Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.BA33285

A Review of Research on Cooperative Path Planning for UAV Swarms Based on Ant Colony Optimization and Deep Reinforcement Learning

Wenjue Yan

School of Mechanical Engineering, Tianjin University of Technology, Tianjin, China
ywj20041105@stud.tjut.edu.cn

Abstract. With the continuous expansion of the application of UAV swarms in tasks such as post-disaster search and rescue, environmental inspection, agriculture and forestry monitoring, and low-altitude security, the path planning problem has evolved from shortest-path search for a single UAV in static and known environments into a cooperative decision-making problem for multiple UAVs in dynamic and unknown environments. Ant colony optimization has the advantages of distributed search, strong global optimization capability, and easy integration into path cost functions, but it suffers from slow convergence, proneness to local optima, and insufficient adaptability to continuous spaces in complex environments. Deep reinforcement learning can achieve online decision-making and adaptive obstacle avoidance through interaction with the environment, making it more suitable for handling dynamic obstacles, partial observations, and multi-UAV cooperative tasks; however, it also faces challenges such as low sample efficiency, complex reward design, and limited generalization ability. Focusing on cooperative path planning for UAV swarms, this paper systematically reviews the research progress of ant colony optimization, deep reinforcement learning, and their hybrid methods, with emphasis on comparing the differences between the two types of methods in terms of environment modeling, coordination mechanisms, path quality, real-time response, and engineering deployability. On this basis, it further summarizes the key bottlenecks in current research, including unified modeling of complex constraints, global-local coordination interfaces, communication and information sharing, and simulation-to-reality transfer. The review concludes that constructing a hierarchical hybrid framework in which ant colony optimization is responsible for global candidate path generation and deep reinforcement learning is responsible for local online correction is an important development direction for improving the performance and engineering applicability of UAV swarm path planning.

Keywords: UAV swarm, path planning, ant colony optimization, deep reinforcement learning, multi-agent coordination

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.BA33285

1. Introduction

Unmanned Aerial Vehicle (UAV) swarms have advantages such as flexible deployment, wide coverage, strong fault tolerance, and high task parallelism, and have become an important research direction in multi-agent autonomous systems. As mission scenarios expand from regular, static environments to urban canyons, woodland areas, fire zones, and unknown regions, path planning is no longer merely a shortest-path search in the geometric sense, but must simultaneously satisfy multiple constraints such as obstacle avoidance, collision avoidance, energy consumption, cooperation, communication, and real-time response. Existing studies show that multi-UAV path planning usually requires joint decision-making in partially observable environments, ensuring not only the flight safety of individual UAVs but also the task efficiency and cooperative benefits at the swarm level [1-4]. Meanwhile, the hyper-heuristic idea combining reinforcement learning with heuristic optimization has been proven suitable for complex optimization problems, which also provides a theoretical basis for the integration of "search + learning" in UAV path planning [1]. Therefore, a systematic review of Ant Colony Optimization (ACO), Deep Reinforcement Learning (DRL), and their integrated approaches has clear research value and engineering significance.

2. Research paradigms for UAV swarm path planning

From the perspective of research development, current UAV swarm path planning has generally formed three technical routes: the first is swarm intelligence optimization methods represented by ant colony optimization, the second is data-driven decision-making methods represented by deep reinforcement learning, and the third is hybrid methods that integrate heuristic search with learning mechanisms. The first is good at global search in large-scale solution spaces and is suitable for static scenarios or scenarios with relatively complete prior maps; the second relies on continuous interaction with the environment to learn policies and is more suitable for unknown environments, dynamic obstacles, and multi-UAV cooperative exploration tasks; the integrated methods that have emerged in recent years emphasize using learning mechanisms to enhance the adaptability of heuristic search, or using heuristic methods to reduce the training difficulty and search blindness of reinforcement learning [1]. Puente-Castro et al. studied reinforcement learning methods for UAV swarms oriented to area coverage tasks, and pointed out that in multi-UAV cooperation, using a unified control network can achieve shorter flight time in certain scenarios, thus reflecting the coupling relationship between parameter sharing and swarm cooperation [4].

Table 1. Overall comparison of ACO, DRL, and hybrid methods in UAV swarm path planning

Method Paradigm	Representative References	Main Advantages	Limitations and Applicable Scenarios
Improved ACO	[5] [6] [7]	Global search; interpretable;easy cost modeling	Sensitive to parameters; slow convergence; static environments
Multi-agent DRL	[2] [3] [8]	Online decision-making; adaptive; cooperative control	High training cost; reward design complexity; limited generalization
Hybrid Methods	[1] [7] [9]	Combine global + local; improved adaptability	Interface design unclear; stability needs improvement

DOI: 10.54254/2755-2721/2026.BA33285

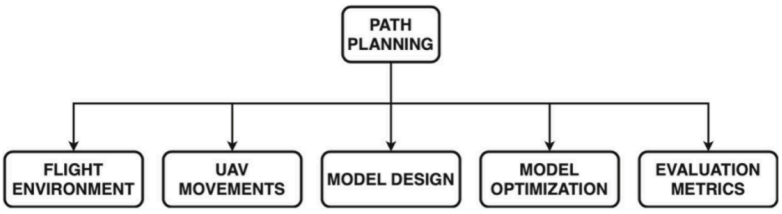


Figure 1. Diagram with the formulation of path planning problems. It summarizes all the inherent and necessary problems to guarantee the validity of the final system

Note. From "UAV swarm path planning with reinforcement learning for field prospecting," by A. Puente-Castro, D. Rivero, A. Pazos, and E. Fernandez-Blanco, 2022, Applied Intelligence, 52, pp. 14101–14118 (https://doi.org/10.1007/s10489-022-03254-4).

3. Research progress of ant colony optimization in UAV path planning

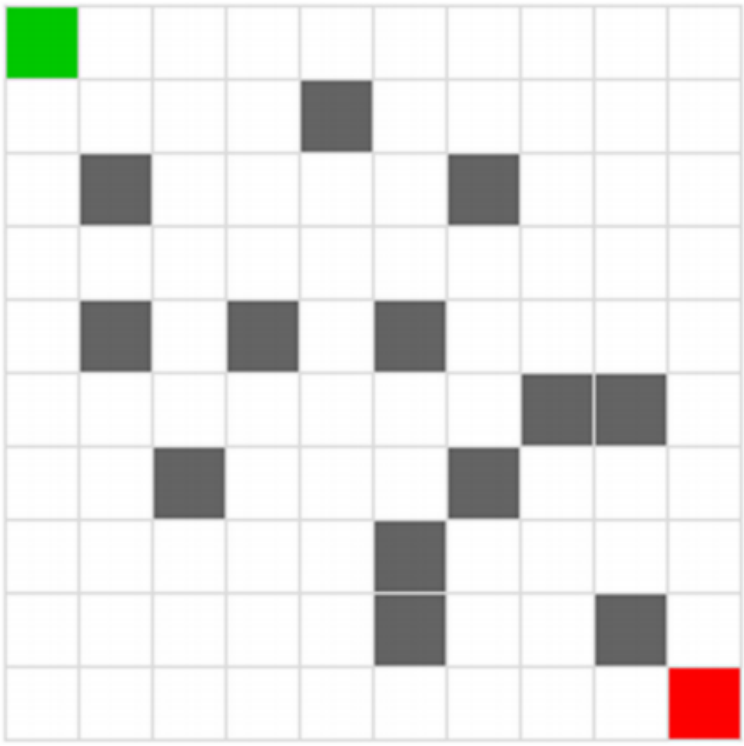


Figure 2. Grid method to establish environment model. The figure illustrates that the unmanned workspace is decomposed into a series of binary information grid cells

Note. From "Unmanned vehicle path planning using a novel ant colony algorithm," by X. Yue and Y. Chen, 2019, EURASIP Journal on Wireless Communications and Networking, 2019, Article 136 (https://doi.org/10.1186/s13638-019-1474-5).

However, traditional ACO also faces significant limitations in UAV scenarios. First, the randomness of the initial search is strong, which easily leads to blind traversal. Second, the positive feedback mechanism may cause premature convergence to a local optimum. Third, the native ACO is mainly applicable to discrete spaces, and in continuous three-dimensional space it often requires probability density modeling or additional repair strategies. To address the above problems, researchers have mainly improved the algorithm from three directions: the heuristic function, the state transition rule, and the pheromone update mechanism. Li Yan et al. introduced Q-learning into ant colony optimization, using reinforcement learning to explore the environment in advance and

DOI: 10.54254/2755-2721/2026.BA33285

converting the Q-table into a non-uniform initial pheromone distribution. At the same time, they introduced an angle heuristic function, an optimized distance heuristic function, and a reward-punishment pheromone update mechanism into the state transition probability, thereby improving convergence speed, deadlock problems, and global search capability [6]. Furthermore, Wang et al. proposed QMSR-ACOR, which uses Q-learning to dynamically select construction strategies and walking strategies in continuous three-dimensional multi-UAV scenarios, and incorporates an elite waypoint repair mechanism, significantly improving path feasibility and robustness in complex obstacle scenarios [7].

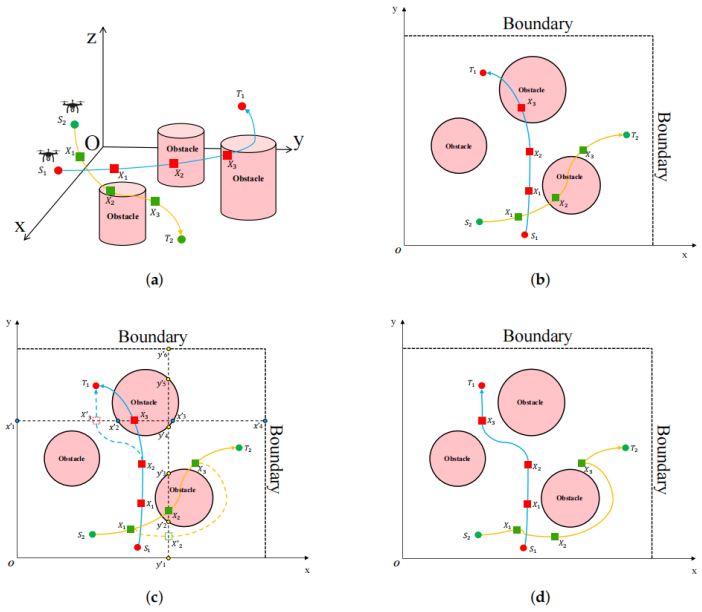


Figure 3. Elite waypoint repair strategy. (a) 3D view. (b) Initial paths. (c) Repair process. (d) Repaired paths

Note. From "Path Planning for Multi-UAV in a Complex Environment Based on Reinforcement-Learning-Driven Continuous Ant Colony Optimization," by Y. Wang, J. Liu, Y. Qian, and W. Yi, 2025, Drones, 9, Article 638 (https://doi.org/10.3390/drones9090638).

4. Research progress of deep reinforcement learning in cooperative path planning for UAV swarms

Unlike heuristic search, deep reinforcement learning learns the optimal policy through "state-action-reward" interaction, and is more suitable for handling online decision-making problems in dynamic environments. In multi-UAV path planning, the core advantage of DRL is that it can use local observations to achieve adaptive obstacle avoidance, and coordinate multi-agent behavior through mechanisms such as centralized training and decentralized execution. Si Pengbo et al. modeled multi-UAV path planning as a partially observable Markov process, proposed a path planning framework based on MAPPO, and combined network pruning to form the NP-MAPPO algorithm, enabling the model to improve training efficiency and deployment efficiency while considering obstacle constraints, inter-UAV collision constraints, and energy consumption constraints [2].

DOI: 10.54254/2755-2721/2026.BA33285

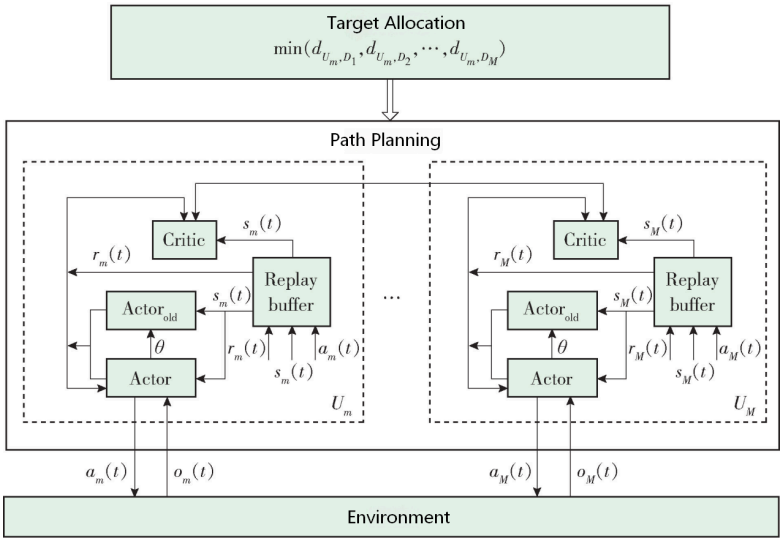


Figure 4. Multi-UAV path planning network framework

Note. From "UAV path planning based on multi-agent deep reinforcement learning" by P. Si, B. Wu, R. Yang, M. Li, and Y. Sun, 2023, Journal of Beijing University of Technology,49(4),pp.449-458 (https://doi.org/10.11936/bjutxb2022080007).

Such studies indicate that the focus of multi-agent DRL is not only on path generation itself, but also on the design of state representation, reward shaping, and cooperative training mechanisms.

5. Research trends in the integration of ACO and DRL

From existing studies, ACO and DRL are not in a substitutive relationship, but are more suitable for forming a hybrid framework with complementary advantages. Li et al. pointed out in their review of RL-HH that the core value of combining reinforcement learning with heuristic methods lies in using learning capability to dynamically adjust search strategies while retaining the efficient exploration ability of heuristic methods in complex solution spaces [1]. This view is also applicable to UAV path planning: ACO is suitable for undertaking tasks such as global candidate path generation, key waypoint construction, and multi-objective cost search, while DRL is more adept at handling local dynamic obstacle avoidance, real-time cooperation, and online correction tasks.

Existing representative works have already reflected this integration idea. On the one hand, Li Yan et al. and Wang et al., starting respectively from discrete grids and continuous three-dimensional space, introduced Q-learning to enhance the strategy selection and search quality of ACO, proving the effectiveness of "learning-assisted search" [6,7]. On the other hand, Mondal et al. proposed the OptiRoute framework, in which reinforcement learning is used to select energy-replenishment rendezvous points in heterogeneous UAV-UGV cooperative routing, and heuristic constraint planning is then used to generate cooperative routes. The results showed that this kind of heuristics-assisted RL framework outperformed the genetic algorithm baseline in metrics such as task completion time, idle time, and energy consumption [9]. Although this study focuses on heterogeneous systems rather than pure UAV swarms, it provides strong inspiration for the architectural division in which "reinforcement learning is responsible for strategy selection, while heuristic methods are responsible for path solving."

DOI: 10.54254/2755-2721/2026.BA33285

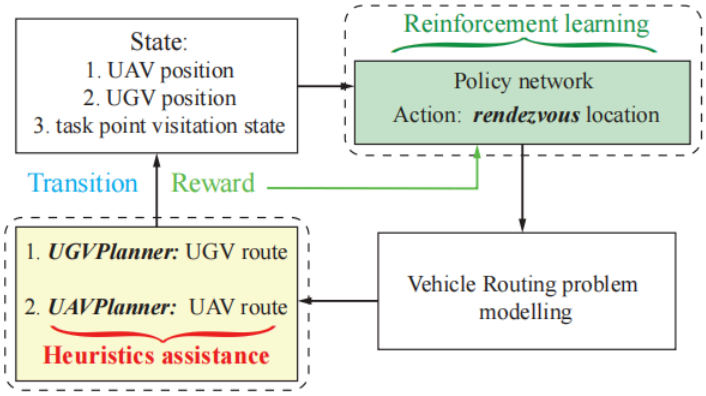


Figure 5. MDP for heuristics-assisted UAV-UGV cooperative routing

Note. From "OptiRoute: A heuristic-assisted deep reinforcement learning framework for UAV-UGV collaborative route planning," by M. S. Mondal, S. Ramasamy, and P. A. Bhounsule, 2023, arXiv (https://doi.org/10.48550/arXiv.2309.09942).

Based on the existing literature, it can be foreseen that a more promising direction is to construct a two-layer planning system of "global guidance + local correction": the upper layer uses improved ACO to generate the global path skeleton, task allocation scheme, or key waypoints, while the lower layer uses multi-agent DRL to achieve dynamic obstacle avoidance, formation adjustment, and cooperative behavior correction based on local observations and shared information. This hierarchical framework can not only reduce the training burden of DRL caused by searching from scratch, but also compensate for the delayed response of ACO when facing dynamic environments, and thus represents an important development trend in UAV swarm path planning under complex scenarios.

6. Main challenges and development directions

Based on the existing studies, future cooperative path planning for UAV swarms still needs to focus on solving the following problems. First, the capability for unified modeling of complex constraints is insufficient. Real-world missions often simultaneously involve multiple constraints such as obstacle avoidance, trajectory smoothing, time synchronization, communication maintenance, and energy consumption control, whereas most existing studies focus on only one of these aspects, and there is still a lack of a unified task-path-control integrated framework. Second, sim-to-real transfer remains weak. Most existing methods are mainly validated in Unity, Matlab, or self-built simulation platforms, and their performance may degrade significantly when facing real sensor errors, wind-field disturbances, and communication delays [2,3]. Third, multi-agent cooperation mechanisms still need to be further developed. Although many current studies use multi-agent models, cooperation often remains at the level of shared states or shared rewards, and issues such as information redundancy, communication topology, and decision consistency have not yet been fully addressed. Fourth, the interface design of hybrid algorithms is still immature. How to express global path cost and local rewards in a unified way, when to trigger local replanning, and how to prevent conflicts between global objectives and local behaviors are still key issues that must be resolved before ACO-DRL hybrid frameworks can move toward engineering applications.

Future research may proceed in three directions. First, graph neural networks, attention mechanisms, and world models can be combined to improve multi-agent state representation and generalization capability. Second, for dynamic and partially known environments, hierarchical

DOI: 10.54254/2755-2721/2026.BA33285

planning frameworks that can be updated online should be developed. Third, while maintaining algorithm performance, model compression, transfer learning, and curriculum learning can be introduced to improve the deployability of algorithms on embedded platforms.

7. Conclusion

Cooperative path planning for UAV swarms is evolving from a traditional static path search problem into a comprehensive optimization problem involving global optimization, local online decision-making, and multi-agent cooperative control. Ant colony optimization has advantages in global search, path interpretability, and multi-objective cost modeling, but it has limitations in continuous spaces, dynamic environments, and under complex constraints. Deep reinforcement learning can better adapt to unknown environments and dynamic obstacles, but it also faces problems such as high training cost, insufficient stability, and difficulty in transfer. Existing studies show that both improving ACO based on learning mechanisms and assisting DRL with heuristic methods have achieved preliminary results. Overall, conducting research around a hierarchical hybrid framework in which "ACO is responsible for global guidance, while DRL is responsible for local cooperation" will be a key direction for improving the quality, real-time performance, and engineering applicability of UAV swarm path planning.

References

[1] Li C, Wei X, Wang J, Wang S, Zhang S. A review of reinforcement learning based hyper-heuristics [J]. PeerJ Computer Science, 2024, 10: e2141. DOI: 10.7717/peerj-cs.2141.
[2] SI, P., WU, B., YANG, R., LI, M., & SUN, Y. (2023). UAV path planning based on multi-agent deep reinforcement learning. Journal of Beijing University of Technology, 49(4), 449–458. https: //doi.org/10.11936/bjutxb2022080007
[3] WANG, W., YOU, M., SUN, L., ZHANG, X., & ZONG, Q. (2024). Intelligent cooperative exploration path planning for UAV swarm in an unknown environment. Chinese Journal of Engineering, 46(7), 1197–1206. https: //doi.org/10.13374/j.issn2095-9389.2023.10.15.002
[4] Puente-Castro A, Rivero D, Pazos A, Fernandez-Blanco E. UAV swarm path planning with reinforcement learning for field prospecting [J]. Applied Intelligence, 2022, 52: 14101-14118. DOI: 10.1007/s10489-022-03254-4.
[5] Yue X, Chen Y. Unmanned vehicle path planning using a novel ant colony algorithm [J]. EURASIP Journal on Wireless Communications and Networking, 2019, 2019: 136. DOI: 10.1186/s13638-019-1474-5.
[6] LI Yan, LIAO Zhanghao, LI Minghui. An improved ant colony optimization algorithm based on reinforcement learning for mobile robot path planning [J/OL]. Computer Engineering and Applications, 2025-12-05. DOI: 10.3778/j.issn.1002-8331.2509-0189.
[7] Wang Y, Liu J, Qian Y, Yi W. Path Planning for Multi-UAV in a Complex Environment Based on Reinforcement-Learning-Driven Continuous Ant Colony Optimization [J]. Drones, 2025, 9: 638. DOI: 10.3390/drones9090638.
[8] Demir K, Tumen V, Kosunalp S, Iliev T. A Deep Reinforcement Learning Algorithm for Trajectory Planning of Swarm UAV Fulfilling Wildfire Reconnaissance [J]. Electronics, 2024, 13: 2568. DOI: 10.3390/electronics13132568.
[9] Mondal M S, Ramasamy S, Bhounsule P A. OptiRoute: A heuristic-assisted deep reinforcement learning framework for UAV-UGV collaborative route planning [EB/OL]. arXiv: 2309.09942, 2023. DOI: 10.48550/arXiv.2309.09942.