International Journal of Intelligent Robotics and Applications (2021) 5:186–202
https://doi.org/10.1007/s41315-021-00172-5

REGULAR PAPER

An obstacles avoidance method for serial manipulator based on reinforcement learning and Artificial Potential Field

Haoxuan Li[1,2] · Daoxiong Gong[1,2] · Jianjun Yu[1,2]

Received: 1 November 2020 / Accepted: 7 April 2021 / Published online: 17 April 2021
© The Author(s) 2021

Abstract
The obstacles avoidance of manipulator is a hot issue in the field of robot control. Artificial Potential Field Method (APFM) is a widely used obstacles avoidance path planning method, which has prominent advantages. However, APFM also has some shortcomings, which include the inefficiency of avoiding obstacles close to target or dynamic obstacles. In view of the shortcomings of APFM, Reinforcement Learning (RL) only needs an automatic learning model to continuously improve itself in the specified environment, which makes it capable of optimizing APFM theoretically. In this paper, we introduce an approach hybridizing RL and APFM to solve those problems. We define the concepts of Distance reinforcement factors (DRF) and Force reinforcement factors (FRF) to make RL and APFM integrated more effectively. We disassemble the reward function of RL into two parts through DRF and FRF, and make them activate in different situations to optimize APFM. Our method can obtain better obstacles avoidance performance through finding the optimal strategy by RL, and the effectiveness of the proposed algorithm is verified by multiple sets of simulation experiments, comparative experiments and physical experiments in different types of obstacles. Our approach is superior to traditional APFM and the other improved APFM method in avoiding collisions and approaching obstacles avoidance. At the same time, physical experiments verify the practicality of the proposed algorithm.

Keywords Reinforcement learning · Artificial Potential Field Method · Robotic arm · Obstacles avoidance

1 Introduction

Obstacles avoidance of manipulators is a hot and important issue in robot control research. At present, commonly used obstacles avoidance path planning methods include APFM (Wang et al. 2018), Fuzzy inference (Le et al. 2016), Rapidly-exploring random trees (RRT) (Tao et al. 2018), Approximate obstacle avoidance algorithm (Zhang et al. 2017a), etc. Among them, APFM is simple in principle, fast in the calculation, and has the ability to obtain environmental information in real-time. Therefore, it shows great advantages in complex environments and is often used in path planning to solve obstacles avoidance problems (Gai et al. 2019). However, APFM also has limitations, for instance, it is easy to fall into a locally optimal solution or produce vibrations, and lacks adaptive ability. In addition, it is also difficult to avoid dynamic obstacles, and the robotic arm is easy to fail avoiding obstacles by using APFM when the target is close to obstacles (Zhiyang and Tao 2017).

Considering the existing problems of traditional APFM, researchers usually use two species of methods to optimize its disadvantages. The first is to modify and optimize the internal parameters of the APFM. Guan et al. proposed a variable-step APFM in order to get rid of the local minima by the improved APFM. The manipulator avoided obstacles to reach the target point by modifying the potential energy function and the variable step length of the joint space for segmented search (Guan et al. 2015); Zhang et al. used a potential energy function containing joint angle information, which was on the basis of traditional APFM function, to guide the path planning of manipulator. At the same time, the manipulator could avoid the local minima by adding virtual obstacles (Zhang et al. 2017b). Li et al. (2019) selected a series of control points affected by the total potential field on the manipulator, and proposed a strategy that only attracted

✉ Haoxuan Li
hxlj97@163.com

1 Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

2 Beijing Key Lab of the Computational Intelligence and Intelligent System, Beijing 100124, China

the last point but all selected points were repelled by obstacles, then the points in the workspace were converted into the torque in the configuration space to make the manipulator move. Chen and Li (2017) modified the APFM function by increasing the gravitational field while reducing the repulsion field when approaching the obstacles, and applied it to the collision avoidance problem of unmanned ships; Lin et al. proposed the concept of rotating repulsive force field to avoid obstacles in three-dimensional space by modifying the function parameters. The target factor was introduced into the repulsive force field to link APFM function and the target, which provided a feasible direction for the robotic arm (Lin and Hsieh 2018). Song et al. (2020) proposed a Predicted-APFM that used time information and predicted potential to plan a smoother path with correct angle limits, speed adjustments, and correctly predicted potentials to improve the feasibility of generating paths. However, the above adjustments of traditional APFM's parameters do not touch the fundamentals of APFM in terms of algorithms. Thus, there are still some problems with those improved methods, such as their slow convergence speed and a large amount of calculation. Meanwhile, they are also weak in dealing with dynamic obstacles and collisions, and have failed to solve the local optimal solution drastically. Therefore, in order to solve the problems of APFM fundamentally, researchers often use the second type of improved method to obtain better results.

The second is to use complementary hybrid strategies to promote APFM. Qian et al. proposed an upgraded APFM based on connectivity analysis in order to solve the dynamic target path planning problem of the Location-Based Service system. By analyzing the connectivity of obstacles, this algorithm obtained the feasible solution domain, and planed the path in the feasible solution domain in advance, which prevented the agent from falling into the local optimal solution. Meanwhile, they set the speed factor and the maximum steering angle to obtain a relatively smooth path (Qian et al. 2018). Chen et al. (2019) avoided the local minima problem by setting virtual obstacles and targets, and used a simulated annealing algorithm to search for the optimal parameters of APFM. Shen et al. (2019) aimed at the dual-arm robot's control, carried out obstacles avoidance motion planning for the master arm against static obstacles firstly, then used the master arm as a dynamic obstacle when the slave arm moved, so that modified the traditional APFM and planed the complete path of the dual-arm robot. In recent years, the rapid development of RL has brought more new possibilities to the traditional APFM. RL has strong adaptability and self-learning capabilities in complicated environments. It can train the model through interactive or offline testing in the environment, and use the reward or punishment function to motivate the agent to learn new strategies. The agent interacts with the environment in real time, obtains the optimal behavior for completing the task through trial and error, and establishes a value function by observing the current state to predict the rewards of different behavior (Christen et al. 2019). Hence, many researchers have tried to combine RL with APFM. Zheng et al. (2015) proposed a multi-agent path planning algorithm based on hierarchical RL and APFM, and used the partial updating capability of hierarchical RL to ameliorate the problems of slow convergence and low efficiency of path planning algorithms. Yao et al. combined the improved Black Hole Potential Field with RL to make the agent automatically adapted to the environment and learned how to use basic environmental information to find the target. At the same time, the trained agent could adapt Curriculum Learning method to generate obstacles avoidance path in complex variable environments (Yao et al. 2020). Noguchi and Maki (2019) proposed an APFM based on binary Bayesian filtering and used RL to create a path in a simulated environment. The second type of improved APFM generally have stronger advantages compared to the first type, especially the strategies of mixing RL and APFM have better performance, which provides support for us to combine these two algorithms.

In this paper, we propose a new Reinforcement learning-Artificial Potential Field Method (RL-APFM) for avoiding obstacles to solve the problems that traditional APFM is not easy to avoid dynamic obstacles, and obstacles closing to target. We define the concept of reinforcement factors and find the optimal obstacles avoidance trajectory through iterative learning of RL, and RL-APFM can make up for deficiencies of traditional APFM by introducing RL. Firstly, traditional APFM detects the environment and creates obstacles avoidance trajectory. Secondly, when the robotic arm is too close to obstacles, DRF is activated, and the FRF is activated when a collision occurs. Finally, RL-APFM iteratively searches for the value of the optimal enhancement factor through RL to get rid of obstacles and reach the target. Meanwhile, we conduct multiple sets of experiments to verify the effectiveness and practicability of RL-APFM.

2 Traditional APFM

2.1 Obstacles avoidance problem

To simplify the analysis, obstacles are usually transformed into cylinders or spheres, and the center of each obstacle is their origin, the occupied volume is their range of influence. Hence, with the aid of the geometric topology principle, this paper approximates obstacles avoidance problem to be simplified and specific.

We assume that the range of obstacles is a sphere, and the robotic arm is simplified to multiple cylindrical linkages. The simplified obstacles avoidance situation is shown

in Fig. 1. Assuming that the coordinate of the sphere's center is $O(x_o, y_o, z_o)$, and the radius is r_o. The radius of one of the robotic arms' links is r_l, then the radius after merging the radius of cylinder and sphere is $r = r_o + r_l$. Assuming that the perpendicular from the center of sphere to connecting rob intersects the link at $F(x_f, y_f, z_f)$, which represents the position of the nearest point from the robotic arm to the obstacle.

Therefore, the distance between robotic arm and the obstacle can be expressed as $d_{of} = \sqrt{(x_o - x_f)^2 + (y_o - y_f)^2 + (z_o - z_f)^2}$. In order to ensure the safety of robotic arm, we set a safety distance δ, and the obstacles avoidance process can be roughly divided into three situations:

1. $d_{of} - r > \delta$ means the robotic arm is far away from the obstacle and it is safe.
2. $0 < d_{of} - r \leq \delta$ means the robotic arm is close to the obstacle, and the robotic arm needs to take appropriate actions to avoid collisions.
3. $d_{of} - r = 0$ means the robotic arm is in contact with or colliding with the obstacle, and the robotic arm needs to react immediately to get rid of the obstacle.

Therefore, the purpose of obstacles avoidance is to make the end effector reach the target position without colliding with obstacles. Even if sudden or unavoidable collisions occur, the manipulator can be quickly avoided to prevent further damage.

2.2 Process of APFM

The concept of artificial potential field was proposed by O. Khatib in 1985, which belonged to the local path planning algorithm. The basic idea of traditional APFM is to fill an artificial potential field in the working space of the robot. The target point generates a virtual "attractional field" in the workspace (defined as U_{att} in Eq. 1), and then generates an "attractional force" to attract the agent to approach it. Obstacles generate a virtual "repulsive field" in the workspace (defined as U_{rep} in Eq. 2), thereby generating a "repulsive force". Under the "attractional field" and "repulsive field", agents will be in a combined field (defined as U in Eq. 3), which can make the agent avoided obstacles and reached the target point (Wang et al. 2018).

$$U_{att}(q) = \frac{1}{2}\xi\rho^2(q, q_{tar}) \tag{1}$$

$$U_{rep} = \begin{cases} \frac{1}{2}\eta\left(\frac{1}{\rho(q,q_{obs})} - \frac{1}{\rho_0}\right)^2 & \rho(q, q_{obs}) \leq \rho_0 \\ 0 & \rho(q, q_{obs}) > \rho_0 \end{cases} \tag{2}$$

$$U(q) = U_{att}(q) + U_{rep}(q) \tag{3}$$

where, ξ is the positive scale factor of attraction, η is the positive scale factor of repulsion, $q, \mathrm{q}_{tar}, q_{obs}$ are the positions of the end effector, the target, and the obstacle in the working space respectively, $\rho(x, y)$ is the Euclidean distance between x and y, so $\rho(q, q_{tar})$ is the distance between the end effector and the target, $\rho(q, q_{obs})$ is the distance between the end effector and the obstacle, and ρ_0 is the effective radius of each obstacle. Attraction F_{att} and repulsion F_{rep} are defined as the negative gradient values of the attractional field and the repulsive field respectively, which showed in Eqs. 4 and 5. The resultant force on the end-effector F is shown in Eq. 6.

$$F_{att}(q) = -\nabla U_{att}(q) = \xi\rho(q_{tar} - q) \tag{4}$$

$$F_{rep}(q) = -\nabla U_{rep}(q) = \begin{cases} \eta\left(\frac{1}{\rho(q,q_{obs})} - \frac{1}{\rho_0}\right)\frac{1}{\rho^2(q,q_{obs})}\nabla\rho(q, q_{obs}) & \rho(q, q_{obs}) \leq \rho_0 \\ 0 & \rho(q, q_{obs}) > \rho_0 \end{cases} \tag{5}$$

$$F(q) = -\nabla U(q) = F_{att}(q) + F_{rep}(q) \tag{6}$$

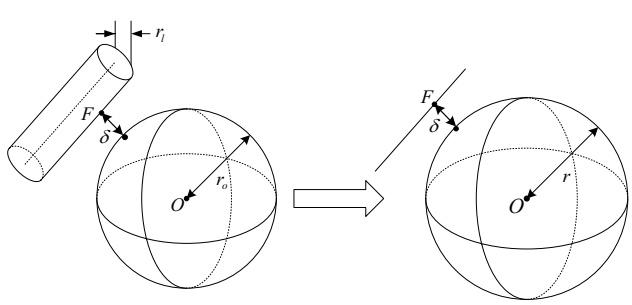

Fig. 1 Simplified obstacles avoidance situation

2.3 Problems of traditional APFM

APFM usually can avoid obstacles, but in some special circumstances, it happens to appear violent fluctuations or even fail to reach the target (as shown in Figs. 2 and 3).

In Fig. 2, the robotic arm collides with an obstacle in the first second, and begins to fluctuate periodically, which indicates that the robotic arm has fallen into a local minimum and cannot get rid of the obstacle. Figure 3 shows the posture of the robotic arm when it fails to avoid obstacles more intuitively.

These above phenomena show that APFM has certain limitations when avoiding obstacles, and the main reasons can be roughly summarized as follow:

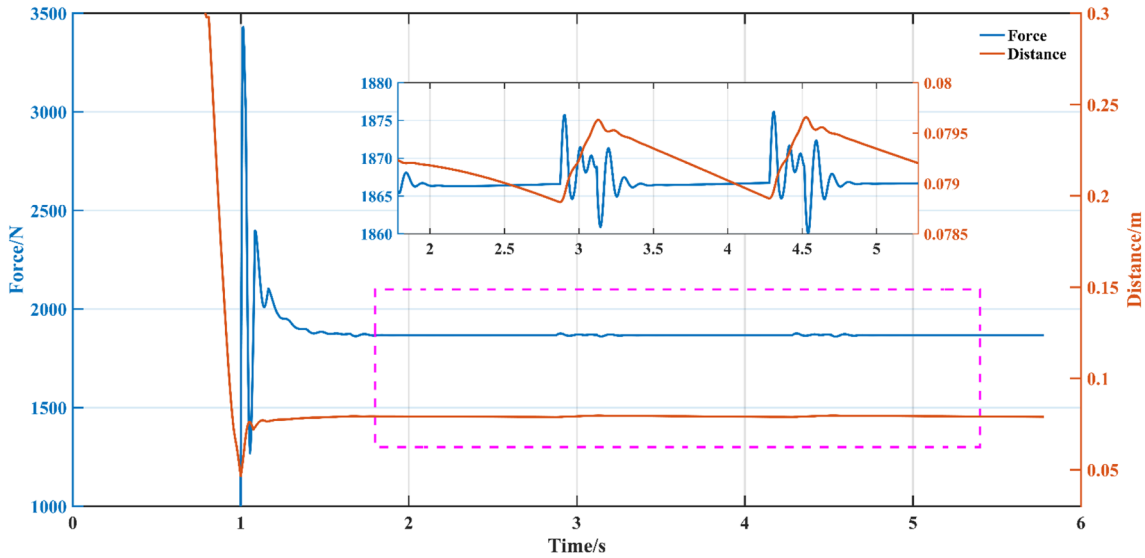

Fig. 2 Significant vibration during obstacles avoidance

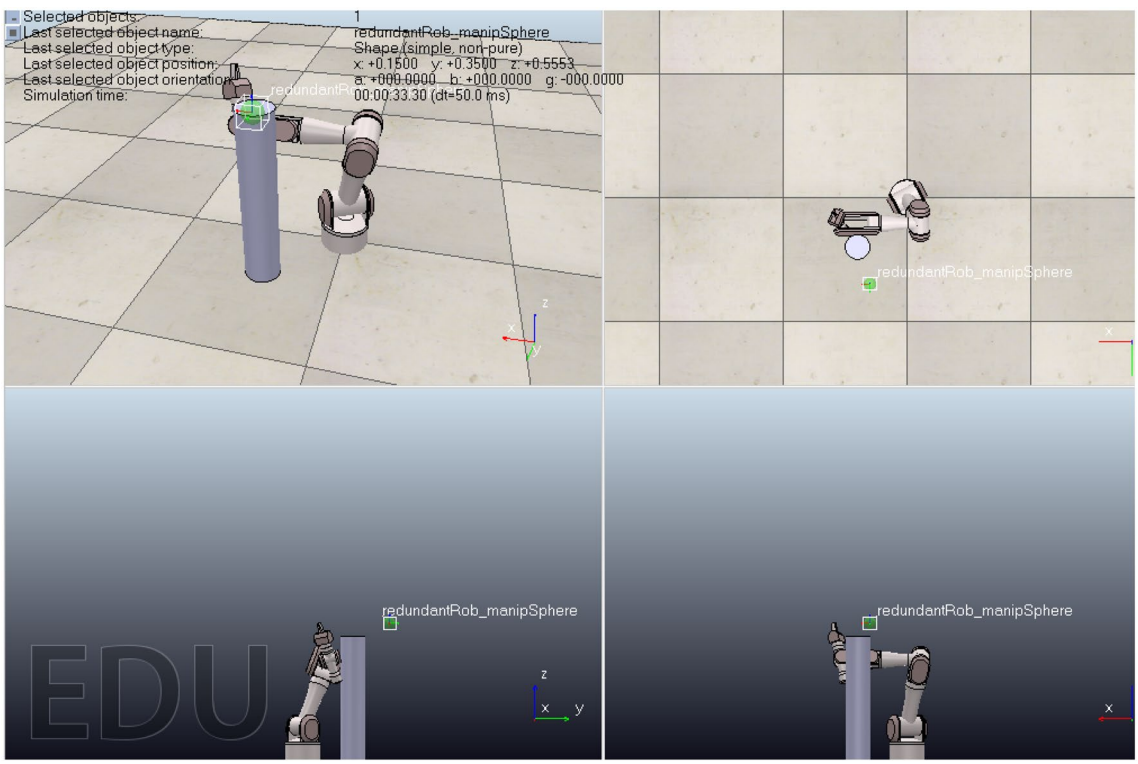

Fig. 3 Obstacles avoidance fail

1. The independent variables of U_{att} and U_{rep} of the traditional APFM are only related to the position of the target, obstacles, and the end effector, but no more other constraints. Therefore, when the position information continuously updates, the efficiency of APFM will be reduced greatly, which means, agents cannot avoid dynamic obstacles and track dynamic targets well.
2. Where the combined field is zero, F_{att} and F_{rep} cancel out each other, which makes it difficult for APFM to

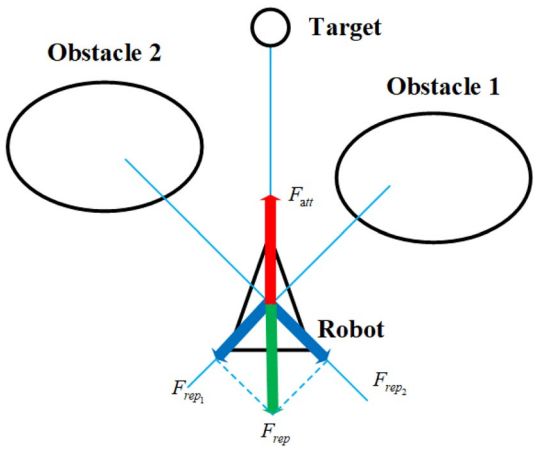

Fig. 4 The situation that combined field is zero

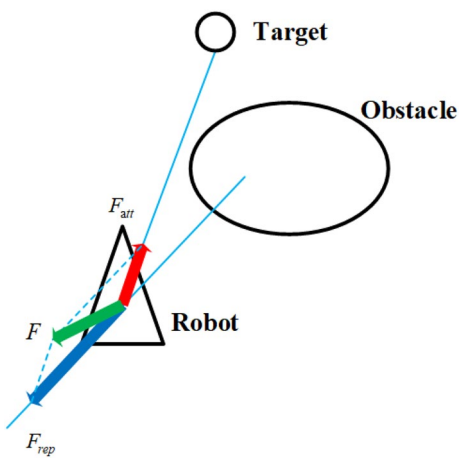

Fig. 5 The situation that obstacles are relatively close to target

judge trajectory, or even causes the local optimal solution problems (as shown in Fig. 4).

3. When the target and obstacles are relatively close to each other, F_{rep} will be far greater than F_{att}, resulting the robotic arm cannot approaching the target (as shown in Fig. 5).

In response to the above problems, we hybrid the RL and APFM, and introduce the concept of Reinforcement learning factors (RLF) including DRF and FRF, which can make the combination of RL and APFM much more effectively.

3 RL-APFM

RL rewards from a specific environment and trains the model through interactive or offline testing. The agent acquires knowledge from the environment and improves their policy to adapt to the environment. In this framework, the agent interacts with the environment through perception and operation. By observing the current state, a function is established to predict different behaviors. At the same time, the strategy generated by the value function maps the current state to the corresponding behavior. The environment reacts to the agent's behavior, and converts the new state information into the agent's response, then the agent receives feedback from the environment and updates the attribute value. Through the above process, the agent can adapt to the environment and react to different states accordingly.

RL tasks are usually described by the Markov decision process: the agent is in the environment $\mathbb{E}$, and the state space is S, where each state $s \in S$ is the description of the environment perceived by the machine; Actions which agent can choose constitute the action space A. If a certain action $a \in A$ acts in the current state s, the potential transfer function P will make the environment transfer from the current state to another state with a certain probability. At the same time, the environment will feedback to the agent according to the potential reward function R that makes the agent obtain a policy π in constant attempts.

In general, the RL task for the robotic arm corresponds to a set of tuples $M = \{S, A, P, R, T, S_0, \gamma\}$, where, S represents the environmental status, A represents the action status, $P : S \times A \times S \to \mathbb{R}$ represents state transition probability, $R : S \times A \times S \to \mathbb{R}$ represents the reward, T represents the working time, S_0 represents the initial state distribution, $\gamma \in (0, 1)$ represents the discount rate, $s_t \in S$ and $a_t \in A$. Define R_t as the discounted sum of future penalties:

$$R_t = \sum_{i=t}^{T} \gamma^{i-t} r(s_i, a_i) \tag{7}$$

We define the Q-function $Q^{\pi}(s_t, a_t)$, which describe the expected return under a policy π when taking action a_t from state s_t (also called as state-action pair) as follows:

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_M[R_t | s_t, a_t] \tag{8}$$

For all strategies, if the expected revenue of one strategy is greater than or equal to the revenue of other strategies, it will be the optimal strategy $Q^*(s_t, a_t)$, i.e.

$$Q^*(s_t, a_t) = \max_{\pi} \mathbb{E}_M[R_t | s_t, a_t] \tag{9}$$

The optimal strategy conforms to the bellman equation and can be expressed by Q value at the next moment as:

$$\mathrm{Q}^*(s_t, a_t) = \mathbb{E}_M[r + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) | s_t, a_t] \tag{10}$$

In order to solve the above problems of traditional APFM, considering the actual situation of obstacles avoidance of the robotic arm, we combine RL with APFM and

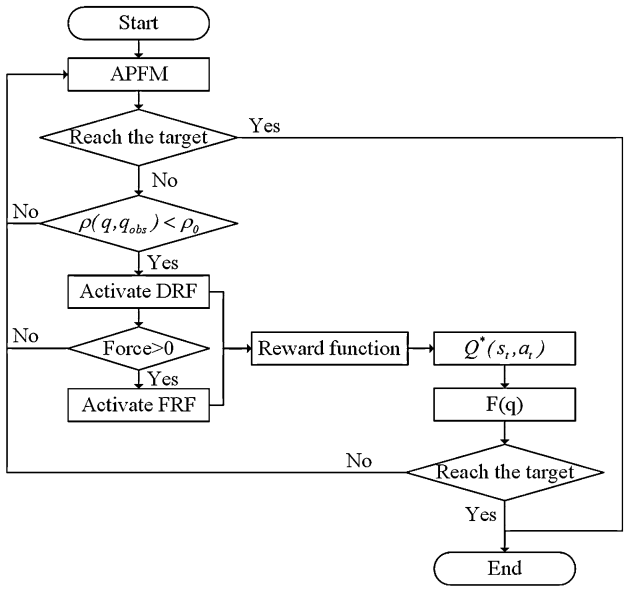

Fig. 6 Flow diagram of RL-APFM

propose RL-APFM. Considering that the obstacles avoidance process mainly needs to be focused on distances and forces, we split the reward function value $R = r(s_t, a_t)$ into two parts, which are DRF and FRF. Their corresponding Q function values are $Q_d^\pi(s_t, a_t)$ and $Q_f^\pi(s_t, a_t)$ respectively, which shown as below:

$$R_{DRF} = \sum_{i=t}^{T} \gamma^{i-t} \frac{1}{\rho(q, q_{obs}) + random(0, 10^{-3})} \tag{11}$$

$$R_{FRF} = \sum_{i=t}^{T} \gamma^{i-t} F \tag{12}$$

$$Q_d^\pi = \mathbb{E}_M[R_{DRF}|s_t, a_t] \tag{13}$$

$$Q_f^\pi = \mathbb{E}_M[R_{FRF}|s_t, a_t] \tag{14}$$

$$Q^\pi(s_t, a_t) = Q_d^\pi(s_t, a_t) + Q_f^\pi(s_t, a_t) \tag{15}$$

where, the reward functions R_{DRF} and R_{FRF} are determined by the distance $\rho(q, q_{obs})$ and the collision force F between the robotic arm and obstacles.

It can be seen that when the distance between the robotic arm and obstacles becomes shorter and the collision force between the robotic arm and obstacles becomes greater, the obtained reward function value should also be larger; otherwise, the reward function value should be smaller. Therefore, take the minimum reward function's value and find the corresponding Q function's value $Q^\pi(s_t, a_t)$ and define it as $Q^*(s_t, a_t)$, then the resultant force generated by APFM in each RL iteration process can be defined as:

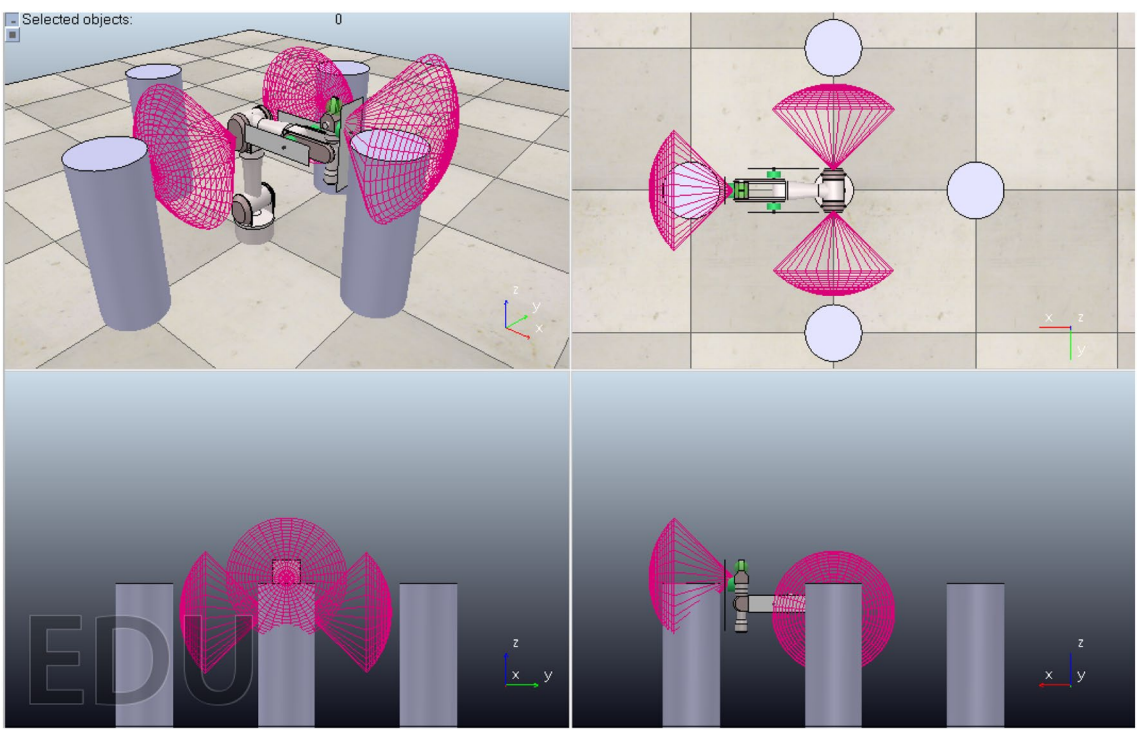

Fig. 7 Static obstacles environment

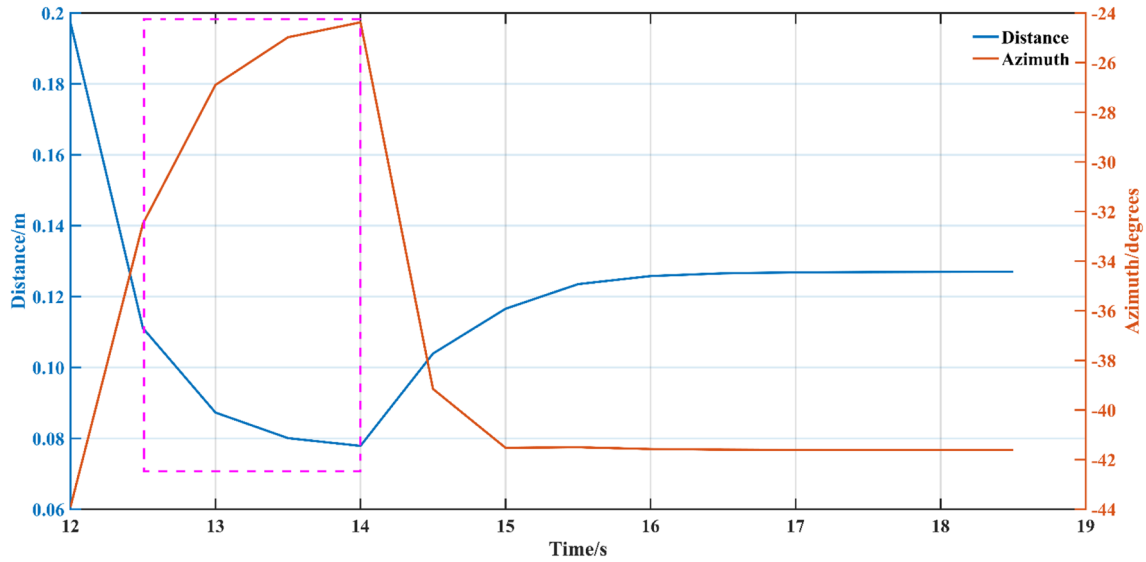

Fig. 8 Result of static obstacles avoidance without collision

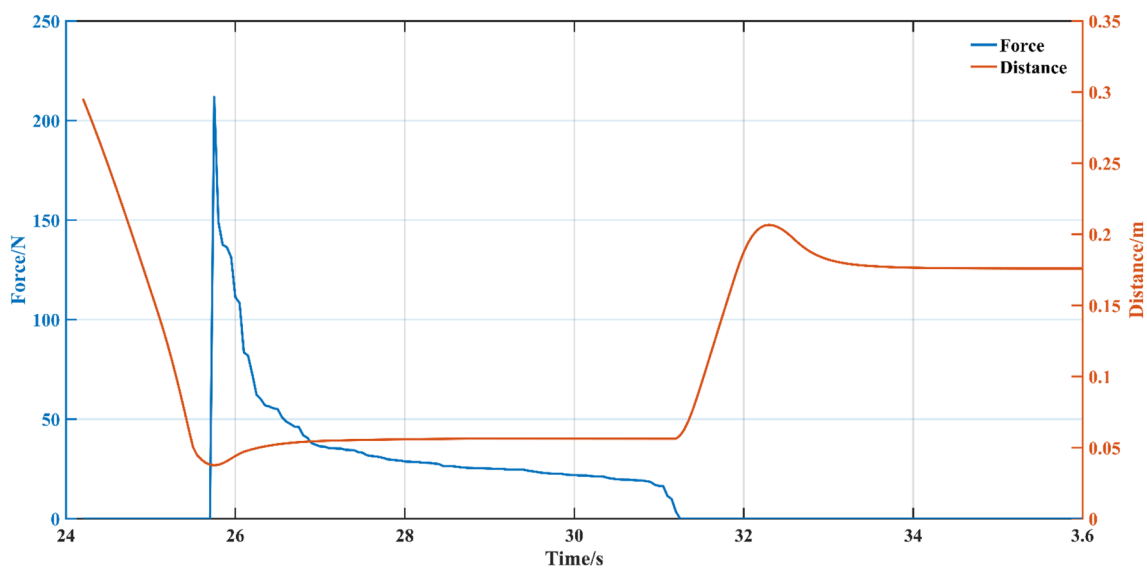

Fig. 9 Result of static obstacles avoidance with collision

$$F(q) = -\nabla U(q) = \frac{Q^*(s_t, a_t)}{1 + Q^*(s_t, a_t)} F_{rep}(q) + \frac{1}{1 + Q^*(s_t, a_t)} F_{att}(q) \tag{16}$$

Through analysis, it can be seen that when the robotic arm gets closer and closer to the obstacle, the value of $Q^*(s_t, a_t)$ becomes larger and larger, and the repulsive force is also increasing, helping the robotic arm to escape from obstacles. Conversely, when $Q^*(s_t, a_t)$ keeps decreasing, it indicates that the robotic arm has gradually escaped from the control range of obstacles, and the attractional field begins to restore its dominant position, guiding the object to continue moving towards the target. Therefore, the ultimate goal of the RL-APFM is to make the $Q^*(s_t, a_t)$ function value infinitely close to zero through iterative learning, which means the reward function value continues to decrease, to get rid of obstacles and reach the target.

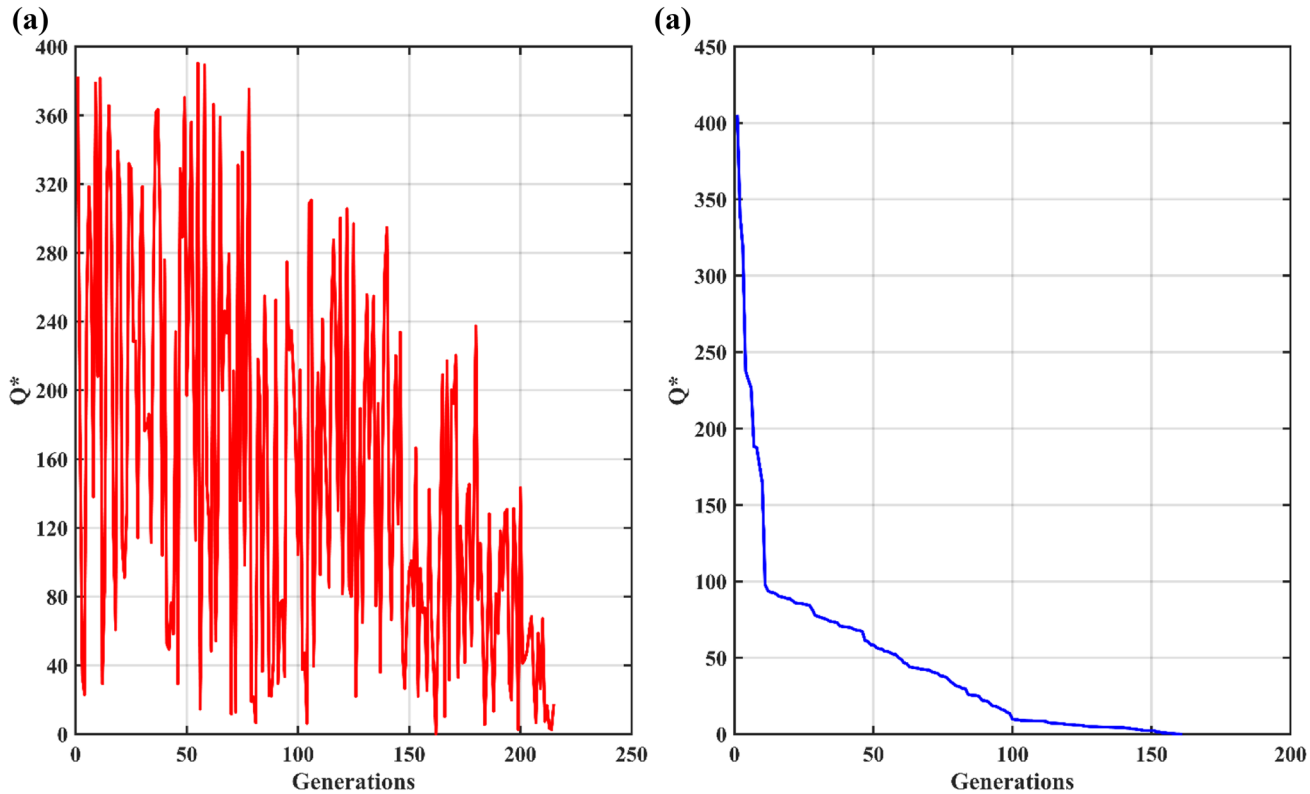

Fig. 10 Function value of Q^* under static obstacles avoidance with collision. **a** Before RL training, **b** After RL training

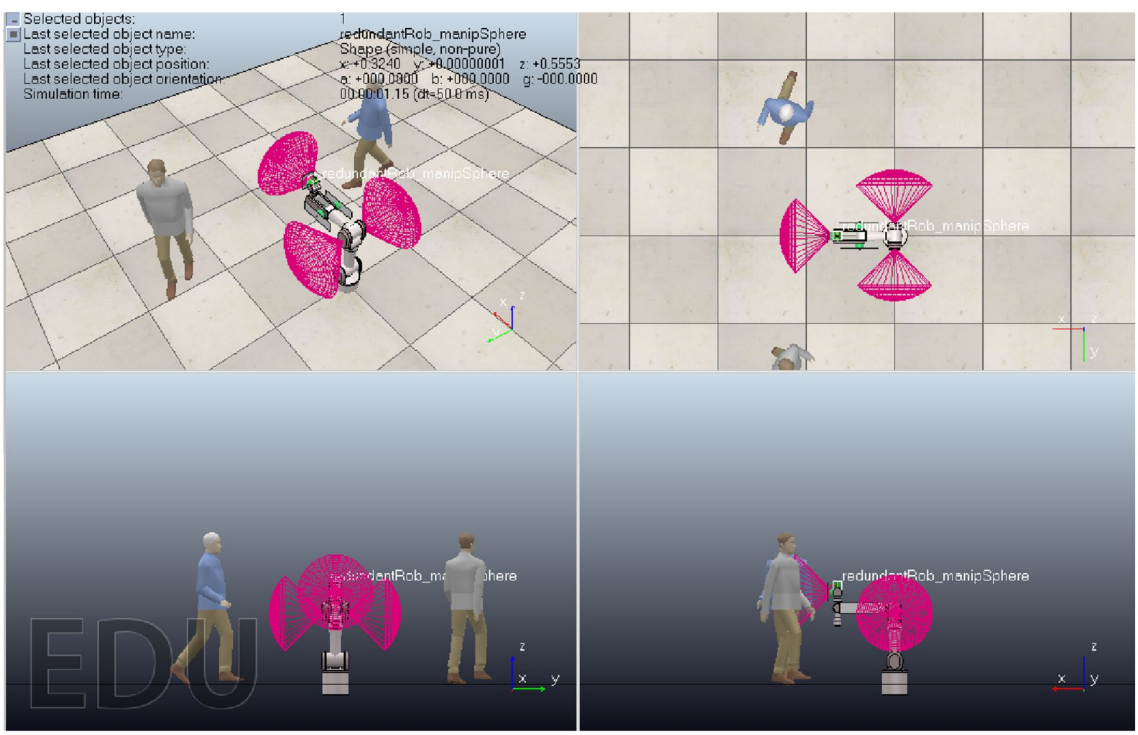

Fig. 11 Dynamic obstacles environment

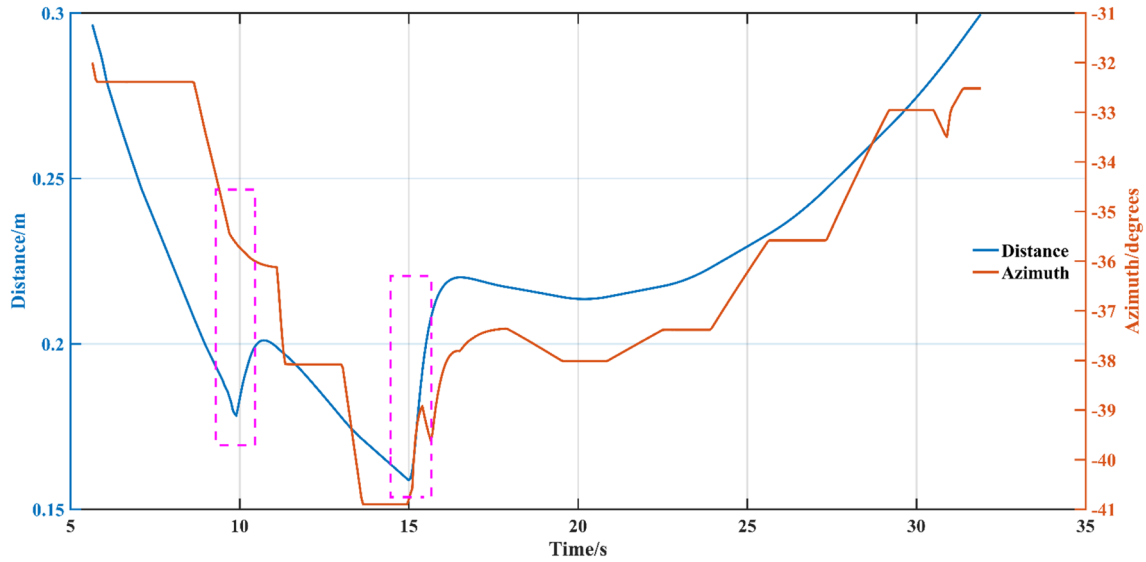

Fig. 12 Result of dynamic obstacles avoidance without collision

RL-APFM can show its unique advantages in dealing with the typical APFM's problems:

1. In addition to position information, RL-APFM introduces DRF and FRF, which makes the independent variables more abundant and more time-sensitive, and can have better performance during avoiding dynamic obstacles;
2. In the areas where the combined field is zero, RL can help RL-APFM to try out the best path direction and clarify the movement trajectory, at the same time, iterative trial-and-error of DRF and FRF can help RL-APFM get rid of locally optimal solutions and achieve the ultimate goal;
3. When the target and obstacles are relatively close, the activation of the DRF can modify the RL-APFM potential energy function to make the end effector continue to approach the target, and the FRF can ensure that the robotic arm can get rid of obstacles after collisions.

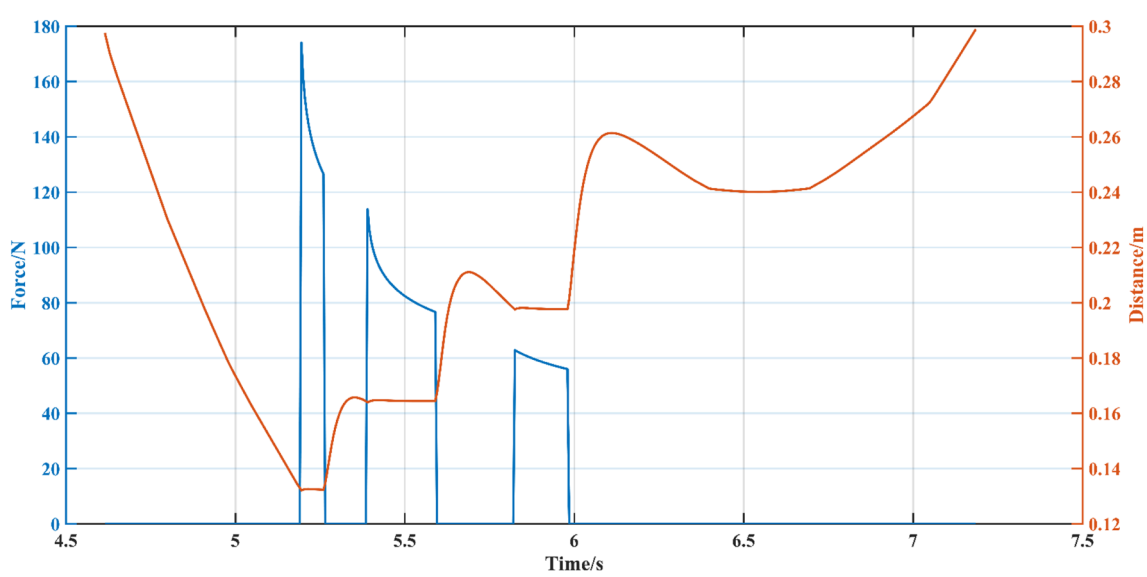

Fig. 13 Result of dynamic obstacles avoidance with collision

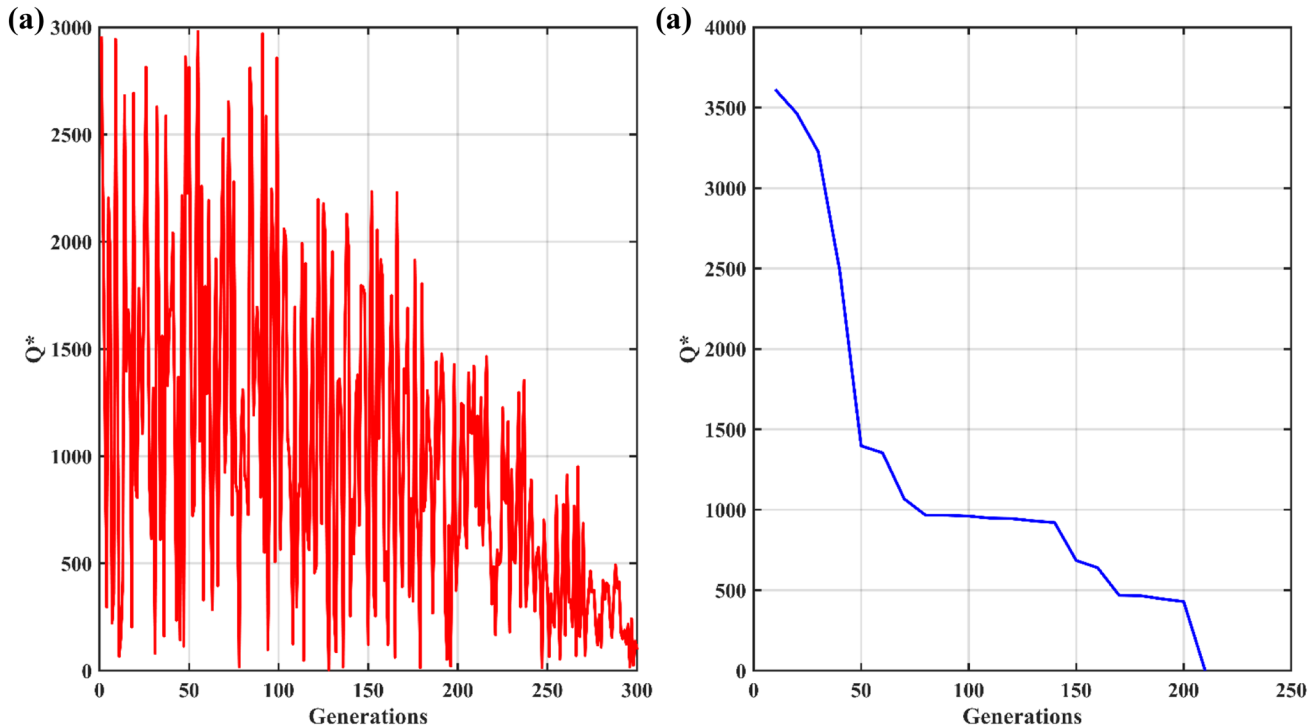

Fig. 14 Function value of Q^* under dynamic obstacles avoidance with collision. **a** Before RL training, **b** After RL training

Therefore, for general obstacles avoidance problems, RL-APFM adopts traditional APFM for path planning, and when the robotic arm enters the effective radius of the obstacles, DRF is activated to stay away from obstacles and avoid collisions. When the robotic arm collides obstacles, FRF is also activated at the same time to escape from obstacles quickly. The flow of the RL-APFM is shown in Fig. 6.

4 Simulations and experiments

4.1 Simulations and analysis

In order to verify the performance of the proposed algorithm, we use Matlab and CoppeliaSim as platforms to conduct an associated simulation of a 7-DOF redundant robotic arm. We place proximity infrared sensors on the front of the forearm and both sides of the elbow joint to detect the real-time distance between the robotic arm and obstacles. Similarly, place force sensors and protect planes on the front of the forearm and both sides of the shoulder joint to detect collisions and protect the robotic arm. Set the effective radius of obstacles $\rho_0 = 0.1$ m.

To verify the obstacles avoidance performance of RL-APFM aiming at static obstacles, we set up a group of static obstacles avoidance simulations. The goal is to move the end effector to a specified position while trying to avoid collisions between the robotic arm and static obstacles. Even in the appearance of an inevitable collision, the robotic arm can react quickly to escape obstacles. Figure 7 shows the static obstacles environment of simulations. Four cylindrical static obstacles are set in the range of 1 m*1 m area, and RL-APFM is used for obstacles avoidance of the robotic arm.

1. When no collision occurs during the obstacles avoidance of the robotic arm, the obstacles avoidance situation is shown in Fig. 8. A proximity infrared sensor detects an obstacle at the 12 *s*, and starts to get rid of the obstacle at the 14 *s*, and then completely escapes from the obstacle at the 16 *s*. While the robotic arm enters the dotted area, its speed slows down significantly and finally can realize obstacles avoidance, which proves that the DRF

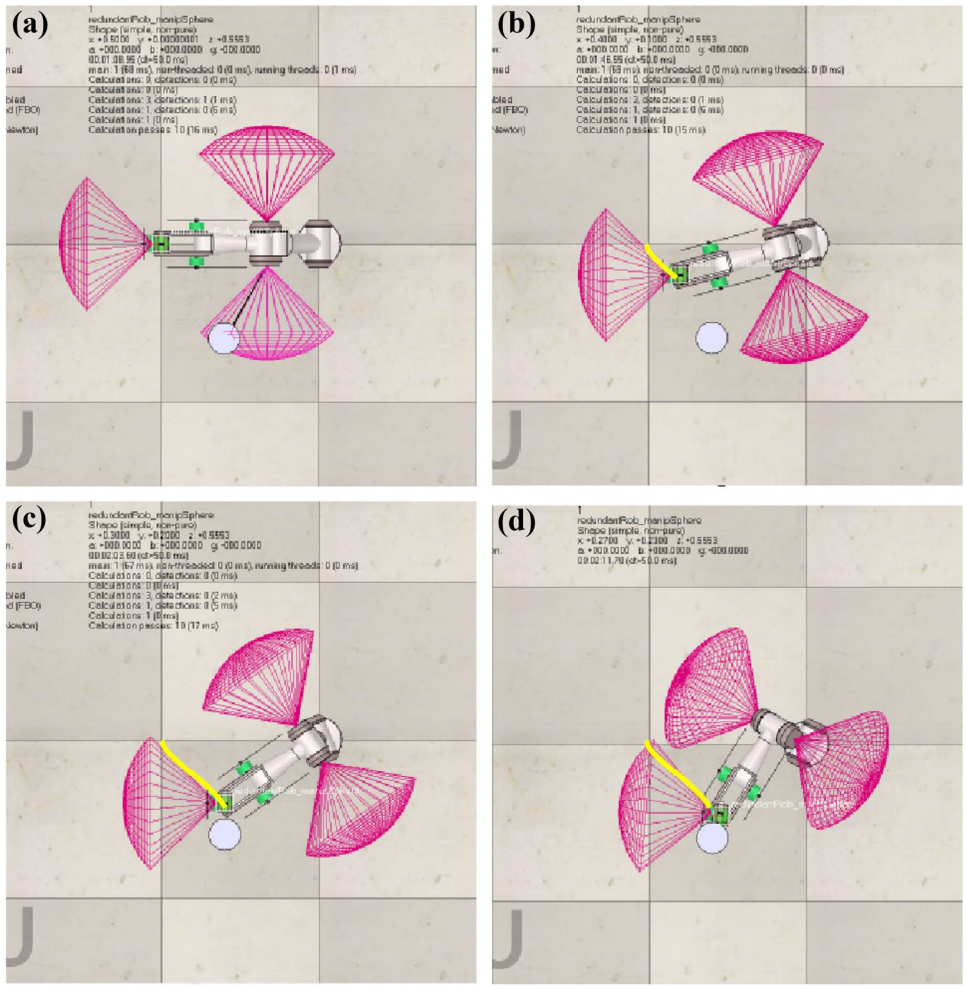

Fig. 15 Obstacles avoidance of static obstacle in traditional APFM

is effective in static obstacles avoidance process, and RL-APFM can realize non-collision obstacles avoidance under static obstacles environment.

2. When a collision occurs during the obstacles avoidance of the robotic arm, the obstacles avoidance result is shown in Fig. 9. A proximity infrared sensor detects an obstacle at 24.2 *s*, and a force sensor detects collision at 25.7 *s*. The robotic arm ends the collision at 31.3 *s* and completely gets rid of the obstacle at 33.7 *s*. When the robotic arm collides at 25.7 *s*, FRF is also activated under the premise that DRF remains activated. After a period of RL training, the robotic arm can finally achieve obstacles avoidance. Figure 10 shows that, after a period of iterative training, Q^*'s real-time value has shown an obvious downward trend relatively, which means that through multiple generations of training, RL-APFM has found the optimal obstacles avoidance solution.

In order to verify the performance of RL-APFM aiming at dynamic obstacles, we also set up a group of dynamic obstacles avoidance simulations. The dynamic obstacles environment is shown in Fig. 11. Two dynamic human-shaped obstacles are set at the widthwise and longitudinal positions of robotic arm and walk through the widthwise and longitudinal sides of robotic arm to simulate the avoidance with dynamic obstacles. Meanwhile, we adopt RL-APFM for obstacles avoidance under dynamic conditions. The target is to move the end effector to a specified position while trying to avoid collisions from dynamic obstacles. Even in the appearance of an inevitable collision or continuous collisions, the robotic arm can still escape obstacles as quickly as possible.

3. When no collision occurs during the obstacles avoidance of robotic arm, obstacles avoidance situation is shown in

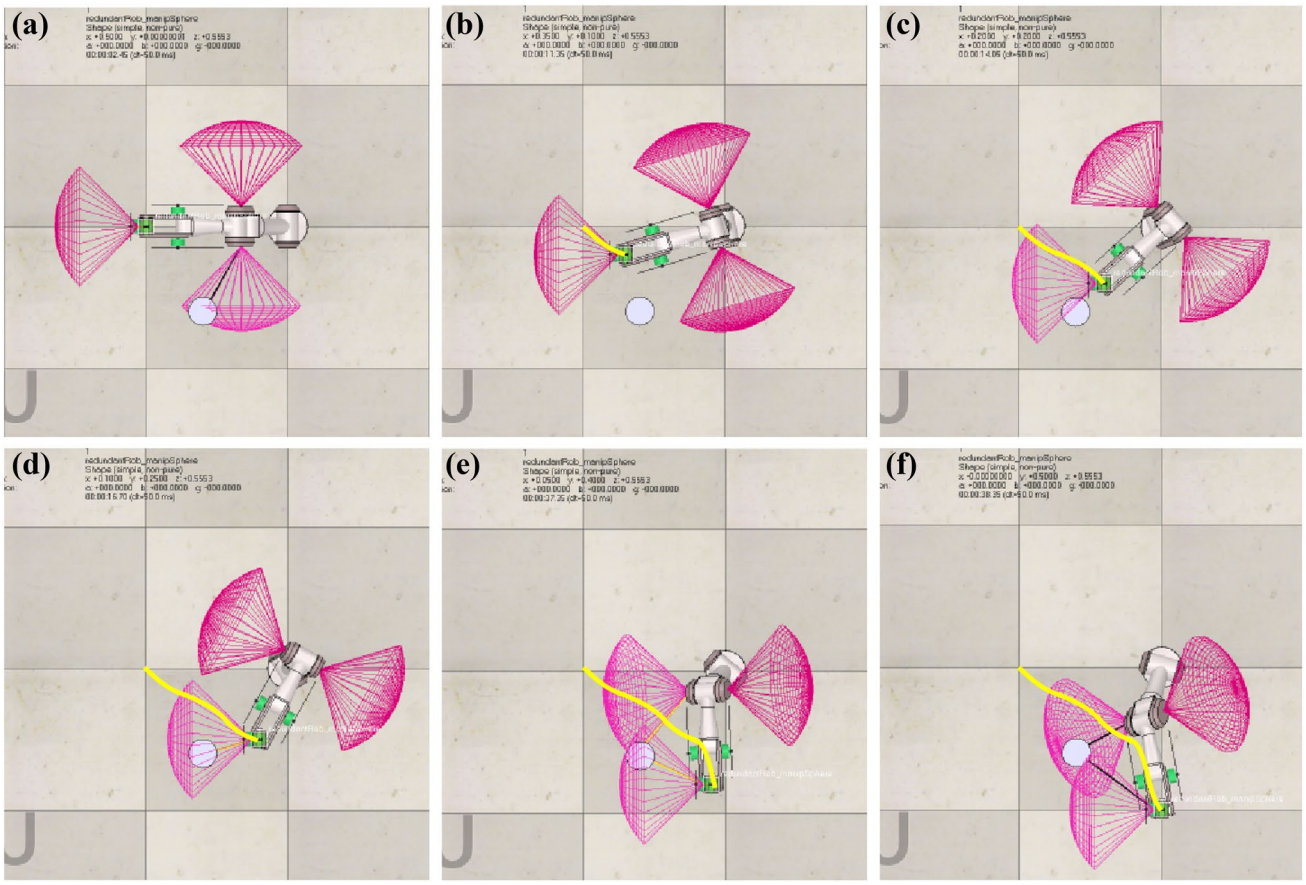

Fig. 16 Obstacles avoidance of static obstacle in RL-APFM

Fig. 12. Two dynamic human-shaped obstacles approach the robotic arm at about the 10 *s* and the 15 *s*, respectively. However, the robotic arm does not collide with them, and begins to escape from obstacles at about the 20 *s*. When a dynamic obstacle is detected approaching, RL-APFM can achieve avoidance more favorably, and avoid collision within dotted areas, which can realize non-collision obstacles avoidance under dynamic obstacles as much as possible.

4. When collisions occur during the obstacles avoidance of robotic arm, considering the elastic collision characteristics of dynamic obstacles, the robotic arm often needs to achieve continuous obstacles avoidance under multiple collisions to accomplish complete obstacles avoidance. The obstacles avoidance result is shown in Fig. 13. The robotic arm occurs three consecutive collisions with dynamic obstacles in the 5–6th seconds. However, the impact force on robotic arm is getting smaller and smaller, and the manipulator starts to escape from obstacles at about 6.5 s.

When dynamic obstacles continuously hit the robotic arm, FRF is activated while keeping DRF in the active state, so that the robotic arm is gradually away from the movement path of obstacles under the guidance of RL-APFM. Figure 14 shows that, after a period of iterative training, Q^*'s real-time value has shown an obvious downward trend relatively, which means through multiple generations of training, RL-APFM can still find the optimal obstacles avoidance solution under dynamic obstacles avoidance with collision.

4.2 Comparative experiments

Aiming at the problems that traditional APFM is prone to fail to avoid dynamic obstacles and obstacles closing to the target, we set up three sets of comparative experiments:

1. We compare the performance differences between RL-APFM and traditional APFM by making the end effector reach the target position closing to an obstacle to prove

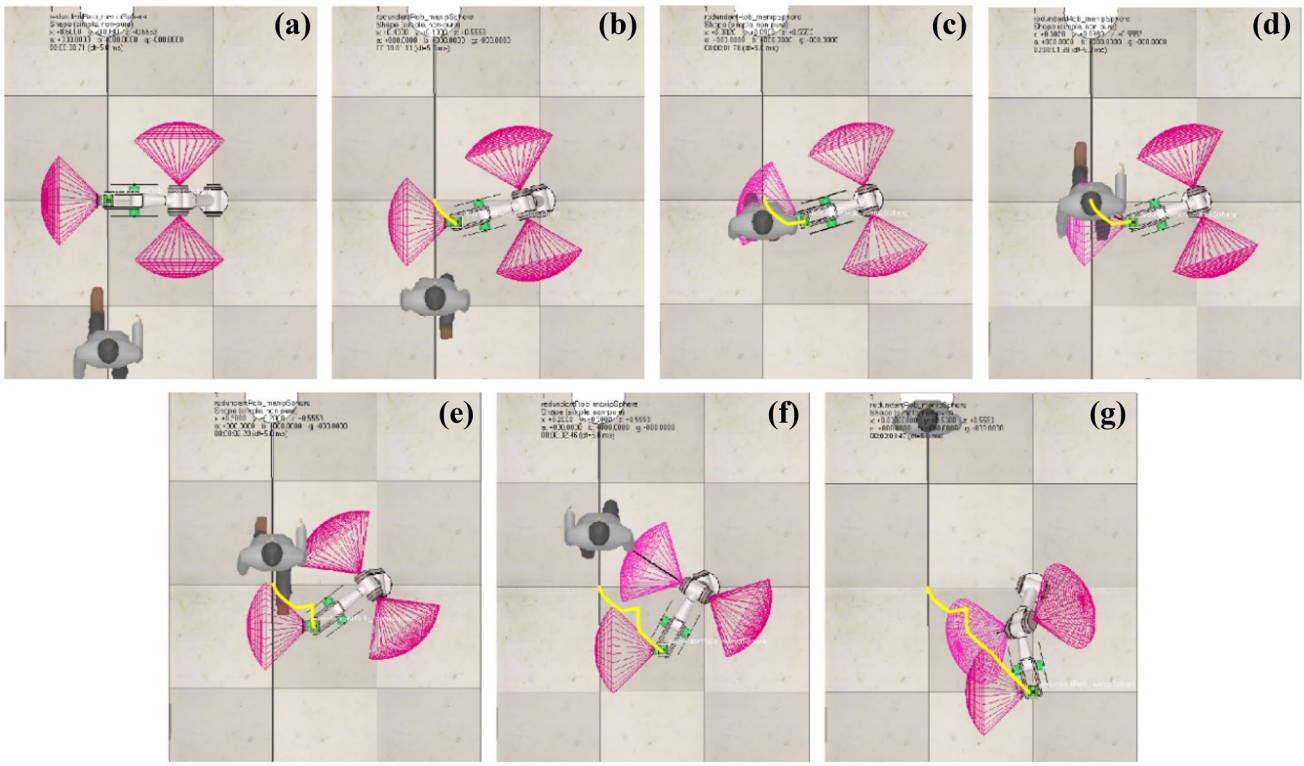

Fig. 17 Obstacles avoidance of dynamic obstacle in traditional APFM

the superiority of RL-APFM. Set the initial position of the end effector $X_{ini} = (0.5, 0, 0.5)$, the target position is $X_{tar} = (0, 0.5, 0.5)$, the unit meters. Set a cylindrical obstacle on the planned path of the robotic arm, and record the obstacle avoidance of APFM and RL-APFM respectively. The results are shown in Figs. 15 and 16.

Figure 15 shows that traditional APFM makes robotic arm collided in Fig. 15c and cannot escape from the obstacle, which causing the end-effector to fall into a locally optimal solution in Fig. 15d. Thus, the robotic arm cannot reach the target. Figure 16 shows that RL-APFM can avoid collisions in the same environment. Even if the target is close to the obstacle, it can realize obstacles avoidance successfully.

2. To prove the superiority of RL-APFM by avoiding dynamic obstacles, we set another comparative experiment, and the goal task of this experiment is to make the robotic arm reach the target position while facing dynamic obstacles. Keep the initial and target positions unchanged, set the approximate humanoid model with a radius of 0.25 m and a height of 1.2 m as a dynamic obstacle. We let the humanoid model walk through the planned path of the robotic arm during its movement. Record obstacles avoidance of traditional APFM and RL-APFM respectively, and the results are shown in Figs. 17 and 18.

   Figure 17 shows that traditional APFM cannot avoid the dynamic obstacle. The robotic arm collides in Fig. 17c and stalls in Fig. 17d. When dynamic obstacle leaves the acting range of robotic arm in Fig. 17e, the end effector can continue to move until it finally reaches the target. Figure 18 shows that RL-APFM can avoid dynamic obstacle and realize non-collision obstacles avoidance in dynamic obstacles environment, which proves its superior performance comparing with traditional APFM.

3. In order to compare the proposed RL-APFM with the method in Guan et al. (2015) and traditional APFM, we construct the same simulation environment with Guan et al. (2015) for comparison. The compare results are shown in the following Table 1.

By comparison, we can find that in the experimental environment of Guan et al. (2015), traditional APFM falls into a local minimum and fails to avoid obstacles. The improved APFM can jump out of the local minimum to complete obstacles avoidance. Under the same simulation conditions, RL-APFM can also achieve obstacles avoidance with shorter

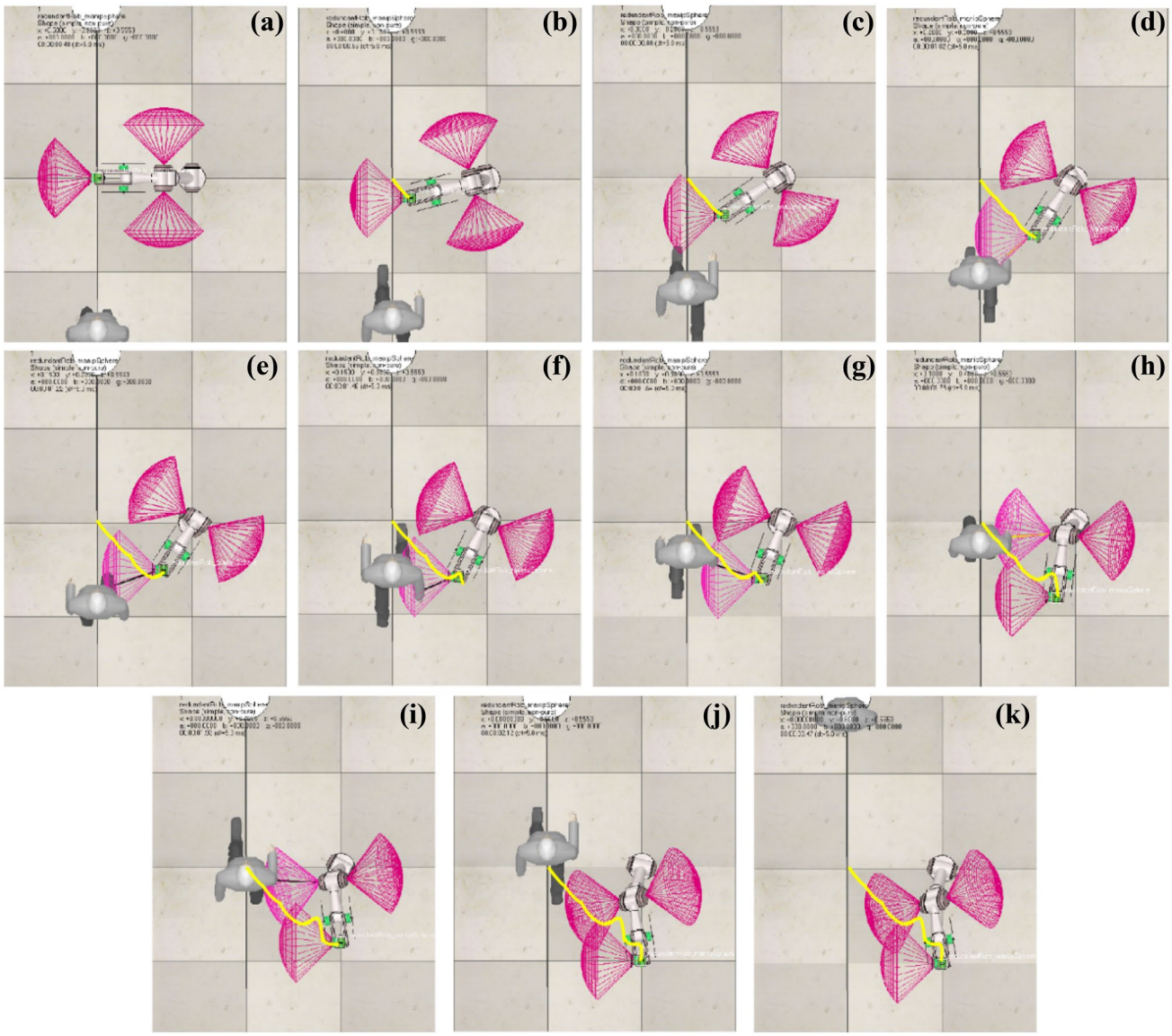

Fig. 18 Obstacles avoidance of dynamic obstacle in RL-APFM

Table 1 Comparison of RL-APFM and other methods

	APFM	Improved APFM in Guan et al. (2015)	RL-APFM
Time/s	–	80	68
Generations	–	100	76

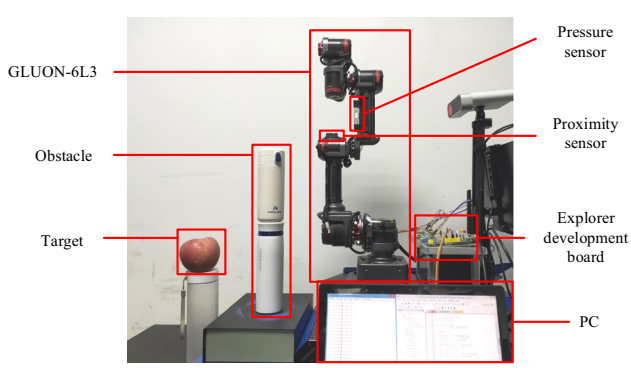

Fig. 19 Physical experiments platform

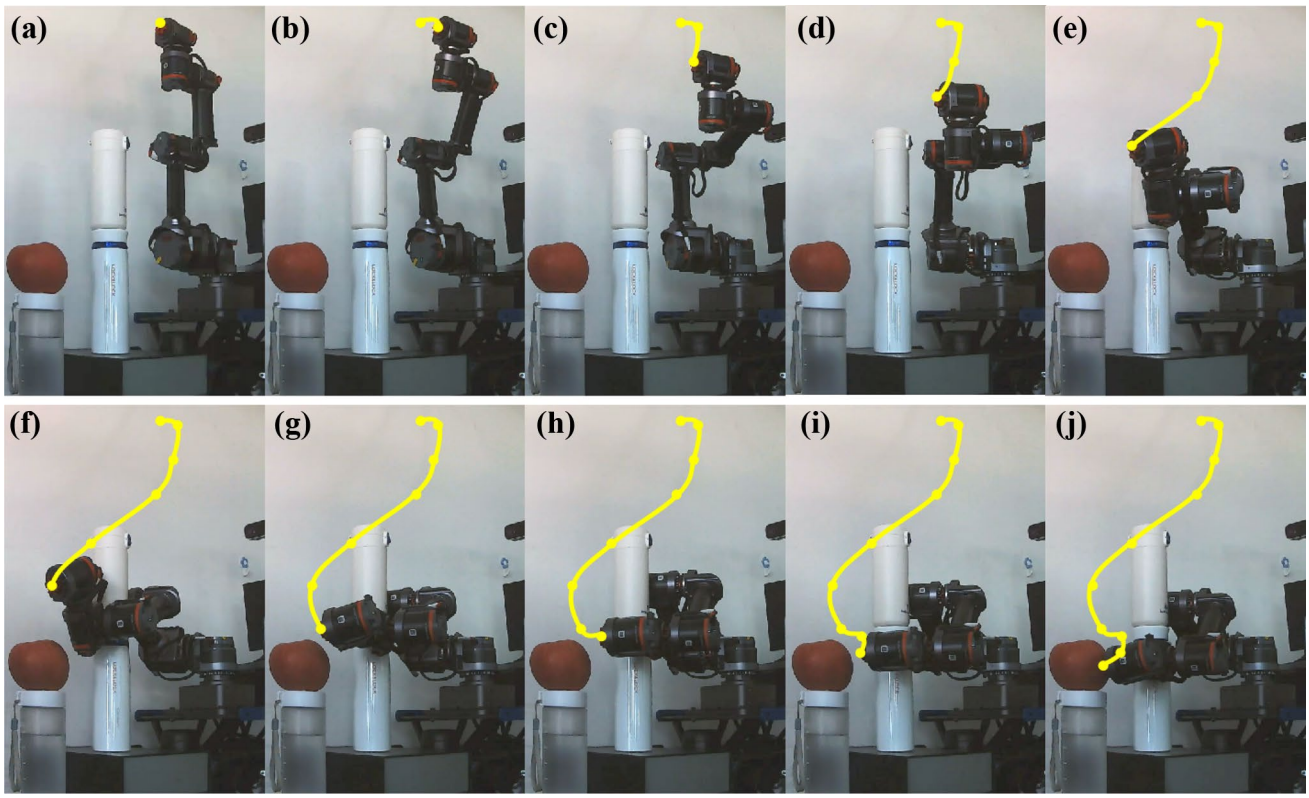

Fig. 20 Movement of the robotic arm facing static obstacles

time and faster iteration, which proves the superior performance of the proposed method.

4.3 Physical experiments

We migrate the proposed algorithm to INNFOS's GLUON-6L3 robotic arm, and install an appropriate amount of proximity sensors and pressure sensors on robotic connecting links. Use the Explorer development board to connect sensors to the PC to read sensors' values. In order to save time and ensure the safety of the robotic arm, we directly apply the RL-APFM whose parameters have been trained on the simulation platform. The purpose of experiments is to make the end effector reach the target position under the premise of avoiding static and dynamic obstacles. The physical experiments platform is shown in Fig. 19.

1. When the robotic arm faces static obstacle closing to the target, the movement process of the robotic arm is shown in Fig. 20, and the yellow line shows the trajectory of the end effector. The experimental results show that the robotic arm using RL-APFM can avoid closing obstacles without collision, which proves the effectiveness of RL-APFM in facing closing static obstacles scenarios.
2. We imitate the collision against the robotic arm in the process of facing dynamic obstacles by hitting pressure sensors. The movements of the robotic arm are shown in Fig. 21. The yellow line shows the trajectory of the end effector before the collision, and the blue line shows the expected motion trajectory of the end of the robotic arm without collision. After hitting the force sensor in Fig. 21d, the robotic arm starts to achieve dynamic obstacles avoidance and finally makes the end effector reach the target, as shown by the red line from Fig. 21e to h. The experimental results show that the robotic arm using RL-APFM can react, avoid collisions, complete obstacles avoidance, and reach the target quickly when facing dynamic obstacles and collisions, which proves the effectiveness of RL-APFM in facing dynamic obstacles and collision scenarios.

Through the above multiple sets of simulations and experiments, it can be seen that the RL-APFM can realize collision and non-collision obstacles avoidance in static and dynamic obstacles environments. Meanwhile, RL-APFM shows more

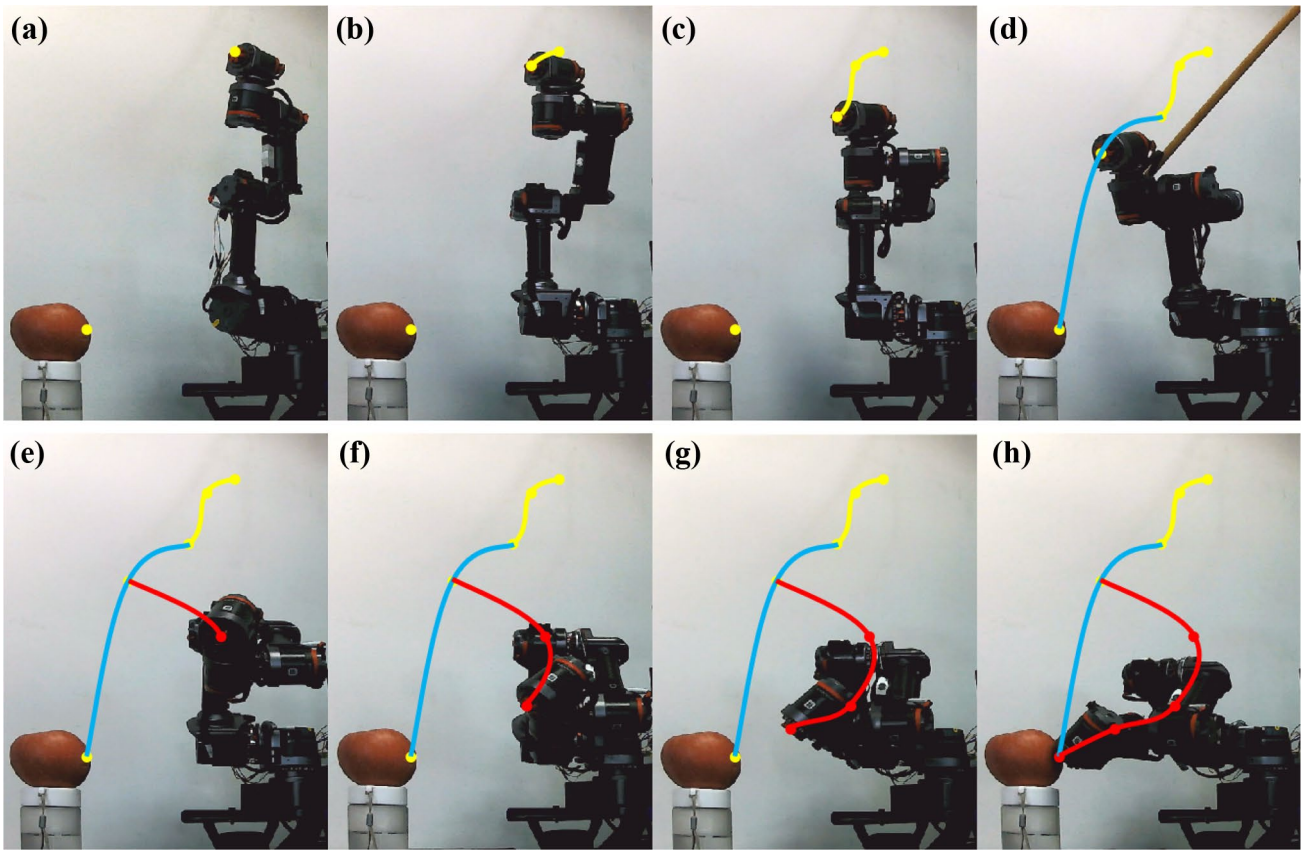

Fig. 21 Movement of the robotic arm facing dynamic obstacles (color figure online)

predominant performance when facing the obstacles closing to the target position. Besides, RL-APFM also has better dynamic obstacles avoidance ability than traditional APFM. Finally, physical experiments prove that RL-APFM can be used on physical robots and has a great obstacles avoidance effect.

5 Conclusion

In this paper, we construct the RL-APFM aiming at the obstacles avoidance problems proved by the serial redundant manipulator and physical robotic arm. RL is grafted into traditional APFM to solve the two main problems of APFM that are prone to occur in the obstacles avoidance process, which include traditional APFM is not easy to avoid dynamic obstacles and obstacles closing to target. In order to combine RL and APFM efficiently, we define the concepts of DRF and FRF, and apply them in the reward function of RL. Then, the optimal strategy determined by the reward function is applied to APFM, and through iterative or offline learning in a specified environment, the robotic arm can achieve obstacles avoidance eventually. In order to verify the performance of RL-APFM, we describe the advantages of RL-APFM over traditional APFM theoretically. At the same time, we design four collision and non-collision obstacles avoidance experiments in static and dynamic obstacles environments and three groups of comparative experiments. The simulation results show that the RL-APFM can avoid dynamic obstacles and realize obstacles avoidance when the target is close to obstacles by virtue of its unique advantages from RL. The comparative experiments' results show RL-APFM's superior performance compared with traditional APFM and another mentioned improved APFM when facing dynamic obstacles environment and the situation that the target and obstacles are approaching. At last, the physical experiments prove the effectiveness and practicability of RL-APFM.

Funding This research was supported by National Natural Science Foundation of China (Grant 61673003) and Natural Science Foundation of Beijing Municipality (Grant 4192010).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Chen, Y., Li, T.: Collision avoidance of unmanned ships based on artificial potential field. In: IEEE Chinese Automation Congress (CAC), pp. 4437–4440 (2017)

Chen, Z., Ma, L., Shao, Z.: Path planning for obstacle avoidance of manipulators based on improved artificial potential field. In: IEEE Chinese Automation Congress (CAC), pp. 2991–2996 (2019)

Christen, S., Stevšić, S., Hilliges, O.: Guided deep reinforcement learning of control policies for dexterous human-robot interaction. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 2161–2167 (2019)

Gai, S.N., Sun, R., Chen, S.J., Ji, S.: 6-DOF robotic obstacle avoidance path planning based on artificial potential field method. In: IEEE 16th International Conference on Ubiquitous Robots (UR), pp. 165–168 (2019)

Guan, W., Weng, Z., Zhang, J.: Obstacle avoidance path planning for manipulator based on variable-step artificial potential method. In: IEEE The 27th Chinese Control and Decision Conference (CCDC), pp. 4325–4329 (2015)

Le, K.D., Nguyen, H.D., Ranmuthugala, D., Forrest, A.: Artificial potential field for remotely operated vehicle haptic control in dynamic environments. Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng. **230**(9), 962–977 (2016)

Li, H., Wang, Z., Ou, Y.: Obstacle avoidance of manipulators based on improved artificial potential field method. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 564–569 (2019)

Lin, H.-I., Hsieh, M.-F.: Robotic arm path planning based on three-dimensional artificial potential field. In: IEEE 18th International Conference on Control, Automation and Systems (ICCAS), pp. 740–745 (2018)

Noguchi, Y., Maki, T.: Path planning method based on artificial potential field and reinforcement learning for intervention AUVs. In: IEEE Underwater Technology (UT), pp. 1–6 (2019)

Qian, C., Qisong, Z., Li, H.: Improved artificial potential field method for dynamic target path planning in LBS. In: IEEE Chinese Control And Decision Conference (CCDC), pp. 2710–2714 (2018)

Shen, D., Niu, W., Liu, J., Li, H., Tan, C., Xue, Y., Sun, Y.: Obstacle avoidance path planning for double manipulators based on improved artificial potential field method. In: Proceedings of the 2nd International Conference on Information Technologies and Electrical Engineering, pp. 1–5 (2019)

Song, J., Hao, C., Su, J.: Path planning for unmanned surface vehicle based on predictive artificial potential field. Int. J. Adv. Rob. Syst. **17**(2), 1729881420918461 (2020)

Tao, T., Zheng, X., He, H., Xu, J., He, B.: An improved RRT algorithm for the motion planning of robot manipulator picking up scattered piston. In: IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), pp. 234–239 (2018)

Wang, W., Gu, J., Zhu, M., Huo, Q., He, S., Xu, Z.: An obstacle avoidance method for redundant manipulators based on artificial potential field. In: IEEE International Conference on Mechatronics and Automation (ICMA), pp. 2151–2156 (2018)

Yao, Q., Zheng, Z., Qi, L., Yuan, H., Guo, X., Zhao, M., Liu, Z., Yang, T.: Path planning method with improved artificial potential field—a reinforcement learning perspective. IEEE Access **8**, 135513–135523 (2020)

Zhang, E., Yu, J., Li, M.: Research on autonomous obstacle avoidance algorithm of teleoperation manipulator in unstructured unknown environment. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 622–627 (2017a)

Zhang, N., Zhang, Y., Ma, C., Wang, B.: Path planning of six-DOF serial robots based on improved artificial potential field method. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 617–621 (2017b)

Zheng, Y., Li, B., An, D., Li, N.: A multi-agent path planning algorithm based on hierarchical reinforcement learning and artificial potential field. In: IEEE 11th International Conference on Natural Computation (ICNC), pp. 363–369 (2015)

Zhiyang, L., Tao, J.: Route planning based on improved artificial potential field method. In: IEEE 2nd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), pp. 196–199 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Haoxuan Li was born in 1997. He received the B.S. degree in Automation from Beijing University of Technology in 2019. His research interests include the bio-inspired robotics, artificial intelligent control, and evolutionary computation.

Daoxiong Gong received the PhD degree in Control Science and Engineering from Beijing University of Technology in 2004. He was an academic visiting scholar with the Laboratory for Robotics and Automation, the department of Electrical and Computer Engineering, Michigan State University, US, from 2012 to 2013. He is a professor with the Faculty of Information Technology, Beijing University of Technology, Beijing, China. His research interests include the bio-inspired robotics, the tele-robotics, intelligent control, and evolutionary computation.

Jianjun Yu received the M.S. degree in Control Science and Engineering and the B.S. degree in Automatic Control from Beijing University of Technology, China, in 2004 and 1989. She is an associate professor with the Faculty of Information Technology, Beijing University of Technology. Her research interests include humanoid robot, intelligent control, and imitation learning.