} 1,(d_{xi} \leqslant R + l_{i1}\ \&\&d_{xi} \leqslant R + l_{i2}) \\ 0,\text{else} \end{cases} \\ k_{yi} = \begin{cases} 1,(d_{yi} \leqslant R + l_{i1}) \\ 0,(d_{yi} > R + l_{i1}) \end{cases} \\ k_{zi} = \begin{cases} 1,(d_{zi} \leqslant R + l_{i1}\ \&\&d_{zi} \leqslant R + l_{i2}) \\ 0,\text{else} \end{cases} \end{cases} \tag{8}$$

$$\begin{cases} \sum_{i=1}^{4} k_{xi}\frac{1}{d_{xi}} + k_{yi}\frac{1}{d_{yi}} + k_{zi}\frac{1}{d_{zi}} & (H \cap O_{\text{obs}} = \varnothing) \\ k_{\text{obs}} & (H \cap O_{\text{obs}} \neq \varnothing) \end{cases} \tag{9}$$

where d_{xi} is the distance between the x-coordinate of the critical point on the virtual robot's rotating shaft i and the x-coordinate of the obstacle, i.e., the left-right distance; d_{yi} represents the longitudinal distance; l_{i1} and l_{i2} stand for the cylinder radius of manipulator i and 1/2 of the cylinder length, respectively; k_{obs} is a collision penalty constant; HO_{obs} is the space set of the virtual robot and obstacle.

Equation (9) takes effect only when the virtual robot enters the range of obstacle repulsion, and its penalty increases as the distance between the critical point on each rotating shaft of the virtual robot and the characteristic direction of the obstacle decreases; in case of robot-obstacle collision, a high penalty will be posed and this round will be ended.

Setting of time function

The time penalty function R_p is set according to the distance traveled under the initial state, calculated as follows (10):

$$R_p = \frac{k_t}{D_{s_0}} \tag{10}$$

where k_t is the time penalty constant.

The designed total reward function is the cumulative sum of the above 3 functions, which can be expressed as bellows (11):

$$R = R_{goal} + R_{obs} + R_p \tag{11}$$

RESULTS

In order to verify the effectiveness of the algorithm, a simulation experiment was performed on the MATLAB platform. After establishing the forward and inverse kinematics models of the manipulator, the position of the manipulator joint 1 was set to (0,0,0), and the initial position configuration of the manipulator was set to (0°, -60°, 120°, 0°), which ensured the flexibility of bidirectional movement to the target position of the fruit stalk picking point and the collection position of the collection box. A deep reinforcement learning algorithm was compiled, and the step size was set to 2°, and the path planning simulation experiment was carried out in different environments.

Picking path planning of deep reinforcement learning algorithm

The simulation experiment scene was set as a 700in×700in operation area of the grape picking robot. See Fig. 3 for the specific distribution of picking points in the operation area. The path planning problem of multiple picking points in obstacle-free space was solved using the deep reinforcement learning algorithm, the grape picking task was completed based on this algorithm, and the convergence curve of the algorithm is displayed in Fig. 4. The grape picking robot's path planning based on the deep reinforcement learning algorithm is exhibited in Fig. 5. From the convergence of the fitness curve of 49 task points in Figure 2, the optimal traversal path was found through the 132nd epoch of the deep reinforcement learning algorithm. The deep reinforcement learning algorithm achieved a relatively good convergence effect, indicating its strong model solving ability.

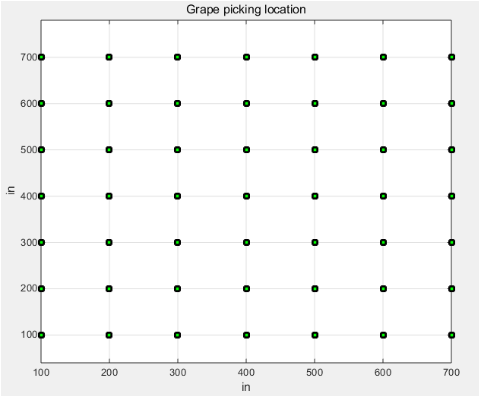

Fig. 3 - Length of operation area and position coordinates of picking points (m)

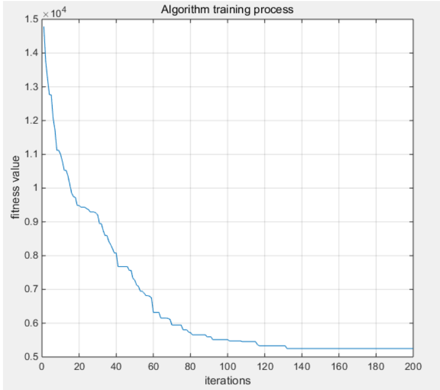

Fig. 4 - Convergence curve of deep reinforcement learning algorithm

Fig. 5 - Path planned by deep reinforcement learning algorithm

Algorithm verification

To explore the stability of the algorithm and verify the effectiveness of the deep reinforcement learning algorithm, the traditional genetic algorithm solving model was also designed in this study. The number of grape picking points was 49, the minimum value of two-dimensional coordinates was 100in, the maximum value of two-dimensional coordinates was 700in, the maximum number of epochs in genetic algorithm parameters was popsize (population size) =100, the tournament size was tournament_size=5, the crossover probability was pc=0.95, and mutation probability was pm=0.1. Model solving was performed to obtain the path diagram of the traditional genetic algorithm, as shown in Fig. 6 The convergence curve of the traditional genetic algorithm is displayed in Fig. 7.

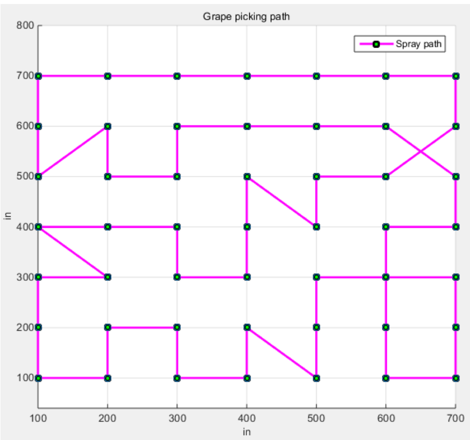

Fig. 6 - Path diagram of traditional genetic algorithm

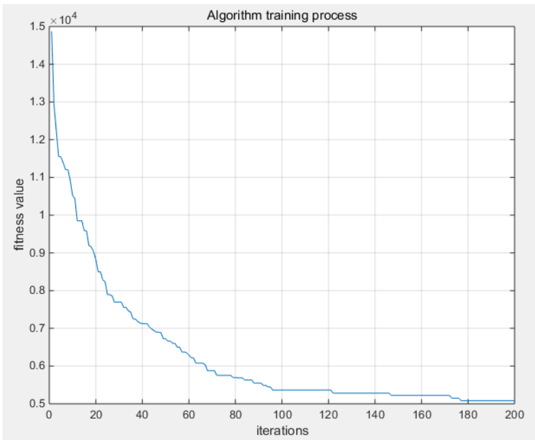

Fig. 7 - Convergence curve of traditional genetic algorithm

It could be observed that compared with the traditional genetic algorithm, the deep reinforcement learning algorithm was more exploratory and convergent, accompanied by the better value of the objective function.

Besides, the deep reinforcement learning algorithm performed better than the traditional genetic algorithm in the number of inflection points, convergence time and convergence algebra. In this study, an improved deep reinforcement learning algorithm was proposed to solve the traversal order of the grape picking robot in each picking area. First, a path quality evaluation function was established, the length of the coverage path was taken as an evaluation index to comprehensively score the coverage path planned in each round, and this score was used as the criterion for the subsequent reward assignment. Then, to enhance the global correlation of the model, the action value in each round was updated on the whole by setting an empirical repository, so as to accelerate the model learning efficiency in the initial stage. Finally, an empirical backtracking mechanism was established, and high-score paths were selected for empirical backtracking to enhance their positive guiding effect on the model. The model training results revealed that the deep reinforcement learning algorithm could complete convergence faster at the relatively optimal solution than the traditional genetic algorithm. Additionally, the navigation test results in the grape orchard manifested that the length of the coverage path obtained by the proposed method was reduced by 272.79in compared with the traditional genetic algorithm, with a reduction rate of 5.09%, and the total time consumption by navigation was 1549.24 s, which was 83.15 s shorter than that consumed by the traditional algorithm. This indicates that the coverage path planning method raised in this study can effectively shorten the length of the grape picking robot's coverage length and improve the navigation efficiency of the grape picking robot.

CONCLUSIONS

Given the large number of picking tasks in orchards and the highly random fruit distribution, a set of virtual obstacle positions based on deep reinforcement learning were set for the sake of real-time efficient path planning. The results of the simulation experiment showed that the picking success rate reached above 95.5%. Besides, a reward function setting method was put forward, i.e., introducing the idea of target attraction and obstacle repulsion of the artificial potential field method, aiming to increase the success rate of obstacle avoidance during picking. Considering that the shortest path planning was affected by the range repulsion of the artificial potential field method, a directional penalty obstacle avoidance function and a robot collision analysis model were established, facilitating the robot to make correct decisions and prevent unnecessary penalties and improving the picking efficiency. To verify the effectiveness of the proposed method, the picking performance contrast experiments under different reward functions were designed. The results manifested that the length of the coverage length was 272.79in, which was 5.09% shorter than that obtained by the genetic algorithm, and the total time consumption by navigation was 1549.24 s, which was 83.15 s shorter than that obtained by the traditional algorithm. The model and method involved in this paper can quickly complete the path planning of grape picking tasks under various conditions, but there is still small probability of picking failure under harsh picking conditions, which has certain shortcomings. In the follow-up study, the picking situation under special conditions will be processed, and the path planning in the case of irregular shape obstacles will be discussed to further improve the robustness of the system.

REFERENCES

[1] Ali, H., & De Silva, C.W. (2019). Path Planning for Agricultural Robots: Algorithms, Challenges, and Applications. *IEEE International Conference on Robotics and Automation* (ICRA), pp. 1234-1241, Canada.

[2] Bac, C.W., Hemming, J., & Van Henten, E.J. (2014). Stem localization of sweet-pepper plants using the support wire as a visual cue. *Computers and Electronics in Agriculture*, Vol. 105, pp. 111–120, United Kingdom.

[3] Brogli, R., & Rucker, D. (2017). Robots in Agriculture: A Review of Current Applications and Future Opportunities. *Precision Agriculture*, Vol. 18, pp. 471-486, Netherlands.

[4] Cao, X. M., Wu, B., & Chen, H., Chen, C.Y., Huang, X. (2023). Path planning for a six degree of freedom harvesting robot (六自由度采摘机器人路径规划). *China Science and Technology Information*, Vol.21, pp. 87-90, Beijing/China.

[5] Chen, Y., & Zhang, J. (2021). Mechanical Design and Control of Robotic Systems for Agricultural Harvesting Tasks. *Frontiers in Robotics and AI*, Vol. 8, pp. 63, Switzerland.

[6] Cong, C., Liang, Y. Q., & Liu, J. D., Chen, H.Y. (2024). Full Automatic Picking Robot for Soft Date Kiwifruit (软枣猕猴桃全自动采摘机器人设计). *Journal of Jiamusi University (Natural Science Edition)*, Vol. 41, pp. 106-109, Jiamusi/China.

[7] Cui, R., Li, Y., & Yan, W. (2016). Mutual information-based multi-AUV path planning for scalar field sampling using multidimensional RRT. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, Vol.46, pp. 993–1004, United States.
[8] Devaurs, D., Siméon, T., & Cortés, J. (2016). Optimal Path Planning in Complex Cost Spaces with Sampling-Based Algorithms. *IEEE Transactions on Automation Science and Engineering*, Vol.13, pp. 415-424, United States.
[9] Gupta, S., & Shankar, A. (2022). Artificial Intelligence for Harvesting Robots: A Comprehensive Review of AI-Based Approaches and Future Trends. *Artificial Intelligence in Agriculture*, Vol.5, pp. 88-105, Beijing/China.
[10] Hernandez, J., & Calderon, R. (2018). Robotic Harvesting of Fruit Crops: A Review of the State of the Art and Future Directions. *Robotics*, Vol. 7, pp. 46, Switzerland.
[11] Jiang, H., & Zhao, X. (2020). Robotic Harvesting Mechanisms: A Comprehensive Review of Designs, Control Strategies, and Applications. *Advanced Robotics*, Vol. 34, pp. 325-343, Japan.
[12] Lehmann, A., & Fendt, J. (2020). Multi-Robot Systems in Precision Agriculture: Opportunities and Challenges for Automated Harvesting. *Journal of Field Robotics*, Vol.37, pp. 1169-1185, United States.
[13] Li, B. H.， Chen, B. D. (2022). An adaptive rapidly-exploring random tree. *IEEE/CAA Journal of Automatica Sinica*, Vol.9, pp. 283-294, Beijing/China.
[14] Li, P., Lee, S. H., & Hsu, H. Y. (2011). Review on fruit harvesting method for potential use of automatic fruit harvesting systems. *Proceedings of the 2011 International Conference on Power Electronics and Engineering Application*, Vol. 12, pp. 351–366, Shanghai/China.
[15] Liu, Z., & He, Z. (2022). Collaborative Harvesting Using Multiple Robots in Greenhouses: A Survey. *International Journal of Advanced Robotics Systems*, Vol. 19, pp. 102, Austria.
[16] Luo, H., & Wei, G. (2015). System design and implementation of a novel robot for apple harvest. *INMATEH-Agricultural Engineering*, Vol. 46, pp. 85-94, Romania.
[17] Luo, L., Tang, Y., Zou, X., Ye, M., Feng, W., & Li, G. (2016). Vision-based extraction of spatial information in grape clusters for harvesting robots. *Biosystems Engineering*, Vol. 151, pp. 90–104, United Kingdom.
[18] Ma, X.J., Shi, X.Y., & Xiao, W.X., Ren, M.T., Bao, M.L. (2024). Multi-objective continuous picking path planning for citrus based on TSP__RRT algorithm (基于 TSP__RRT 算法的柑橘多目标连续采摘路径规划). *Acta Agriculturae Universitatis Jiangxiensis*, Vol. 46, pp. 490-501, Nanchang/China.
[19] Mohanan, M. G. & Salgoankar, A. (2018). A survey of robotic motion planning in dynamic environments. *Robotics and Autonomous Systems*, Vol. 100, pp. 171-185, Netherlands.
[20] Nguyen, T., & Liu, W. (2020). Trajectory Planning for Autonomous Agricultural Robots in Dynamic Environments. *Journal of Field Robotics*, Vol.37, pp. 955-973, United States.
[21] Nof, S. Y., & Sgobbi, F. (2022). Challenges in Robotic Harvesting of Soft Fruits: A Survey on Perception, Control, and Automation. *Robotics and Autonomous Systems*, Vol. 134, pp. 103617, Netherlands.
[22] San Juan, V., Santos, M., & Andújar J. M. (2018). Intelligent UAV Map Generation and Discrete Path Planning for Search and Rescue Operations. *Complexity*, Vol.17, pp. 6879419, United States.
[23] Sun, Y. X., Zhang, C. R., Sun, P. C. (2020). Safe and smooth motion planning for mecanum-wheeled robot using improved RRT and cubic spline. *Arabian Journal for Science and Engineering*, Vol.45, pp. 3075-3090, Saudi Arabia.
[24] Wang, J., & Zhang, Y. (2021). Deep Learning for Agricultural Robotics: Applications in Harvesting, Monitoring, and Control. *Computers and Electronics in Agriculture*, Vol.183, pp.105964, United Kingdom.
[25] Wei, W., Ouyang, D. T., Lu, S., & Feng Y.X. (2011). Multiobjective path planning under dynamic uncertain environment. *Jisuanji Xuebao/Chinese Journal of Computers*, Vol. 34, pp. 836–845, Beijing/China.
[26] Xie, L., & Liu, H. (2019). A Review of Vision-Based Approaches for Fruit Detection and Localization in Robotic Harvesting. *Computers and Electronics in Agriculture*, Vol.158, pp. 49-61, United Kingdom.
[27] Xu, N., & Zhou, W.M. Research on Global Coverage Path Planning of Picking Robot Based on Adaptive Ant Colony Algorithm (采摘机器人全局覆盖路径规划研究——基于自适应蚁群算法). *Journal of Agricultural Mechanization Research*, Vol. 45, pp. 213-216, Haerbin/China.
[28] Zhang P., Song J., Gong S., & Muham, P.D. (2014). A kinematic analysis and simulation based on ADAMS for eggplant picking robot. *INMATEH-Agricultural Engineering*, Vol. 43, pp. 51-60, Romania.

[29] Zhao, J.Z., Wang, H.C., & Yu, Y.H., Li, J.P. (2023). Design of Safety Protection for Deep Reinforcement Learning Picking Robot(深度强化学习采摘机器人安全防护设计). *Computer Simulation*, Vol. 40, pp. 448-453, Beijing/China.

[30] Zhou, J., & Yu, C. (2023). Future Directions in Agricultural Robotics: Emerging Technologies for Smart Harvesting. *Journal of Field Robotics*, Vol.40, pp. 501-520, United States.