Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 14, No. 1, 2025

Application of Reinforcement Learning to UAV Decision Making and Path Planning

Chuqiao Cheng *

Queen Mary School Hainan, Beijing University of Posts and Telecommunications, Lingshui Li Autonomous County, Hainan Province, China

*** Corresponding author Email:** chengchuqiao@bupt.edu.cn

Abstract: The aim of this study is to systematically explore the framework and potential of reinforcement learning for UAV decision-making and path planning. The article firstly analyses the core advantages of reinforcement learning over traditional planning methods. its ability to cope with environmental uncertainty, achieve multi-objective optimisation, and complete end-to-end decision-making. Then, by focusing on three typical application scenarios, namely autonomous navigation and obstacle avoidance, multi-UAV collaboration, and resource-constrained planning, the paper analyses the technical path and effectiveness of reinforcement learning in solving key problems in specific domains. Despite the challenges of sample efficiency, security, and interpretability, cutting-edge developments such as simulation-to-reality migration and constrained reinforcement learning are continuing to drive the technology towards practicality. The systematic discussion in this study aims to provide theoretical references for cross-disciplinary research on reinforcement learning and robotics, and to lay an important technical foundation for building the next-generation intelligent, robust and adaptive unmanned aerial systems.

Keywords: Reinforcement Learning; Drones; Path Planning.

1. Introduction

The rapid development of the UAV industry has promoted its large-scale application in logistics and transport, geographic mapping, agricultural monitoring, emergency rescue and security defence. With the continuous expansion of application scenarios, UAVs are facing increasingly complex operating environments, especially in urban canyons, indoor spaces, and unstructured environments with dynamic obstacles, where traditional decision-making and planning methods based on a priori information about the environment face significant challenges. These complex environments put forward higher requirements on the autonomous decision-making capability of the vehicle, and there is an urgent need to develop new intelligent decision-making methods.

In recent years, the research on the application of reinforcement learning in UAV decision-making and path planning has made significant progress, and its application scope has been extended from single-aircraft obstacle avoidance to complex scenarios such as multi-aircraft collaboration. First, in terms of single-aircraft navigation in complex environments, researchers have used deep reinforcement learning (DRL) algorithms to train UAVs to achieve end-to-end autonomous flight and dynamic obstacle avoidance by relying only on on-board visual sensors, which significantly improves the survivability in unknown GPS denial environments [1]. Second, in the field of multi-UAV cooperative control, the multi-intelligence reinforcement learning (MARL) method based on a centralised training-distributed execution framework is successfully applied to solve the cooperative search and target rounding problem of UAV swarms under communication-constrained conditions, demonstrating excellent group intelligence [2]. Finally, in terms of persistent operation under resource constraints, the study enables UAVs to autonomously learn the optimal path that takes into account both task completion and energy efficiency by integrating constraints such as energy consumption and flight time into the reward function, effectively extending their operational flight time in inspection, mapping and other tasks [3]. Together, these research results demonstrate the effectiveness and great potential of reinforcement learning in solving complex real-world problems of UAVs.

This study systematically explores the application of reinforcement learning in UAV decision-making and path planning. Firstly, it analyses the advantages of reinforcement learning over traditional methods in terms of environment adaptability, multi-objective optimization and end-to-end decision-making; secondly, it focuses on the solutions and effectiveness in three major scenarios, namely, autonomous navigation and obstacle avoidance, multi-machine collaborative control, and resource-constrained planning; and lastly, in view of the challenges of sample efficiency and safety, it explores the key directions of development, such as simulation migration and constraint optimization, which provides theoretical references to the research and development of intelligent UAV systems. Theoretical references are provided.

2. Advantages of Reinforcement Learning over Traditional Methods

Traditional methods rely heavily on accurate environment models and face the bottleneck of inefficient replanning and a lack of adaptability in dynamic and uncertain environments. In contrast, RL autonomously learns the optimal policy by interacting with the environment, which has many advantages, as shown in Table 1. Firstly, it has excellent environment generalisation and dynamic adaptation capabilities, and the trained strategies can make real-time and robust obstacle avoidance decisions for dynamic obstacles directly based on local observations, without the need for explicit global maps and frequent replanning. Secondly, RL achieves comprehensive optimality under multi-objective, and its

reward function mechanism can co-optimise multiple constraints, such as path length, energy consumption, task time and safety, so as to plan a path that is not the shortest but the most globally efficient one, which is especially suitable for resource-sensitive tasks such as long flight time. Finally, it provides a unified framework for complex collaborative decision-making. Based on architectures such as centralised training-distributed execution, multi-UAS can emerge efficient group intelligence based on local information alone under communication constraints, solving the distributed collaboration problem that is difficult to handle by traditional methods. As a result, RL elevates UAVs from tools that execute predefined procedures to intelligences that can autonomously perceive, decide and learn in uncertain environments.

Table 1. Comparison between Traditional Methods and Reinforcement Learning for UAV Path Planning

Characteristic	Traditional Methods	Reinforcement Learning Approach
Environmental Model	Requires a precise global map/model.	Does not require an environmental model; learns through interaction based on partial perception.
Dynamic Handling	Poor handling of dynamic obstacles; requires frequent re-planning.	Inherently suitable for dynamic environments, the learned policy can adapt to changes.
Uncertainty Robustness	Sensitive to sensor noise and model inaccuracies.	Can learn robust policies against uncertainties through extensive training.
Decision Dimension	Typically limited to low-level path planning.	Enables hierarchical decision-making, integrating high-level and low-level control.
Optimality	Finds analytically or asymptotically optimal paths in static environments.	Seeks the solution with maximum expected return in uncertain environments, often yielding the most robust and safe, rather than the shortest, path.

3. Application Scenarios of Reinforcement Learning in UAV Decision Planning

3.1. Autonomous Navigation and Obstacle Avoidance

Compared with the traditional path planning algorithms, reinforcement learning gives a higher level of intelligence to autonomous navigation and obstacle avoidance of UAVs. Traditional methods usually decompose navigation into multiple independent modules, such as 'perception-map construction-planning-control', and the error of each module will be transmitted step by step. Reinforcement learning enables 'end-to-end' or 'sense-action' mapping. Superior adaptability to dynamic and unknown environments is one of the most significant advantages of RL.

In an indoor/forest scenario with GPS denial, Zhang Yuntao uses the end-to-end framework of deep reinforcement learning (DRL) to directly map visual images into continuous obstacle avoidance actions, which solves the problem of the traditional Visual-SLAM method that requires pre-built maps and is difficult to adapt to unknown environments; the method trains the intelligences with DDPG, which eliminates the need for manual feature engineering, and has the advantages of 'zero map initiation and real-time replanning'. The method uses DDPG to train the intelligent body, eliminating manual feature engineering, with the benefits of 'zero map startup, real-time replanning', and improving the task completion rate by 19.98% [4].

For the GPS-deficient environments such as urban canyons and warehouses, Li Yang et al. proposed an end-to-end obstacle avoidance decision-making model based on DDPG, which inputs LiDAR point clouds or depth maps, and outputs continuous speed/attitude commands, solving the problems of discontinuous control and restricted direction of 3D-VFH; experiments show that DDPG trajectories are smoother and the collision rate is reduced by more than 35% [5].

3.2. Multi-UAV Collaboration

Reinforcement learning empowers multi-UAV collaborative systems with distributed intelligence beyond traditional centralised or regularised approaches. Its core advantage lies in the fact that, through a framework such as centralised training-distributed execution, each UAV can make decisions relying only on local observations, thus achieving efficient collaboration in complex environments such as communication constraints. As shown in the collaborative search task, the method can significantly improve area coverage and reduce repeated search and energy consumption; while in collaborative localisation and formation maintenance, RL can actively dispatch UAVs to estimate the status of out-of-place members through collaborative observation, so as to quickly recover the formation without external assistance.

Tailin Zhou uses the Multi-intelligent Proximal Policy Optimisation (MAPPO) algorithm to learn a 'who-goes-where' distributed search strategy for swarm UAVs, which solves the problem of high repeat coverage and low target discovery rate of traditional parallel line/sector search in dynamic obstacle and communication-constrained environments. UAVs to maintain synergy by relying only on local observations, with the benefits of increasing area coverage from 65% to 85% for independent flights and reducing the average flight distance by 20%, which significantly saves energy and shortens the golden time for search and rescue [6].

Wan et al. used a deep Q-network (DQN) to establish a 'cooperative positioning scheduling' POMDP model for GPS denial environments, which solves the problem of maintaining formation after some UAVs in swarm formation lose their positioning; updating the belief state by EKF and outputting the optimal observation-manoeuvre strategy by

DQN, which has the benefit of enabling three lost UAVs to re-locate within 6 s [7]. positioning UAVs converge back to a triangular formation with error <0.5m within 6s, the whole process does not require external base stations or pre-built maps, which significantly improves the robustness of the swarm under interference confrontation [7].

3.3. Energy Optimised Path Planning

Reinforcement learning is able to autonomously learn flight strategies that take into account aerodynamics, wind field environment and its own dynamics by integrating energy consumption terms in its reward function. This data-driven approach allows the UAV to not only complete its mission, but to perform it in the most globally energy-efficient manner. By interacting with the environment, the intelligence learns to make fine trade-offs between flight speed, acceleration, climb angle and energy consumption, resulting in intelligent behaviours such as using inertia to glide, choosing a leeward path or maintaining the most economical cruise speed. This is of key significance for improving the endurance and mission reliability of long endurance UAVs in mapping and inspection missions.

Zhou et al. use a Dual Deep Q Network + Prioritised Experience Replay (DDQN-PER) to write real-time power consumption, link stability and data collection gain into a reward function, and learn a joint energy-saving-communication strategy for multiple UAVs performing mobile edge computing, which solves the problem of 'speed-height-hover' in a discrete action space. The PER mechanism ensures the reuse of important experience, improves the convergence speed by 30%, reduces the total energy consumption by 20.2% compared with the shortest path method in the simulation of a 200-node IoT region, and improves the timeliness of data collection by 14%, which verifies the potential of reinforcement learning for path optimisation in energy-constrained scenarios [8].

Baierlein et al. proposed the Multi-UAV Deep Deterministic Policy Gradient (MADDPG) data harvesting framework, which takes the 'number of harvested bits - propulsion energy - collision penalty' as a whole as a reward, and allows the swarm to learn 'when to hover - when to collide' in the wind field. "The centralised training-distributed execution architecture enables zero-shot migration of strategies to different wind speeds, and the outdoor 3-aircraft experiments show that the total number of harvested bits is increased by 37% and the flight time is extended by 18% under the same power level, which significantly improves the performance of long endurance UAVs in emergency communication and disaster area inspection missions. communication and disaster area inspection tasks [9].

4. Challenges and Development

4.1. Challenges

Although reinforcement learning empowers UAVs with powerful environmental adaptation and decision-making capabilities, it still faces a number of challenges. The first challenge is the high sample complexity of the algorithms. Intelligent systems usually need to undergo a large number of interaction trials before converging to an effective strategy, which leads to high time and economic costs and unacceptable operational risks associated with direct training on physical UAV platforms. Second, the safety of the training process cannot be ignored. The inherent randomness of the exploration phase is prone to induce catastrophic failures (e.g., crashes), which severely constrain online learning in the physical world. Finally, the lack of interpretability and verifiability of policy models constitutes a central obstacle to their deployment in safety-critical scenarios. The black-box nature of deep neural networks makes it difficult to trace and diagnose their decision logic, which in turn makes it difficult to adequately ensure the robustness of the policies under extreme or adversarial conditions, and creates fundamental difficulties in the assessment and certification of system reliability.

4.2. Future Developments

To address the above challenges, research in this area is focusing on improving the practicality, security and reliability of algorithms, and several clear technology trends have been formed. To address the bottleneck of sample efficiency and security, simulation-to-reality transfer learning has become a mainstream research paradigm. This approach effectively bridges the 'simulation-reality' gap by pre-training in high-quality simulation environments and adopting domain randomisation techniques to enhance the generalisation of the model to the uncertainty of reality. Meanwhile, Constrained Reinforcement Learning (CREL) and Imitation Learning (IL) provide critical safety guarantees for the training process: CREL translates safety requirements into hard constraints in the optimisation problem, while IL derives a priori strategies from expert data such as traditional planners, both of which significantly reduce the physical risks associated with direct exploration. In terms of reward design, Reverse Reinforcement Learning (RRL) demonstrates the value of reducing the reliance on human design experience by deriving reward functions from expert tutorials. Finally, to address the issue of model trustworthiness, the fusion of interpretable AI for decision traceability and formal validation methods to rigorously prove the robustness of strategies is becoming an important way to build reliable UAV intelligences. The synergistic development of these techniques together drives the transformation of reinforcement learning from theoretical algorithms to practical engineering solutions.

5. Conclusion

This study systematically explores the application value and practical path of reinforcement learning in UAV decision-making and path planning. Firstly, the unique advantages of reinforcement learning in terms of environment adaptability, multi-objective optimisation and end-to-end decision-making are explained by comparing with traditional planning methods. Secondly, it focuses on analysing its solutions and effectiveness in three typical scenarios: autonomous navigation and obstacle avoidance, multi-machine cooperative control and resource constraint planning. Finally, for the existing challenges such as sample efficiency and safety, key development directions such as simulation migration and constraint optimisation are explored, providing theoretical references and technical paths for building a new generation of intelligent UAV systems.

References

[1] Kong X, Zhang R, Chen L, et al. Energy-aware path planning for long-endurance UAVs using constrained deep reinforcement learning. IEEE Transactions on Vehicular Technology, 2023, 72(5): 5891-5904.

[2] Loquercio A, Kaufmann E, Ranftl R, et al. Learning high-speed flight in the wild. Science Robotics, 2020, 6(59): eabg5810.

[3] Zhou T, Wang L, Wang S. Distributed multi-UAV cooperative search and tracking based on deep reinforcement learning. Aerospace Science and Technology, 2022, 131: 107989.

[4] Zhang Yuntai. Research on UAV obstacle avoidance path planning based on deep reinforcement learning. Master's Thesis, Southeast University, 2022.

[5] Li Yang, Wang Chen, Zhang Wei. Research on 3D obstacle avoidance algorithm for UAV based on DDPG. Journal of Northwestern Polytechnical University, 2022, 40(5): 901-908.

[6] Zhou Tailin. Research on cooperative search algorithm for UAV swarm based on multi-agent reinforcement learning. Master's Thesis, Zhejiang Normal University, 2025.

[7] Wan Kaifang, Li Xiang, Wang Zili. Cooperative localization and formation control of UAVs based on deep reinforcement learning under GPS-denied environments. Acta Aeronautica et Astronautica Sinica, 2025, 46(2): 1-10.

[8] Zhou J, Liu M L, Zhao Y F, et al. Energy-efficient UAV path planning based on double deep Q-network for mobile edge computing. Computer Applications Research, 2021, 38(6): 1699-1703.

[9] Baierlein N, Ayan Ö, Göktepe E, et al. Multi-agent deep reinforcement learning for energy-efficient data harvesting with UAV swarms. IEEE Open Journal of the Communications Society, 2021, 2: 1171-1187.