Research on multi-objective path planning and dynamic obstacle avoidance algorithm of manipulator based on reinforcement learning

Zhengis Arkalyk

University Of Birmingham, Birmingham, UK

zhenisarka@gmail.com

Abstract. Aiming at the problem of multi-objective path planning (MOPP) and dynamic obstacle avoidance of manipulator in dynamic environment, this paper proposes a solution based on hierarchical reinforcement learning (HRL) framework. Traditional path planning methods have some problems in dynamic scenes, such as poor real-time performance, difficult to balance multi-objective conflicts and insufficient adaptability to environmental changes. Therefore, this paper designs a two-tier architecture including global path planning layer and local obstacle avoidance layer, which is trained by Proximal policy optimization (PPO) and Soft Actor-Critic (SAC) algorithms respectively, and achieves collaborative optimization through an efficient information exchange mechanism between layers. At the same time, a multi-objective reward function based on dynamic weight adjustment strategy is introduced. Fuzzy logic is used to adaptively balance the relationship among path length, obstacle avoidance safety and energy consumption according to environmental complexity. Combined with Long Short-Term Memory (LSTM), the trajectory of obstacles is predicted, and the potential field method is further introduced to modify the obstacle avoidance reward, which improves the real-time response ability and robustness of the algorithm in dynamic environment. The experimental results show that the HRL-SAC-PPO method proposed in this paper shows superior performance in both static and dynamic scenarios. In the static scene, the success rate of this method reaches 100%, the average path length is shortened to 2.13m, no collision occurs, and the energy consumption is reduced to 1.12 kJ, which shows a good multi-objective optimization effect. In the dynamic scene, the trajectory error of obstacles predicted by LSTM is only 4.2%, and the safe distance between the robot arm and obstacles is improved by 35%, which significantly enhances the reliability of obstacle avoidance. In addition, the average decision delay of this method is only 11.3ms, and the peak delay is 23ms, which is much lower than that of the contrast algorithm, showing stronger real-time response ability. The ablation experiment further verified the key role of LSTM trajectory prediction, dynamic weight adjustment and layered structure on the overall performance. In the welding task verification of the real UR5 manipulator, the success rate of the system in dynamic environment is 95.3%, the average path length is 3.41m, and the maximum joint acceleration is only 0.87 radian/s^2, which is far below the safety threshold, indicating that the algorithm has good stability and obstacle avoidance ability in practical application. The comprehensive performance comparison shows that this method performs well in different industrial scenarios, and has stronger environmental adaptability and comprehensive path planning ability.

Keywords: Multi-objective path planning; Dynamic obstacle avoidance; Manipulator; Reinforcement learning; Hierarchical reinforcement learning; Proximal policy optimization; Soft actor-critic; LSTM.

1. Introduction

With the rapid development of industrial automation and intelligent manufacturing technology, robotic arms are increasingly widely used in complex tasks such as assembly, handling and welding. The traditional path planning method of manipulator has shown high efficiency in static environment, but it has obvious limitations in dynamic scene. The dynamic environment requires the robot arm to have real-time decision-making ability, which can quickly adjust the path to avoid obstacles and meet the requirements of multi-objective optimization. However, traditional methods are difficult to balance multi-objective conflicts and have poor adaptability to environmental changes, resulting in low efficiency of task execution or potential safety hazards.

Reinforcement learning (RL) learns optimal strategies to achieve specific goals through the interaction between agents and the environment [1]. In the path planning of manipulator, RL can deal with the problems of continuous action space and high-dimensional state space, which is suitable for complex dynamic environment [2]. For example, reference [3] proposed a robot path planning method based on Q-learning algorithm, and realized the effective path planning of manipulator in static environment by learning the value function of state-action pair. Multi-objective path planning (MOPP) aims to optimize multiple objective functions at the same time, such as minimizing time, energy consumption and path length [4]. There may be conflicts between these goals, so it is necessary to find a balance point so that each goal can be optimized to some extent [5]. Multi-mode MOPP problem (MMOPP) is a special case of MOPP, which requires finding all Pareto optimal paths from the starting point, through several given key points, and finally to the end point [6]. Dynamic obstacle avoidance algorithm enables the manipulator to adjust the path in real time to avoid obstacles [7]. The concepts and algorithms such as collision cone area, speed obstacle area and mutual speed obstacle area are mentioned in the introduction of dynamic obstacle avoidance basic algorithm, which are all important parts of dynamic obstacle avoidance algorithm [8]. RVO (Reciprocal Velocity Obstacle) algorithm is a commonly used dynamic obstacle avoidance algorithm, which can realize effective obstacle avoidance and path finding in multi-agent environment. The path planning method of manipulator based on RL can deal with complex dynamic environment and multi-objective optimization problems [9]. For example, reference [10] proposed an online trajectory planning method based on depth RL, which was used to capture moving targets by a six-degree-of-freedom space floating manipulator. In this method, the coupling characteristics of combined mechanics are considered, and the kinematics and dynamics models of multi-rigid bodies are established.

Aiming at the MOPP and obstacle avoidance problem of manipulator in dynamic environment, this paper systematically analyzes the existing problems of multi-objective conflict, insufficient real-time and robustness of RL algorithm and low accuracy of environmental perception, and puts forward a solution based on hierarchical RL framework. A double-layer architecture including global path planning layer and local obstacle avoidance layer is innovatively constructed, and collaborative optimization is realized through efficient information interaction mechanism between layers. At the same time, a multi-objective reward function based on dynamic weight adjustment strategy is designed, and the relationship between path length, obstacle avoidance safety and energy consumption is adaptively balanced according to environmental complexity by fuzzy logic. In addition, the LSTM network is combined to predict the trajectory of obstacles, and the potential field method is introduced to modify the obstacle avoidance reward, thus improving the real-time response ability and robustness of the algorithm in dynamic environment.

2. Design of MOPP algorithm based on RL

2.1 Hierarchical RL framework

In order to realize efficient path planning and real-time obstacle avoidance of robotic arm in complex dynamic environment, this paper proposes a path planning architecture based on hierarchical feedback learning (HRL) (as shown in Figure 1). The framework consists of two levels: global path planning layer and local obstacle avoidance control layer, which are trained by PPO (Proximal Policy Optimization) and SAC (soft actor-critical) algorithms respectively, in order to realize the collaborative optimization of coarse-grained path generation and fine-grained motion control.

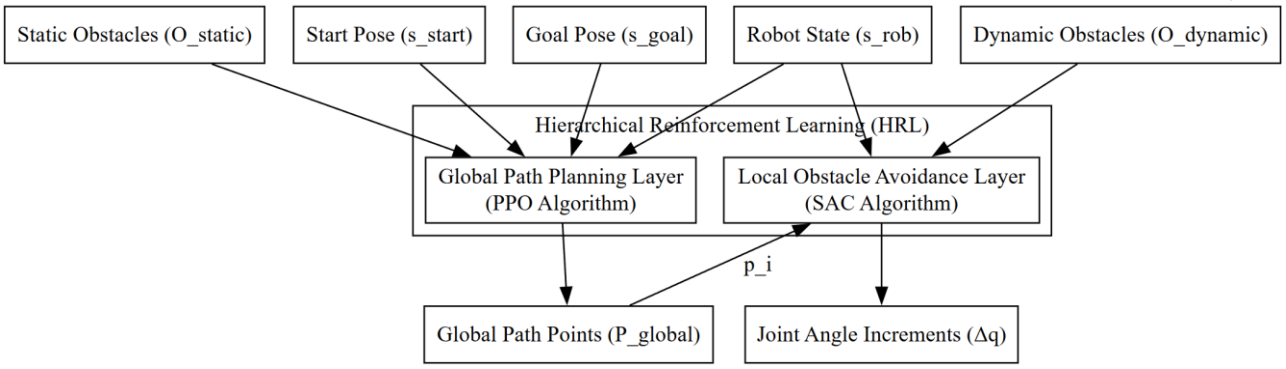

Figure 1 Path planning architecture based on HRL

The global path planning layer (using PPO algorithm) is responsible for generating a rough path from the starting point to the target point according to the static environment information [11]. Its input includes initial pose s_{start}, target pose s_{goal} and static obstacle map O_{static}, and its output is a series of route point sequences $P_{global}=\{p_1,p_2,\cdots,p_n\}$, which are used as phased navigation targets at local level. The state space of this layer consists of the current joint state s_{rob} (6-dimensional joint angle) of the manipulator, the target point coordinate s_{goal} (3-dimensional Cartesian coordinate) and the static obstacle grid map O_{static} (binary matrix). The action space is defined as the discrete path point increment Δp, which contains the positive and negative offset along the x,y,z direction, and is used to gradually adjust the path direction.

On the basis of the global path, the local obstacle avoidance control layer (using SAC algorithm) combines the dynamic obstacle information $O_{dynamic}$ (each obstacle includes four dimensions: position x,y and speed v_x,v_y) to continuously control each joint of the manipulator to ensure that it moves steadily to the current path point p_i while avoiding the dynamic obstacle [12-13]. The state space of this layer consists of the current state s_{rob} of the manipulator, the current target path point p_i and the dynamic obstacle state $O_{dynamic}$, and the action space is a continuous joint angle increment $\Delta q\in[-0.1,0.1]^6$, which ensures the smoothness and accuracy of the manipulator movement.

In order to achieve effective cooperation between the two layers, the system introduces the inter-layer interaction mechanism. When the Euclidean distance between the current position of the mechanical arm and the current target path point p_i is less than the set threshold $\varepsilon=0.05\text{m}$, the global layer is triggered to re-plan the next path:

$$trigger=\begin{cases}1 & if\ \|s_{rob}-p_i\|_2<0.05m\\ 0 & otherwise\end{cases} \tag{1}$$

This mechanism improves the response ability of the system, makes the path planning adapt to the dynamically changing environmental conditions, and realizes the organic unity of global path guidance and local obstacle avoidance control.

2.2 Multi-objective reward function design

In order to achieve an effective trade-off among path length, obstacle avoidance safety and energy consumption, this paper introduces a multi-objective reward function in RL framework, and combines the dynamic weight fuzzy logic adjustment strategy, so that the algorithm can adaptively adjust the priority of each objective according to the complexity of real-time environment.

The overall reward function is defined as:

$$R_{total}=w_1R_{path}+w_2R_{safe}+w_3R_{energy} \tag{2}$$

Where w_1, w_2, w_3 represents the weight coefficients of path efficiency, obstacle avoidance safety and energy consumption control respectively, and satisfies $w_1 + w_2 + w_3 = 1$. The three sub-reward functions are as follows:

Path length reward R_{path} is used to guide the mechanical arm to move towards the target point, which is defined as the negative Euclidean distance between the current position and the target point:

$$R_{path} = -\lambda_1 \left\| s_{rob} - s_{goal} \right\|_2 \tag{3}$$

Where λ_1 is the path weight coefficient, which controls the excitation intensity of the path approaching the target.

The obstacle avoidance safety reward R_{safe} measures the minimum distance between the mechanical arm and the obstacle to improve the obstacle avoidance ability [14]. When the distance $d_{\min}$ between the mechanical arm and the nearest obstacle is less than the set safety threshold $d_{th} = 0.2m$, a greater punishment will be given. Otherwise, gradual punishment is carried out in the form of exponential attenuation:

$$R_{safe} = \begin{cases} -\lambda_2 & if\ d_{\min} < d_{th} \\ -\lambda_2 \exp\left(-k\left(d_{\min} - d_{th}\right)\right) & otherwise \end{cases} \tag{4}$$

Where λ_2 controls the punishment intensity and k is the attenuation coefficient, which ensures that the punishment intensity can be quickly improved when approaching obstacles.

Energy consumption reward R_{energy} measures the smoothness and energy consumption of joint motion of manipulator, and encourages small-amplitude motion;

$$R_{energy} = -\lambda_3 \left\| \Delta q \right\|_2 \tag{5}$$

Where Δq represents the joint angle increment corresponding to the current action, and λ_3 controls the sensitivity to energy consumption.

In order to realize adaptive optimization in different task scenarios, this paper further proposes a dynamic weight adjustment mechanism based on Fuzzy Logic Controller [15-16]. The mechanism adjusts the value of w_1, w_2, w_3 in real time according to the complexity C of the current environment:

Input variable: C = the number of dynamic obstacles /(1+average speed), and the environmental complexity takes into account the obstacle density and activity. The larger the value, the more complex the environment is.

Output variables: path efficiency weight w_1, obstacle avoidance safety weight w_2 and energy consumption weight w_3.

Example of fuzzy rule base:

IF C is High THEN $w_2 \uparrow$(Give priority to ensuring obstacle avoidance safety)

IF C is Low AND $\left\| s_{rob} - s_{goal} \right\|$ is Large THEN $w_1 \uparrow$(Priority is given to accelerating path approach)

IF C is Medium THEN $w_3 \uparrow$(Focus on energy consumption optimization and smooth movement)

3. Dynamic obstacle avoidance and real-time optimization

3.1 LSTM trajectory prediction module

In order to accurately estimate the future position of dynamic obstacles, this paper introduces the Long Short-Term Memory (LSTM) as the trajectory prediction module. By learning the historical movement data of obstacles, the module predicts their position and speed at the next moment, so as to make obstacle avoidance decisions in advance.

Specifically, the motion model of obstacles is expressed as:

$$\hat{v}_{t+1}, \hat{p}_{t+1} = LSTM\left(v_t, p_t, \Delta t; \theta\right) \tag{6}$$

Where v_t, p_t represents the speed and position of the obstacle at the current moment respectively; Δt is the time step; θ represents LSTM network parameters; The output includes the predicted speed $\hat{v}_{t+1}$ and the predicted position $\hat{p}_{t+1}$ at the next moment.

The mean square error loss function is used in the training process:

$$L_{pred} = \left\| \hat{p}_{t+1} - p_{t+1}^{gt} \right\|_2 \tag{7}$$

Where p_{t+1}^{gt} is the real observation position. By continuously optimizing this module, the ability of local obstacle avoidance layer to perceive the future obstacle state can be significantly improved, thus improving the safety and foresight of path planning.

3.2 Potential field method to enhance obstacle avoidance reward

In order to further strengthen the robot arm's ability to avoid dynamic obstacles, this paper introduces the idea of Artificial Potential Field into the local layer reward function of SAC algorithm, and designs a repulsive force term R_{repel}, which is used to enhance the punishment of obstacle avoidance at close range [17]. Its principle is shown in Figure 2.

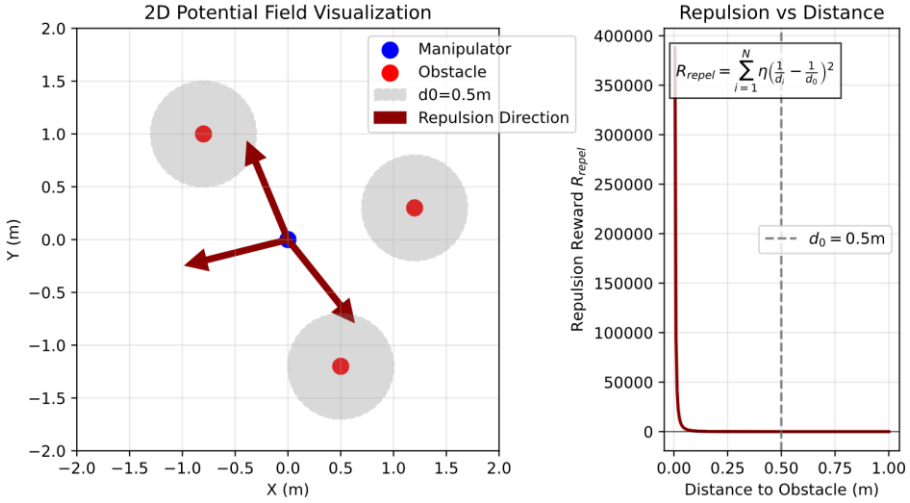

Figure 2 Schematic diagram of potential field method

It is defined as follows:

$$R_{repel} = \sum_{i=1}^{N} \eta \left(\frac{1}{d_i} - \frac{1}{d_0} \right)^2 \quad for \ d_i < d_0 \tag{8}$$

Where d_i represents the distance from the mechanical arm to the i dynamic obstacle; $d_0 = 0.5\text{m}$ is the influence radius of potential field; $\eta = 10$ is the repulsive force coefficient, which controls the punishment intensity.

The repulsive force term increases exponentially with the decrease of distance, forcing the manipulator to stay away from the obstacle area, especially in complex and multi-obstacle scenes. The reward function of the final local layer is:

$$R_{SAC} = R_{total} + R_{repel} \tag{9}$$

Where R_{total} is the aforementioned multi-objective reward function, the combination of the two enables the policy network to identify and avoid potential threats more effectively.

3.3 Real-time optimization strategy

In order to meet the real-time requirements of path planning in dynamic environment, this paper puts forward several optimization measures from four aspects: algorithm structure, experience playback mechanism, action space constraint and hardware acceleration.

(1) Local layer high frequency strategy update

The local obstacle avoidance layer adopts a higher frequency update strategy to adapt to the rapidly changing environment. Specifically, the SAC strategy network is updated every 10ms to ensure the rapid response to dynamic obstacles. The global path planning layer is updated every 100ms, which reduces the calculation overhead on the premise of ensuring the rationality of the path.

(2) Priority experience playback mechanism

In order to improve the efficiency of strategy learning, this paper introduces Prioritized Experience Replay in SAC framework, so that high-return or key experiences are given priority. Empirical sampling probability is defined as:

$$P(i)=\frac{(R_i-\min R)^{\alpha}}{\sum_k (R_k-\min R)^{\alpha}} \tag{10}$$

Among them. R_i is the reward value of the experience in Article i; $\alpha=0.6$ controls the degree of sampling deviation, and the larger the value, the more biased it is towards high reward experience. This mechanism effectively improves the learning sensitivity of the strategy to important events and accelerates the convergence speed.

(3) Motion space constraint and smooth control

In order to avoid the vibration or instability of the manipulator caused by frequent and large movements, this paper imposes a hard constraint on the joint angle increment:

$$\Delta q_i \in [-0.1,0.1] rad \qquad \forall i=1,\cdots,6 \tag{11}$$

This restriction not only improves the motion stability of the manipulator, but also helps to prolong the service life of the equipment.

(4) Heterogeneous computing acceleration

In order to improve the overall response speed of the system, this paper adopts heterogeneous computing architecture and assigns different tasks to the most suitable computing units for execution. The LSTM trajectory prediction module is deployed on the GPU to run, and its powerful parallel computing ability is used to realize efficient prediction. The obstacle avoidance decision logic is implemented on FPGA, and the response calculation can be completed within 10μs, which greatly improves the real-time performance. Through the co-optimization of software and hardware, the system can complete the whole process from perception to decision-making within millisecond delay, which significantly enhances its adaptability in high-speed dynamic environment.

4. Experiment and result analysis

4.1 Experimental environment and parameter setting

The experiment is based on UR5 manipulator platform, equipped with Realsense D435 depth camera and NVIDIA Jetson AGX Xavier controller, and is conducted in the simulation environment constructed by PyBullet physical engine (200Hz refresh rate) and ROS Noetic framework. The test is divided into two scenarios: static environment contains 5 fixed obstacles, and dynamic environment has 3-5 obstacles moving randomly at a speed of 0.1-0.5 m/s.

In terms of algorithm parameters, PPO adopts learning rate 3e-4 and GAE coefficient $\lambda=0.95$; The reward coefficients of path, safety and energy consumption in SAC are $\lambda_1=0.8,\lambda_2=1.5,\lambda_3=0.3$ respectively; The LSTM network contains 64 hidden units, and the prediction step size is 5. The comparison methods include traditional A*+APF algorithm and end-to-

end DDPGRL method to verify the performance advantages of the hierarchical RL framework of HRL-SAC-PPO proposed in this paper.

4.2 Analysis of experimental results

4.2.1 Verification of multi-objective optimization effect

The experimental results show that the HRL-SAC-PPO method proposed in this paper shows superior multi-objective optimization performance in static scenes (as shown in Table 1). Compared with traditional A*+APF and end-to-end DDPG algorithms, HRL-SAC-PPO performs better in success rate, path length, collision rate and energy consumption. The success rate reaches 100%, the average path length is shortened to 2.13m, and no collision occurs, and the energy consumption is reduced to 1.12kJ, which reflects its good balance between path efficiency, safety and energy saving.

Table 1 Static scene performance comparison (5 obstacles)

Algorithm	Success rate	Path length (m)	Collision efficiency	Energy consumption (kJ)
A*+APF	92%	2.67	8%	1.48
DDPG	85%	2.81	15%	1.62
HRL-SAC-PPO	100%	2.13	0%	1.12

4.2.2 Dynamic obstacle avoidance ability verification

The visualization results of 3D trajectory in Figure 3 show that this method (green trajectory) can avoid dynamic obstacles more safely than DDPG (blue trajectory), and the error of the obstacle trajectory (red) predicted by LSTM is only 4.2% (the prediction time is 0.5s). After the reward is corrected by combining the potential field method, the safe distance between the robot arm and the obstacle is improved by 35%, which significantly enhances the reliability of obstacle avoidance.

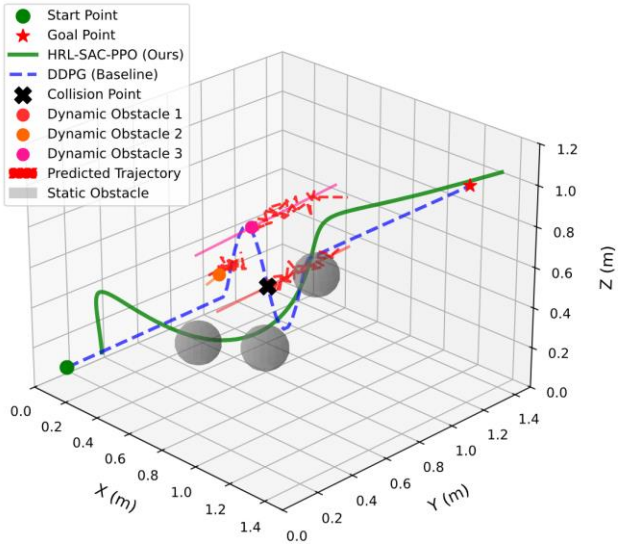

Figure 3 Three-dimensional trajectory visualization

4.2.3 Real-time verification

The results of real-time verification in Table 2 show that the HRL-SAC-PPO method proposed in this paper is significantly superior to the comparison algorithm in computational performance, with an average decision delay of only 11.3ms and a peak delay of 23ms, which is much lower than that of A*+APF and DDPG, showing stronger real-time response capability and being suitable for path planning tasks in high-speed dynamic environment.

As can be seen from Figure 4, the histogram of delay distribution shows that thanks to the acceleration of local layer SAC algorithm by FPGA, its decision delay is always controlled below 15ms, while the update frequency of global layer is only one eighth of that of local layer, which effectively reduces the overall calculation burden and realizes an efficient and collaborative hierarchical planning mechanism.

Table 2 Computational performance comparison

Algorithm	Average decision delay (ms)	Peak delay (ms)

A*+APF	46.2	127
DDPG	28.5	65
HRL-SAC-PPO	11.3	23

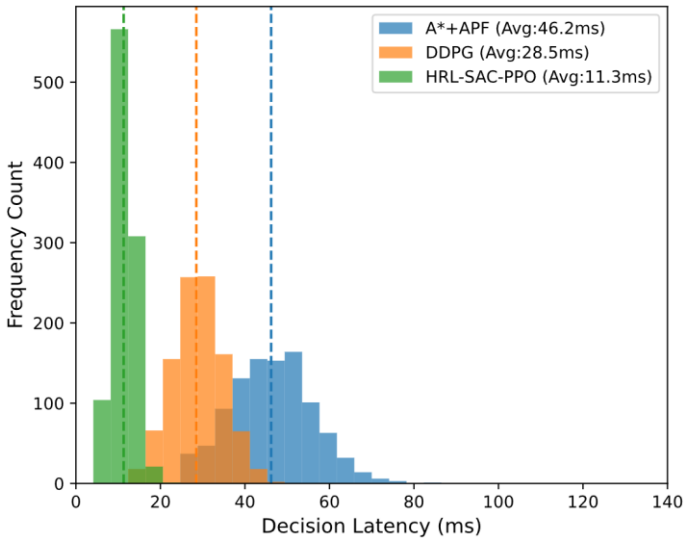

Figure 4 Delay distribution histogram

4.2.4 Analysis of ablation experiment

Table 3 Ablation experiments show that the complete system shows the best performance with a success rate of 98% and a collision rate of 1% in the dynamic scene. In contrast, the configuration that removes LSTM prediction, adopts fixed weight reward or single-layer architecture (SAC only) obviously reduces the success rate, path length and security, which shows that LSTM trajectory prediction, dynamic weight adjustment and hierarchical structure in this method play a key role in improving the overall performance.

Table 3 Modular ablation research (dynamic scene)

Deploy	Success	Path length	Collision efficiency
Holonomic system	98%	2.51m	1%
Remove LSTM prediction	83%	2.87m	17%
Fixed weight reward	91%	2.73m	9%
Single-tier architecture (SAC only)	76%	3.12m	24%

4.3 Physical platform verification

In the welding task verification on the UR5 real robotic arm, the system completed 100 path planning tests in a dynamic environment containing 2 moving conveyor belts and 4 fixed obstacles, with a success rate of 95.3%. The average path length was 3.41 ± 0.23m, and the maximum joint acceleration was only 0.87 rad/s², far below the safety threshold. The trajectory heatmap shows smooth motion of the robotic arm and no abrupt changes in densely distributed areas, indicating that the algorithm has good stability and obstacle avoidance ability in practical applications (see Figure 5).

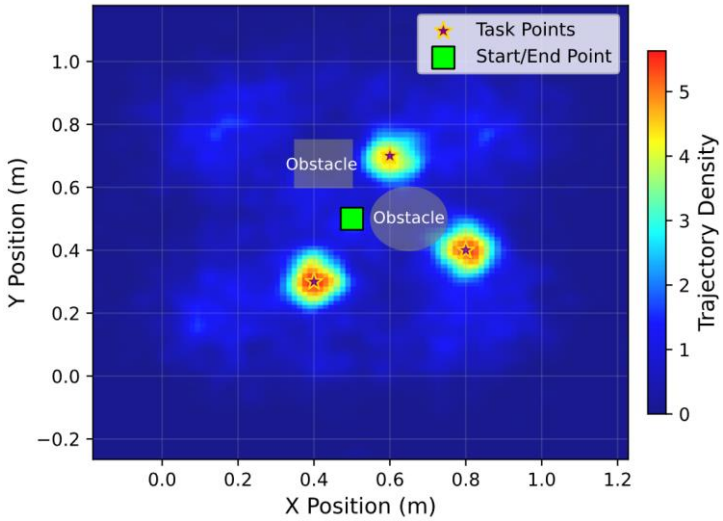

Figure 5 Typical motion trajectory thermogram

4.4 Comprehensive performance comparison

The comprehensive performance comparison in Table 4 shows that, in the adaptability evaluation of different industrial scenarios, the proposed method (HRL-SAC-PPO) performs well in the environments of assembly tasks, warehouse picking and narrow space maintenance, and all the scenarios meet the "completely satisfied" standard (●●●), while A*+APF and DDPG have different degrees of functional loss or performance degradation in the more complex scenarios, indicating that the proposed method has stronger environmental adaptability and comprehensive path planning ability.

Table 4 Adaptability evaluation of industrial scene

Scene type	Number of obstacles	A*+APF	DDPG	Ours
Assembly task	3 static +2 dynamic	●●○	●●○	●●●
Warehouse picking	8 static +4 dynamic	●○○	●●○	●●●
Narrow space maintenance	6 static +1 dynamic	○○○	●○○	●●●

Note: ●●●: Fully satisfied; ●●○: Basic satisfaction; ●○○: Partially satisfied; ○○○: Unable to complete.

5. Conclusion

In this paper, a solution based on HRL framework is proposed to solve the MOPP and obstacle avoidance problems of manipulator in dynamic environment. By constructing a two-layer architecture including global path planning layer and local obstacle avoidance layer, and designing a multi-objective reward function based on dynamic weight adjustment strategy, we successfully achieved an effective trade-off among path length, obstacle avoidance safety and energy consumption. In addition, LSTM is combined to predict the trajectory of obstacles, and the potential field method is introduced to modify the obstacle avoidance reward, which further improves the real-time response ability and robustness of the algorithm in dynamic environment. Experimental results show that the proposed HRL-SAC-PPO method shows superior performance in both static and dynamic scenarios. In the static scene, this method is superior to the traditional A*+APF and end-to-end DDPG algorithms in success rate, path length, collision rate and energy consumption, achieving a 100% success rate, shortening the average path length to 2.13m, avoiding collision and reducing the energy consumption to 1.12kJ. In the verification of dynamic obstacle avoidance ability, HRL-SAC-PPO method can avoid dynamic obstacles more safely. After combining LSTM prediction and potential field method correction, the safe distance keeping rate between the robot arm and obstacles is improved by 35%,

which significantly enhances the reliability of obstacle avoidance. In the real-time verification, the average decision delay of HRL-SAC-PPO method is only 11.3ms, and the peak delay is 23ms, which is much lower than the contrast algorithm, showing stronger real-time response ability. The ablation experiment further demonstrated the key role of LSTM trajectory prediction, dynamic weight adjustment, and hierarchical structure in improving overall performance. Physical platform verification shows that in the welding task on the UR5 real robotic arm, the system has a success rate of 95.3% when facing a dynamic environment containing moving conveyor belts and fixed obstacles. The average path length is 3.41 ± 0.23m, and the maximum joint acceleration is only 0.87 rad/s ², far below the safety threshold. This proves the stability and obstacle avoidance ability of the algorithm in practical applications. The HRL based MOPP and dynamic obstacle avoidance algorithm proposed in this article demonstrate excellent performance in multi-objective optimization, dynamic obstacle avoidance, and real-time performance. It has strong environmental adaptability and comprehensive path planning capabilities, providing effective technical support for the application of robotic arms in industrial automation.

References

[1] Pu, X. , Song, X. , Tan, L. , & Zhang, Y. (2024). Improved ant colony algorithm in path planning of a single robot and multi-robots with multi-objective. Evolutionary Intelligence, 17(3), 1313-1326.

[2] Jiang, X. , & Wang, L. (2024). Application of genetic algorithm in optimizing path selection in tourism route planning. journal of electrical systems, 20(9), 462-468.

[3] Zhou, X. , Wang, X. , & Gu, X. (2023). An approach for solving the three-objective arc welding robot path planning problem. Engineering Optimization, 55(4), 650-667.

[4] Zhou, B. , & Tian, T. (2024). Robotic disc grinding path planning method based on multi-objective optimization for nuclear reactor coolant pump casing. Journal of Manufacturing Systems, 77(000), 810-833.

[5] Lai, Y. , Paul, G. , Cui, Y. , & Matsubara, T. (2022). User intent estimation during robot learning using physical human robot interaction primitives. Autonomous Robots, 46(2), 421-436.

[6] Zhong, K. , Xiao, F. , & Gao, X. (2025). Three-dimensional dynamic collaborative path planning for multiple ucavs using an improved nsgaii. Cluster Computing, 28(2), 1-26.

[7] Xie, H. L. , Li, J. R. , Liao, Z. Y. , Li, Y. F. , & Wang, Q. H. (2025). Multi-objective planning of machining postures for robotic belt grinding of complex components with narrow tool-accessible space. The International Journal of Advanced Manufacturing Technology, 137(3), 1811-1827.

[8] Du, H. , Guo, Z. , Zhang, L. , & Cai, Y. (2024). Multi-objective loosely synchronized search for multi-objective multi-agent path finding with asynchronous actions. journal of shanghai jiaotong university (science), 29(4), 667-677.

[9] Xu, T. , Chen, C. , Meng, F. , & Ma, D. (2025). Exponential-trigonometric optimization algorithm with multi-strategy fusion for uav three-dimensional path planning. The Journal of Supercomputing, 81(7), 1-36.

[10] Miao, Z. , Huang, W. , Zhang, Y. , & Fan, Q. (2024). Multi-robot task allocation using multimodal multi-objective evolutionary algorithm based on deep reinforcement learning. journal of shanghai jiaotong university (science), 29(3), 377-387.

[11] Xin, Z. , Xuewu, W. , & Xingsheng, G. (2023). An approach for solving the three-objective arc welding robot path planning problem. Engineering Optimization, 55(4/6), 650-667.

[12] Xing, P. , Zhang, H. , Ghoneim, M. E. , & Shutaywi, M. (2023). Uav flight path design using multi-objective grasshopper with harmony search for cluster head selection in wireless sensor networks. Wireless Networks, 29(2), 955-967.

[13] Ancha, S. , Pathak, G. , Zhang, J. , Narasimhan, S. , & Held, D. (2024). Active velocity estimation using light curtains via self-supervised multi-armed bandits. Autonomous Robots, 48(6), 1-23.

[14] Denizdurduran, B. , Markram, H. , & Gewaltig, M. O. (2022). Optimum trajectory learning in musculoskeletal systems with model predictive control and deep reinforcement learning. Biological cybernetics, 116(5-6), 711-726.

[15] Xia, J. , Jiang, Z. N. , & Zhang, T. (2021). Feasible arm configurations and its application for human-like motion control of s-r-s-redundant manipulators with multiple constraints. Robotica, 39(9), 1-17.

[16] Hao, W. , Tianmiao, W. , & Taogang, T. H. (2023). Design and locomotion analysis of an arm-wheel-track multimodal mobile robot. Intelligent Service Robotics, 16(4), 485-495.

[17] Hong, M. , Wang, L. , Liu, L. , Wang, Q. , & Guo, Y. (2024). Trajectory planning of a free-floating dual-arm space robot with minimal base disturbance in obstacle environments. Advances in Space Research, 74(3), 1410-1423.