Journal of Electronic Research and Application, 2021, Volume 5, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

# Applications and Challenges of Deep Reinforcement Learning in Multi-robot Path Planning

**Tianyun Qiu*, Yaxuan Cheng**

School of Information, Beijing Wuzi University, Beijing 101149, China

***Corresponding author:** Tianyun Qiu, 946992724@qq.com

**Abstract:** With the rapid advancement of deep reinforcement learning (DRL) in multi-agent systems, a variety of practical application challenges and solutions in the direction of multi-agent deep reinforcement learning (MADRL) are surfacing. Path planning in a collision-free environment is essential for many robots to do tasks quickly and efficiently, and path planning for multiple robots using deep reinforcement learning is a new research area in the field of robotics and artificial intelligence. In this paper, we sort out the training methods for multi-robot path planning, as well as summarize the practical applications in the field of DRL-based multi-robot path planning based on the methods; finally, we suggest possible research directions for researchers.

**Keywords:** MADRL; Deep reinforcement learning; Multi-agent system; Multi-robot; Path planning

***Publication date:*** November 2021; ***Online publication:*** November 30, 2021

## 1. Introduction

Multi-agent systems (MAS) are distributed computing approaches that can be used to a wide range of issues, including robotics, distributed decision making, traffic control, and corporate management. MADRL is a DRL application for MAS, in which agents interact in the same environment, with each agent formulating a policy at each time step using DRL algorithms to collaborate with other agents to achieve a goal. Path planning is the foundation for mobile robot navigation and control, and several robots can complete big and complicated tasks more effectively than a single robot. Many classical methods are used to solve path planning problems for multiple robots, such as graph search algorithms and heuristic algorithms, commonly known as A* algorithms, artificial potential field methods, and ant colony optimization algorithms. Next, path planning methods are classified into centralized and decentralized methods according to the structure of the robot.

## 2. Multi-agent deep reinforcement learning

DRL is considered an important component for building general-purpose artificial intelligence systems and has been successfully combined with techniques such as search and planning. It has also been combined with MAS in recent years to form the emerging field of MADRL. This section provides a brief classification of MADRL.

Analysis of emergent behaviors is a direct extension of the single-agent DRL algorithm in a multi-agent environment, where there is no communication between the agents, either for training or execution. In a multi-agent environment, all the agents independently interact with the environment and form behavioral strategies independently. In this approach, the number of actions grows exponentially, which

makes the problem difficult to solve.

Learning communication is the sharing of information and experiences observed by each of the agents through communication. It is assumed that the agent interacts with other agents through its own local observations with explicit information between them. After training and learning, the agents must make behavioral decisions about which agents to cooperate with based on the transmitted information. This approach is suitable for both fully cooperative tasks and incompletely observed environments.

Learning cooperation is a problem of cooperative strategies for agents to learn complex, partially observable domains without explicit communication. The cooperative problem can be formulated as a decentralized partially observable Markov decision process (Dec-POMDP), where all the agents try to maximize the discounted sum of joint rewards, and the agents do not communicate explicitly and learn cooperative behavior only from their respective observations.

Agent modelling is a technique in which an agent anticipates the behavior of a modelled agent by using models to reason about the behavior of other agents. The Deep Recurrent Opponent Network (DRON) consists of two networks: one to evaluate the Q value and another to learn the adversary agents' strategy, as well as many expert networks operating simultaneously to improve the algorithm's capabilities. The Agents modelling agents' algorithms are more resilient and have a broader range of application scenarios, but modelling complexity is high and practical applications are limited. Deep reinforcement learning-based path planning for multi-agent system.

### 2.1. Classification of multi-robot path planning training methods

The multi-robot path planning training methods can be grouped into three:

(1) Centralized. The centralized training scheme assumes that the actions of all the agents are determined by a central server that knows the intentions of all the agents. The centralized approach has three limitations. First, the computational complexity of centralized control and scheduling is high as the number of robots increases; second, the centralized approach cannot scale to large-scale systems because of communication problems; and third, the centralized system is susceptible to failures in the central server, communications between robots, or sensors on any individual robot.

(2) Decentralized. Decentralized systems are more flexible, more efficient in task completion, and more fault-tolerant. In decentralized training methods each agent learns its own strategy by considering the observable states of other agents, such as shape, velocity and position, maps its own observations to actions to make decisions independently, and each agent calculates paths individually and then adjusts these paths to avoid conflicts. Decentralized approaches are also divided into reaction-based and trajectory-based approaches.

(3) Shared parameter. The path planning is decentralized, but the learning process is centralized. This approach is able to exploit the capabilities of each robot and compensate for the shortcomings of centralized path planning, thus adapting to unknown planning environments, while the method does not require knowledge of the dynamics model of the environment, has no communication requirements, and can be used in both cooperative and competitive environments. However, it also requires a central coordinator, it is not fully autonomous, and it is poorly scalable. Therefore, it does not support the training of a large number of agents and has a long training period.

### 2.2. DRL multi-robot path planning method

**Table 1** summarizes the DRL multi-robot path planning methods and the advantages and limitations of each method. From the information in **Table 1**, it can be summarized that shared parameter type algorithms such as MADDPG and ME-MADDPG can be used in dynamic and complex environments [1-4]; decentralized architectures such as DQN and DDQN can be considered in stable environments [5-7]; large

robotic systems facing a large number of dynamic obstacles can be considered using algorithms such as A2C, A3C and TDueling [8-11].

**Table 1.** DRL multi-robot path planning method

| References | Objectives | Main Algorithm | Advantages | Limitations |
|---|---|---|---|---|
| [1] | Solving the problem of slow learning of decentralized path planning in unknown environments. | Q-learning, kernel smoothing, neural networks | 1. Overcoming the disadvantages of slow and time-consuming learning in RL.<br>2. Not seriously affected by sensor fluctuations. | The efficiency of the method decreases as the number of agents increases. |
| [2] | Solve the problem of multi-robot collision-prone under vision-based. | DQN | 1. No need to manually mark feature values.<br>2. High collision avoidance success rate. | 1. Slow learning efficiency.<br>2. Unable to scale to large-scale robot teams.<br>3. Can't detected Target information. |
| [3] | Solving Path Planning Problems in Warehouse Environments. | DQN | Using a priori knowledge and rules to guide path planning, which improves the learning efficiency of the algorithm. | Suitable for small-scale robotics teams. |
| [4] | Solving path planning problems in hybrid dynamic environments. | A2C | 1. Does not require homogeneous agent;<br>2. Higher success rate and more stable performance. | Only simulation experiments are conducted, and the effectiveness is not proven in a real environment. |
| [5] | Solving the path planning problem of multi-robot systems in the process of information sharing. | TDueling | 1. Solves the problem of source competition and obstacle avoidance;<br>2. The algorithm has higher accuracy and better robustness. | In a three-dimensional environment, the training time and complexity are significantly increased and generalization is lacking. |
| [6]<br>[7] | Addressing Lifelong MAPF in High Density Structured Environments | IL, A3C, LSTM | 1. Suitable for large-scale robotic teams;<br>2. Short training time. | The robots do not communicate with each other and do not take advantage of the collaboration of multi-robot systems. |
| [8] | Solving the collision problem in path planning. | MAPPO | 1. Effectively apply to real-world tasks.<br>2. No need to create a barrier map of the environment. | 1. Algorithm training time is long.<br>2. Validation on only a few teams of agents. |

| | | | | |
|---|---|---|---|---|
| [9] | Solving the problem of target assignment and path planning for collaborative multi-UAV control systems. | MADDPG | 1. Dealing effectively with dynamic environments.<br>2. Real-time planning performance can be guaranteed. | 1. Long training time.<br>2. Validation on only a few teams of agents. |
| [10] | Solving UAV ground target tracking problems in obstacle environments. | MADDPG | 1. High real-time planning.<br>2. No need to converge in the optimization process. | 1. Long training time;<br>2. Vision-based DRL methods cannot detect the target location information. |
| [11] | Solving multi-agent body motion planning problems in dynamic and complex environments. | ME-MADDPG | 1. Fast convergence and high convergence value.<br>2. Good stability and adaptability. | Validity validated on only a few teams of agents. |

## 3. Challenges and research directions of DRL-based multi-robot path planning

Recently, the research on DRL-based multi-robot path planning has made great progress, but there are still some problems in the algorithm because of the complexity of the environment, so the future research on DRL-based multi-robot path planning can be carried out from the following aspects.

(1) Improved generalization capability. DRL-based multi-robot path planning problems often use neural networks to process sensor data, and although the results perform well in the training environment, they lack generalizability from one environment to another and from the simulated environment to the real environment.

(2) Improving the sampling efficiency of the algorithm thus speeding up the learning speed. When the environment is more complex, the amount of data will be larger and the number of interactions between the intelligent body and the environment will be more. Most of the existing studies use empirical playback techniques to improve the sampling efficiency of the algorithm, but cannot guarantee the data in real time.

(3) Set a more efficient reward function and thus improve the exploration efficiency. In the path planning problem, model-free DRLs will rely on exploration to find the optimal policy, but because there is sparsity in rewards, the agent will only be rewarded when it reaches the goal, so the agent will perform many meaningless exploration behaviors in the environment. The reward function must be designed precisely, otherwise the reward function will be learned over-fittingly.

(4) Model-based learning. Model-free DRL algorithms require a large amount of sampled data for training, which is often difficult to obtain through interaction. Therefore, it is possible to consider using existing data from realistic environments to build environmental models, which can then be used to train the agents. Combining model-based algorithms and model-free deep reinforcement learning path planning problems is one of the key ways to improve the efficiency of reinforcement learning in the future.

## 4. Conclusion

In recent years, mobile robotics has been a prominent study topic, and path planning is a key technology for robots and self-driving cars. The literature on DRL-based multi-robot path planning in the previous five years is briefly summarized in this study. In addition, future research prospects for DRL-based multi-robot

path planning are discussed. Many DRL path planning algorithms are still in the laboratory stage, and there are still big disparities between them in terms of current development state. Also, path planning conditions in real environments, including environmental uncertainties, communication and data delays, which pose many challenges for the DRL "trial-and-error" training approach, and also for future DRL research in the field of path planning. These are challenges for DRL's "trial-and-error" training approach, and are also problems to be solved in future DRL research in the field of mobile robot path planning.

## Disclosure statement

The author declares no conflict of interest.

## References


[1] Lin J, Yang X, Zheng P, et al., 2019, End-To-End Decentralized Multi-Robot Navigation in Unknown Complex Environments Via Deep Reinforcement Learning 2019, IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, 2493-2500.

[2] Qie H, Shi D, Shen T, et al., 2019, Joint Optimization of Multi-UAV Target Assignment and Path Planning based on Multi-Agent Reinforcement Learning. IEEE access, 7: 146264-146272.

[3] Li B, Wu Y, 2020, Path Planning for UAV Ground Target Tracking Via Deep Reinforcement Learning. IEEE Access, 8: 29064-29074.

[4] Wu D, Wan K, Gao X, et al., 2021, Multiagent Motion Planning Based on Deep Reinforcement Learning in Complex Environments 2021, 6th International Conference on Control and Robotics Engineering (ICCRE). IEEE, 123-128.

[5] Cruz DL, Yu W, 2017, Path Planning of Multi-Agent Systems in Unknown Environment with Neural Kernel Smoothing and Reinforcement Learning. Neurocomputing, 233: 34-42.

[6] Xin J, Zhao H, Liu D, et al., 2017, Application of Deep Reinforcement Learning in Mobile Robot Path Planning 2017, Chinese Automation Congress (CAC). IEEE, 7112-7116.

[7] Yang Y, Juntao L, Lingling P, 2020, Multi-Robot Path Planning based on a Deep Reinforcement Learning DQN Algorithm. CAAI Transactions on Intelligence Technology, 5(3): 177-183.

[8] Liu Z, Chen B, Zhou H, et al., 2020, Mapper: Multi-Agent Path Planning with Evolutionary Reinforcement Learning in Mixed Dynamic Environments 2020 IEEE, RSJ International Conference on Agent Robots and Systems (IROS). IEEE, 11748-11754.

[9] Wang D, Deng H, 2021, Multirobot Coordination with Deep Reinforcement Learning in Complex Environments. Expert Systems with Applications, 180: 115128.

[10] Sartoretti G, Kerr J, Shi Y, et al., 2019, Primal: Pathfinding Via Reinforcement and Imitation Multi-Agent Learning. IEEE Robotics and Automation Letters, 4(3): 2378-2385.

[11] Damani M, Luo Z, Wenzel E, et al., 2021, PRIMAL $ _2 $: Pathfinding Via Reinforcement and Imitation Multi-Agent Learning-Lifelong. IEEE Robotics and Automation Letters, 6(2): 2666-2673.