Uav Path Planning and Obstacle Avoidance Based on Deep Learning and Reinforcement Learning

Zhixi Shi *

School of Computer Science and Technology,Changsha University of Science & Technology, Changsha, China

* Corresponding Author Email: 202308010305@csust.edu.cn

Abstract. With the rapid development of the Internet of Things and low-altitude economy, unmanned aerial vehicles (Uavs) have been widely used in data collection, environmental monitoring and other fields.Traditional path planning algorithms have defects such as insufficient generalization ability and poor real-time performance in complex and dynamic environments. Deep reinforcement learning technology provides a new solution for autonomous decision-making and collaborative operation of unmanned aerial vehicles (Uavs). In this paper, the UAV path planning technology based on deep learning and reinforcement learning is systematically reviewed. In the aspect of single UAV, the improved Q-Learning algorithm realizes fast local replanning through dynamic exploration factor and artificial potential field reward function. DDPG, DySAC, SMLTO and other algorithms improve the safety of obstacle avoidance and trajectory smoothness in three-dimensional dynamic environment. In terms of multiple Uavs, the MRF-DQN algorithm combines preferential experience replay and maximum reward frequency function to improve the efficiency of collaborative task completion. The DPPDQN algorithm realizes efficient task allocation and path optimization in large-scale regions through double pre-segmentation and CNN feature extraction. The experimental results show that the above algorithms are superior to the traditional methods in terms of collision rate, planning delay, energy efficiency and task completion rate, which provides reliable technical support for multi-UAV cooperative operation.

Keywords: Deep learning; Reinforcement learning; Unmanned aerial vehicle; Path planning.

1. Introduction

In recent years, unmanned aerial vehicle (UAV) technology has evolved from a single operation to multi-UAV collaboration, which has outstanding value in disaster rescue,agricultural monitoring and other scenarios. The number of registered Uavs in China reached 958,000 in 2022, and the global market size is expected to exceed 52.3 billion US dollars in 2025. However,problems such as "black flight" interference, dynamic obstacle avoidance in complex environments, and cooperative allocation of multiple Uavs restrict large-scale application. Traditional path planning algorithms rely on global prior information,which lack adaptability in sudden threats and large-scale scenarios. Deep learning and reinforcement learning autonomously learn optimal strategies through trial-and-error interaction between agents and the environment, which provide technical paths to solve the above problems. In the field of UAV path planning, the existing research has formed a multi-dimensional technical system, which provides the basis for path planning based on traditional mathematical algorithms and intelligent optimization algorithms.

However, there are limitations such as dimension disaster and local optimization in dynamic and complex environments. The deep learning and reinforcement learning algorithms show significant advantages, such as the improved Q-Learning algorithm significantly reduces the replanning time of sudden threats [1], and the DPPDQN algorithm maintains a high task completion rate in large-scale areas through regional pre-segmentation [2]. Dyna Q-Learning combined with adaptive fuzzy PID realizes zero steady-state error trajectory tracking [3]. This paper systematically reviews the research status of core technologies in the field of UAV. It covers single and multiple UAV path planning and obstacle avoidance technologies in complex situations. It combs the advantages, limitations and typical research results of improved Q-Learing algorithm, DRL-based optimization method, DySAC algorithm, SMLTO algorithm, D-PSO algorithm, improved DQN algorithm and multi-UAV path

planning algorithm based on deep reinforcement learning. This paper provides a comprehensive research background support for subsequent technology optimization and innovation.

2. Path planning for single UAV

2.1. Improved Q-Learing Algorithm

Aiming at the real-time performance and safety requirements of UAV path planning in a sudden threat environment, Han Junchi proposed an improved Q-Learning algorithm in his research [1] . It designed a dynamic exploration factor strategy that adjusted adaptively with the standard deviation coefficient of iteration steps, and integrated the artificial potential field method to construct the reward function of sudden threat. The local path re-planning mechanism is used to realize the rapid response to sudden threats. The algorithm achieves stable convergence after 700 rounds of training in a 20×20 grid environment, and there is no negative reward fluctuation in the convergence process. In the face of sudden threats, the replanning time is only 2.12s, the planning path is shortened by 190.15m compared with the traditional Q-Learning, and the collision rate is controlled within 5%. This study shows that the dynamic exploration factor can effectively balance the exploration and utilization capabilities of the algorithm, and solve the defects of slow convergence and poor stability of traditional Q-Learning. The local re-planning mechanism gives UAV A better response speed to sudden threats, and its real-time performance is significantly better than the 57.33s re-planning delay of A * algorithm. However, the algorithm still has obvious limitations. It relies on the raster environment modeling method, which is difficult to be fully adapted in unstructured and complex terrain, and the repulsion gain coefficient η in the reward function needs to be manually debugged, which limits the generalization ability to different threat density scenarios.

In the study of Yin Y et al. [4], also based on the Q-Learning algorithm, the path planning optimization in the emergent threat environment was realized by introducing the "threat avoidance priority" mechanism, which quantize the threat level of the emergent threat as the weight of the reward function. In the decision-making process of obstacle avoidance, the UAV can dynamically adjust the priority of obstacle avoidance according to the severity of threats. In the experimental scenarios with three types of sudden threats, the algorithm reduces the UAV track conflict rate by 37%, stabilizes the replanning delay within 3s, and the task completion rate reaches 92%. This case verifies that the Q-Learning algorithm can effectively adapt to multiple types of emergent threat scenarios through reward function optimization, improve the pertinence and rationality of obstacle avoidance decision-making, and provide a feasible idea for path planning in different threat level environments. However, this study does not make full use of the multi-aircraft collaboration information in the single UAV scenario. When facing dense emergent threats, the UAV is easy to fall into the local optimal path, and it is difficult to realize the global optimal obstacle avoidance decision.

For the path planning problem in complex 3D environments with mixed static and dynamic obstacles, Yasmine Zamoum et al. systematically compared the performance of two reinforcement learning algorithms, Deep Q-learning and Dyna Q-learning [3]. In the framework of Dyna Q-learning, the training starts near the target point and gradually expands the exploration area, instead of randomly initializing in the whole environment. This strategy effectively overcomes the problem of slow convergence in the sparse reward environment of traditional reinforcement learning. The experiment was carried out in a 3D grid environment with random obstacles. The results show that the standard Deep Q-learning algorithm takes about 16 hours to train and cannot stably reach the target, while the improved Dyna Q-learning algorithm only takes about 144 seconds to complete the training, which is 29% shorter.

And it can stably generate obstacle-free feasible paths. This case proves that Dyna Q-learning based on model learning and planning mechanism is significantly superior to Deep Q-learning without model in terms of sample utilization efficiency and convergence speed. Its progressive training strategy can quickly guide the agent to focus on high-value areas, which is especially suitable for UAV navigation tasks with discrete state space. However, the performance of Dyna Q-learning

relies heavily on grid environment modeling, which has insufficient scalability in continuous or high-dimensional state space. At the same time, the deterministic environment model assumption adopted by Dyna Q-learning may not be applicable in highly dynamic or random real scenarios, and it is difficult to deal with sudden threats or nonlinear dynamic disturbances. Therefore, it is necessary to improve the continuity of environmental representation and model adaptability, so as to face the unstructured and strong dynamic real environment with sudden threats.

2.2. Optimization method based on DRL

Aiming at the problems of large environmental uncertainty, high real-time requirements and poor strategy generalization ability in UAV autonomous obstacle avoidance, Fan Hongyue proposed an optimization method based on Deep Reinforcement Learning (DRL) [5]. The core point of the method is to construct a probabilistic graphical model to represent the environment state space and introduce an octree adaptive division strategy. A multi-objective adaptive reward mechanism integrating safety, efficiency and energy consumption indicators is designed, and the rapid obstacle avoidance of sudden threats is realized through the end-to-end policy learning of the DDPG algorithm framework. In the simulated forest environment, the algorithm converges stably after training. In the high-density obstacle scene, the replanning delay in the face of sudden threats is controlled within 48ms, the collision rate is only 4.3%, the energy efficiency is maintained over 86%, and the planning path is 15% shorter than the traditional Artificial Potential Field Method (APF). The study shows that the combination of probabilistic graphical model and octree partition can effectively balance modeling accuracy and computational efficiency, and solve the shortcomings of traditional algorithms in complex environments, such as complex modeling and lack of real-time performance. The fusion of multi-objective reward mechanism and DDPG algorithm gives UAV better robustness for unexpected threat obstacle avoidance. The collision rate of the proposed algorithm is 8.3 percentage points lower than that of the APF algorithm, and the real-time performance is significantly better than the 92ms planning delay of the APF algorithm. However, the performance of the algorithm under extreme weather conditions has not been fully verified, and it is highly dependent on the accuracy of sensor perception. When the threat detection is delayed due to interference such as smoke, the success rate of replanning will drop to 78%, and the generalization ability is greatly affected by environmental interference factors.

2.3. DySAC (Dynamic Soft Actor-Critic) algorithm

For the path planning problem in unknown dynamic obstacle environment, Cao Xinwen proposed the DySAC algorithm [6] in his study, which introduced the LSTM structure to learn the temporal motion trajectory of obstacles, and designed the dual experience replay mechanism with random in the early stage and priority in the later stage. Experiments show that the proposed algorithm performs well in the dynamic simulation environment, with a success rate of 98.6%, a collision rate of only 1.3%, and the fastest convergence speed, which proves its advantages in timing perception, obstacle avoidance safety and training efficiency. However, the algorithm still has limitations such as high model complexity, insufficient adaptability to the real complex environment, and simplified UAV dynamics modeling. Its actual deployment ability needs to be further verified and improved.

2.4. SMLTO algorithm

Tang J proposed a local trajectory optimization algorithm based on safety mechanism called SMLTO. SMLTO considers the unknown area outside the field of view of the UAV sensor as a potentially dangerous space instead of the traditional free space, and introduces a new distance calculation and control point adjustment method. By keeping a preset safe distance threshold between each control point on the whole planned trajectory and the obstacles, the space safety and obstacle avoidance ability of the UAV are significantly improved when it enters the unknown area from the known area. The experimental results show that compared with EGO-Planner and other benchmark algorithms, the proposed method can plan a safer and smoother trajectory, make the UAV find hidden

obstacles in advance in complex scenes such as corners and narrow passages, and improve the success rate of obstacle avoidance under high-speed flight. It proves that the algorithm has significant advantages in improving the adaptability and security of UAV to unknown environments and dynamic obstacles. However, the performance of the SMLTO algorithm is limited by the limited field of view of the airborne sensors, which cannot perceive the lateral and rear obstacles, and depends on the more accurate visual positioning. In the environment with severe motion or texture loss, the planning may fail due to positioning failure.

2.5. D -PSO algorithm

Zhang Wenjing proposed the D-PSO global planning algorithm [8], which skillfully combines the fine search of Dijkstra algorithm with the global optimization ability of improved particle swarm optimization algorithm. In order to solve the problem that the traditional artificial potential field method is easy to get into local deadlock, the virtual target point and the repulsion field correction factor are introduced to realize smooth and reliable local real-time obstacle avoidance. It performs well in the simulation experiment, the detection speed reaches 49.21 frames per second, the model volume is reduced by 80%, and the planned path is significantly improved in convergence speed and smoothness. It is proved that the D-PSO algorithm can effectively balance the accuracy and real-time performance under the limited airborne computing resources, and realize the autonomous closed-loop of UAV from environment perception to safe obstacle avoidance. However, the verification of D-PSO is completed in the simulation environment, lacking the complex interference test in the real environment. At the same time, the obstacle setting in the experiment is idealized, and its generalization ability and robustness still need to be further tested in the complex dynamic real scene.

3. Path planning for multiple Uavs

3.1. Improved DQN algorithm

The MRF-DQN algorithm proposed by Han Junqi provides an effective solution for multi-UAV target convergence path planning [1]. The core innovation of the algorithm is to introduce a preferential experience playback mechanism, and to improve the utilization rate of high-value experience by allocating sampling probability through time-sequence differential error. At the same time, a global maximum reward frequency (MRF) function is designed to optimize the cooperative reward distribution of multiple Uavs, and the flight speed of Uavs is dynamically adjusted by combining the time cooperative strategy. In the experimental verification of three UAV target gathering tasks, the algorithm only needs 4000 rounds to achieve convergence, which is greatly reduced compared with the 7000 rounds of the traditional DQN algorithm. The total path step is controlled within 47, the replanning time under sudden threat is 8.32s, and the time error of multi-UAV cooperation is less than 5%. The study proves that the priority experience replay mechanism can enhance the learning efficiency of the algorithm for key experience, the MRF function effectively avoids the local optimum problem in the process of multi-UAV collaboration, and the time collaboration strategy ensures the realization of accurate synchronous assembly of multiple Uavs, which significantly improves the efficiency and accuracy of collaborative planning. However, the algorithm uses a centralized training architecture, and the computational complexity will increase significantly when the number of Uavs is more than 10. Moreover, the adaptability to the mutation of UAV speed is insufficient, which easily leads to the deviation of coordination time and affects the assembly effect.

In the study of Elfatih et al. [9], in order to solve the communication bottleneck problem of multi-UAV assembly, a distributed communication architecture is integrated on the basis of DQN algorithm, and the global state sharing of traditional centralized architecture is replaced by local information exchange. At the same time,a "collaborative reward function" similar to the MRF function is designed to optimize the action consistency of multiple Uavs. In the experiment of 5 UAV assembly tasks, the proposed algorithm reduces the communication delay by 40%, and the assembly success rate reaches

95%. When a single UAV fails, the fault tolerance rate of the system is improved by 30%. This case verifies the feasibility of reinforcement learning combined with distributed architecture in multi-UAV assembly scenario, which effectively breaks through the communication limitation of centralized architecture through local information exchange and improves the stability of the system in complex communication environment. However, the convergence speed of the distributed information interaction method is 15% slower than that of the centralized architecture, and the reliability of local information transmission is reduced in the strong electromagnetic interference environment, which affects the stability of multi-UAV cooperative assembly.

Gao Y designed the DPPDQN algorithm for large-scale complex environments [2], which combines the double pre-segmentation mechanism and CNN feature extraction, and realizes the cooperative path planning and efficient data collection of multiple Uavs in the scene with dense nodes, many obstacles, and only local observation. Experiments show that the proposed method is superior to traditional algorithms in terms of energy consumption control, task completion rate and path efficiency, showing good adaptability and scalability. The algorithm can effectively reduce path conflicts through intelligent clustering and region division, and enhance the decision-making ability of UAV in unknown environments by means of local perception. However, its disadvantages are high requirements for computing resources, insufficient robustness in extreme dynamic communication environments, and large training overhead in ultra-large-scale scenarios.

3.2. Multi-UAV Path Planning Algorithm based on Deep Reinforcement learning

Bayerlein et al. proposed a deep reinforcement learning-based path planning algorithm for multi-UAV wireless data collection [10], which provides an innovative solution for multi-UAV collaborative data collection tasks in dynamic and complex environments. The algorithm adopts a centralized global-local map processing mechanism, which dynamically rearranges the environment map centered on the current position of the UAV, and constructs a compressed global map and a high-resolution local map to jointly input the convolutional neural network to achieve the generalization ability of scene parameters. In other words, the learned control policy can directly adapt to the changes of multiple parameters such as the number of Uavs, equipment distribution, and flight time without re-training. In the experiment of dense urban environment, three Uavs cooperate to achieve 100% data collection rate, and all Uavs can land safely on time. Compared with the traditional scalar input method, the map input method significantly improves the training efficiency and policy performance. The study proves that the centralized map representation can significantly improve the agent's ability to understand the spatial relationship, the global-local two-stream architecture effectively balance the requirements of long-distance planning and close-range obstacle avoidance, and the parameter generalization mechanism greatly reduces the cost of repeated training of the scene in the actual deployment, which significantly improves the adaptability and practicability of the multi-UAV system in dynamic tasks. However, the algorithm still has some limitations: its action space is limited to two-dimensional discrete grid movement, without considering continuous control and three-dimensional height adjustment, and its training process relies on a large number of simulation data, which leads to high computational cost. At the same time, the performance will be slightly decreased when the scene parameters are extremely beyond the training distribution, reflecting that there are still boundaries in its generalization ability.

Westheider et al. proposed a multi-UAV adaptive path planning method based on deep reinforcement learning and counterfactual credit assignment mechanism [11]. By introducing COMA framework and centralized map representation explicitly, the credit assignment problem in multi-agent cooperation is solved, so that UAV groups can efficiently monitor unknown terrain in three-dimensional space. Experiments show that the proposed method is superior to traditional non-learning methods on synthetic and real thermal data, which can reduce map uncertainty and improve monitoring accuracy faster, and can adapt to different team sizes and communication constraints without re-training, which proves that the proposed method has strong generalization ability and adaptability in complex collaborative tasks. However, it also exposes its dependence on computing

resources and simulation training, and its scalability in larger scale or dynamic unknown environments still needs to be further verified.

4. Challenges and Development

At present, UAV path planning and coordination technology based on deep learning and reinforcement learning still faces multiple challenges. First of all, the generalization ability of the algorithm is insufficient. Most models are trained in simulation environments or specific scenarios, and their performance is easy to deteriorate when facing unknown factors such as complex electromagnetic interference, random wind field, and dynamic obstacles in real scenes. Secondly, there is an inherent contradiction between real-time performance and computational complexity, and deep learning algorithms require a large amount of computing power to support. In high-density obstacles or large-scale cluster collaboration scenarios, the reasoning delay is easy to exceed the obstacle avoidance response threshold of UAV, and the computing power limit of airborne processor further exacerbates this contradiction. At the same time, the data set construction is imperfect, and there is a lack of unified benchmark data set. Most of the existing studies use self-built data, and the inconsistency between labeling standards and data distribution makes it difficult to compare the performance of the algorithm horizontally. Moreover, it is difficult to balance multi-objective optimization, and there are mutual constraints among the objectives such as the shortest path, the lowest energy consumption, and safe obstacle avoidance. The design of reward function is difficult to take into account comprehensively, and local optimal solutions are easy to appear.

In the future, UAV path planning technology will develop in the direction of lightweight, cross-domain adaptive, and distributed cooperation. At the algorithm level, model pruning, quantization and meta-learning technologies should be combined to reduce the computational overhead and improve the generalization ability of the model in complex scenarios such as extreme weather and strong electromagnetic interference. At the perception level, the deep combination of multi-sensor fusion (vision, laser, radar) and feature extraction networks such as CNN should be used to strengthen the real-time perception and rapid decision-making ability of UAV in dynamic environments. In the aspect of multi-level collaboration, the decentralized architecture is used to reduce communication dependence and realize autonomous task allocation and path coordination between Uavs. The adaptive reward function and multi-objective reinforcement learning algorithm are used to achieve the dynamic balance of safety, efficiency, and energy consumption, and promote the large-scale landing of UAV technology in disaster rescue, urban inspection, low-altitude logistics and other scenarios.

5. Conclusion

This paper systematically reviews the UAV path planning and technical system based on deep learning and reinforcement learning. The relevant algorithms show obvious advantages in path safety, real-time response ability, energy efficiency and collaborative task completion rate, which effectively make up for the lack of adaptability of traditional methods in dynamic and complex environments. The improved Q-Learning and DDPG frameworks show good real-time performance and convergence efficiency in known or partially known structured environments, but their modeling dependence and generalization ability are still insufficient when facing highly dynamic and unstructured real scenes. The enhanced algorithms represented by DySAC and SMLTO improve the adaptability and security in dynamic obstacles and unknown areas, but also limit their deployment in resource-constrained platforms or extreme environments due to model complexity and sensor dependence. For multi-UAV collaboration, algorithms such as MRF-DQN and DPPDQN have significantly improved the efficiency and scalability of task collaboration through experience replay, region segmentation and feature extraction. However, their centralized or high-computational architectures still face the challenge of scalability and real-time performance when facing large-scale

clusters or strong communication constraints. In general, the current algorithms have made significant progress in simulation and specific scenarios, but there are still further breakthroughs to be made in cross-scenario generalization, lightweight deployment, and robust cooperation in dynamic unknown environments.

In the future, with the integration of deep learning, reinforcement learning and transfer learning, edge computing, and the development of multimodal sensor data fusion technology, the simultaneous realization of lightweight and high generalization of algorithms will become a reality, which will promote the application of Uavs in emergency rescue, smart city inspection, precision agriculture and other fields.

References

[1] Han J C. Multi-UAV collaborative path planning based on reinforcement learning. Xi'an University of Technology, 2025.

[2] Gao Y. Research on Path Planning Algorithm for Multi-UAV Data Collection Based on Reinforcement Learning. Qilu Industrial University, 2025.

[3] Zamoum Y, Baiche K, Benkeddad Y, et al. Modern artificial intelligence technics for unmanned aerial vehicles path planning and control. Bulletin of Electrical Engineering and Informatics, 2025, 14(1): 153-172.

[4] Yin Y, Wang X, Zhou J. Q-Learning-based Multi-UAV Cooperative Path Planning Method. Binggong Xuebao/Acta Armamentarii, 2023, 44(2): 484-495.

[5] Fan H Y. Optimization Method of Autonomous Obstacle Avoidance Path Planning for UAV Based on Deep Reinforcement Learning. New Technology and New Products in China, 2025, (15): 8-10.

[6] Cao X W. Research on UAV Path Planning Method Based on Deep Reinforcement Learning. Sichuan University, 2024.

[7] Tang J. Research on trajectory planning and target tracking strategy for Unmanned Aerial Vehicle in multiple obstacle environment. Southwest Jiaotong University, 2023.

[8] Zhang W J. Research on UAV obstacle detection and path planning algorithm. Xi'an Petroleum University, 2023.

[9] Elfatih N M, Ali E S, Saeed R A. Navigation and Trajectory Planning Techniques for Unmanned Aerial Vehicles Swarm//Artificial Intelligence for Robotics and Autonomous Systems Applications. Cham: Springer International Publishing, 2023: 369-404.

[10] Bayerlein H, Theile M, Caccamo M, et al. multi-UAV path planning for wireless data harvesting with deep reinforcement learning. IEEE Open Journal of the Communications Society, 2021, 2: 1171-1187.

[11] Westheider J, Ruckin J, Popović M. Multi-uav adaptive path planning using deep reinforcement learning //2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023.