https://doi.org/10.31449/inf.v50i12.13609 Informatica **50** (2026) 401–410 **401**

Adaptive UAV Inspection and Path Planning for Distribution Networks Using Multi-Agent Deep Reinforcement Learning

Tingting Yang [1*], Xin Wang [2]
[1]BasicTeaching Department, Shandong Business Institute, Yantai, Shandong 264000, China
[2]School of Architectural Engineering, Shandong Business Institute, Yantai, Shandong 264000, China
E-mail: yangtt2025sdbi@outlook.com
[*]Corresponding author

Keywords: Distribution network, UAV inspection, adaptive, path planning

Received: January 22, 2026

Efficient inspection of distribution networks is crucial for ensuring the stable operation of the power system. To address the limitations of existing methods in complex environments, this paper proposes an adaptive UAV inspection path planning strategy based on multi-agent deep reinforcement learning, specifically employing the Multi-Agent Actor-Attention-Critic (MAAC) algorithm. This method constructs a reinforcement learning environment with a designed reward function to enable UAVs to collaboratively learn optimal inspection paths. Simulation experiments conducted on the AirSim platform demonstrate the proposed method's superior performance compared to Enhanced Particle Swarm Optimization (EPSO) and Double Deep Q-Network (DDQN). The MAAC-based model achieved a higher cumulative discounted reward (4.56 vs. 2.21 for EPSO and 4.37 for DDQN) and a reduced average running time (812.85 seconds vs. 833.45 and 923.41 seconds, respectively), validating its advantages in both solution quality and computational efficiency. The results indicate strong adaptability and high computational efficiency, offering significant potential for practical UAV inspection applications. Future work will focus on integrating advanced techniques to further improve robustness and learning efficiency.

Povzetek: Članek predlaga prilagodljivo strategijo načrtovanja poti za pregledovanje z brezpilotnimi letali (UAV), ki temelji na večagentnem globokem ojačevalnem učenju z algoritmom Multi-Agent Actor-Attention-Critic (MAAC). Metoda ustvari okolje za ojačevalno učenje s prilagojeno funkcijo nagrajevanja, ki omogoča UAV sodelovalno učenje optimalnih pregledovalnih poti. Simulacijski poskusi na platformi AirSim kažejo, da je model MAAC boljši od metod EPSO in DDQN.

1 Introduction

The distribution network is a vital component of the power system, and its healthy and stable operation directly affects reliable power supply and the normal functioning of society [1]. As distribution networks grow in size and complexity, traditional manual inspection methods struggle to meet the need for efficient and accurate inspections. The rise of UAV technology [2] provides a new solution for the intelligent inspection of distribution networks. By equipping UAVs with various sensor devices, efficient inspection of distribution network equipment can be achieved, improving inspection quality and efficiency [3].

As the inspection network gradually evolves towards 5G technology, drones are highly vulnerable to cyberattacks when transmitting massive amounts of inspection data. For instance, research on communication security in the context of 5G vehicular networks indicates that Distributed Denial of Service (DDoS) attacks are a major factor affecting network stability [4]. To ensure the confidentiality and integrity of data transmission, authentication schemes based on fog computing (FCA-VBN, ECA-VFog) [5-6] and lattice-based quantum-resistant authentication scheme (L-CPPA) [7] have been proposed successively to enhance the security of vehicular networks. Furthermore, the application of artificial intelligence (AI) technology in the vehicular ad-hoc network (VANET) environment provides new ideas for improving the secure processing and utilization of traffic data [8]; and the emergency condition authentication mechanism designed using Chebyshev polynomials also offers new insights into drone communication security in complex environments [9]. These studies provide important theoretical foundations for building drone inspection communication networks.

Despite the paramount importance of communication security, traditional path planning algorithms for UAV inspection have certain limitations in practical applications [10]. When facing complex environments and large-scale problems, their computational efficiency is low, often leading to excessively long path planning times, which impacts real-time performance and efficiency.

In recent years, many researchers have made various improvements to traditional methods to address the complex challenges in UAV inspection path planning. Wang et al. proposed a hybrid algorithm based on Nash

bargaining theory and Particle Swarm Optimization, introducing a cooperative game model with an optimized cost function to plan the optimal inspection path for UAVs in complex urban pipe corridors. Phung et al. [11] proposed an enhanced Discrete Particle Swarm Optimization algorithm to solve the Traveling Salesman Problem, improving algorithm performance through deterministic initialization, random mutation, and edge exchange techniques. Luis et al. developed a real-time path planning algorithm for UAV contact inspection tasks in indoor environments [12], which processes point cloud data to significantly reduce algorithm execution time. Guerrero et al. proposed a method combining grid division techniques with the Zermelo-TSP approach [13], automatically obtaining inspection coordinates of points of interest through a grid division algorithm and using the Zermelo-TSP method to calculate the time-optimal path for inspecting all points of interest in the shortest time. Li et al. proposed an improved Bidirectional Ant Colony and Discrete Honey Badger Algorithm specifically for solving multi-UAV path planning problems under multi-wind-field conditions [14]. Wang et al. [15] proposed an Odd-Even Layer Genetic Algorithm to effectively solve UAV path planning problems.

With the rise of deep learning, many path planning methods based on deep learning have been proposed. Yan et al. proposed a Deep Reinforcement Learning method for UAV path planning based on global situational information [16], combining a greedy strategy with heuristic search rules for action selection, achieving path planning in dynamic environments. Lattice-Based Batch Authentication for Secure and Scalable VANETs in the context of vehicular ad-hoc networks, batch authentication protocols have emerged as essential for ensuring both efficiency and security in large-scale, dynamic environments. Recent proposals such as CLA-FC5G, D-BlockAuth, and FCA-VBN leverage fog computing and blockchain to decentralize trust and reduce latency. Notably, lattice-based cryptosystems—exemplified by L-CPPA and FC-LSR — provide post-quantum security while supporting conditional privacy preservation and lightweight signature aggregation. These schemes allow multiple messages or entities to be verified simultaneously, drastically cutting down communication overhead and computational delay. ECA-Vfog further optimizes this paradigm for vehicular fog computing, demonstrating how certificateless designs can eliminate key escrow issues while maintaining batch verification capabilities. Although these studies focus on vehicular networks, their core mechanisms—scalable verification, resilience to quantum attacks, and privacy-aware design—offer a valuable blueprint for securing future multi-UAV systems, where numerous drones must authenticate each other and exchange inspection data reliably in real time. Addressing the sparse reward problem in UAV path planning, Lv et al. proposed an information-theoretic exploration algorithm specifically for UAV platforms [17], generating intrinsic rewards based on state entropy and action entropy to compensate for the scarcity of extrinsic rewards. Wang proposed an improved distributed DRL framework [18], decomposing the UAV navigation task into two simpler sub-tasks, each handled by a designed LSTM-based DRL network using only partial interaction data, thus solving the problem of algorithm non-convergence. Yue et al. proposed an end-to-end intelligent UAV mission planning method based on DRL [19], introducing three policy training techniques—domain randomization, policy entropy maximization, and shared lower-level network parameters—to enhance the model's learning performance and generalization capability. Bayerlein et al. proposed a Multi-Agent Reinforcement Learning method, formulating the path planning problem as a decentralized partially observable Markov decision process [20], and solving it by training a Double Deep Q-Network to approximate the optimal UAV control policy. Summary of related works on uav inspection path planning as show in Table 1.

As synthesized in Table 1 (See the appendix), existing UAV path planning methods have made considerable advances in static or structured environments, with notable contributions in optimization-based, heuristic, and more recently, deep reinforcement learning approaches. However, several persistent limitations are evident: (1) most methods are designed for single-agent systems, lacking mechanisms for scalable multi-UAV collaboration; (2) many algorithms struggle with adaptability in dynamic or large-scale environments, often requiring extensive prior environmental modeling; (3) real-time performance remains a challenge for optimization-based methods in complex settings [21]; and (4) while DRL-based methods show promise, they often face issues such as sparse rewards, training instability, and poor generalization [22]. These gaps collectively highlight the need for an adaptive, collaborative, and computationally efficient path planning framework capable of operating in dynamic, large-scale distribution networks. To address these challenges, this paper proposes a multi-agent deep reinforcement learning approach based on the MAAC (Multi-Agent Actor-Attention-Critic) algorithm, which integrates an attention mechanism to enable efficient inter-agent coordination and adaptive decision-making in real-time inspection scenarios.

To address these gaps, this study aims to develop an adaptive, collaborative path planning framework for multi-UAV inspection in complex distribution networks. The primary research goal is to leverage multi-agent deep reinforcement learning, specifically the MAAC algorithm with an attention mechanism, to enable real-time, coordinated decision-making among UAVs. We seek to answer the following research questions: (1) How can a multi-agent reinforcement learning model be effectively designed to balance inspection efficiency, energy consumption, and obstacle avoidance in dynamic environments? (2) Does the incorporation of an attention mechanism improve coordination and performance compared to existing optimization and single-agent RL methods? (3) Can the proposed approach achieve higher cumulative reward and lower computational time in realistic simulation scenarios? The intended outcome is a validated, scalable planning strategy that enhances autonomous inspection capabilities, with concrete

performance metrics demonstrating improvements in both solution quality and operational efficiency.

This paper proposes a UAV inspection path planning strategy based on reinforcement learning [23]. By constructing a reinforcement learning environment, the UAV can gradually optimize its inspection path through interaction with the environment. Reinforcement learning algorithms offer strong adaptability and high computational efficiency, effectively addressing complex environments and large-scale problems [24-25]. Through simulation comparative experiments, this paper validates the significant advantages of the proposed method in terms of cumulative discounted reward and algorithm running time, demonstrating its application potential in UAV inspection path planning.

The remainder of the paper is structured as follows: Section 2 introduces the methodology, including environment modeling, the MAAC-based multi-agent framework, and the training process. Section 3 presents the experimental setup and comparative results. Finally, Section 4 concludes the study and suggests future work.

2 Methodology

Building on the challenges outlined in Section 1, particularly the low computational efficiency and poor adaptability of existing methods in complex, large-scale environments. This section introduces a multi-agent deep reinforcement learning approach to UAV path planning. The proposed method aims to overcome these limitations through collaborative decision-making and adaptive learning. First, we model the inspection environment as a reinforcement learning problem by defining the state space, action space, and reward function. These algorithms exhibit low computational efficiency and excessively long path planning times when dealing with complex environments and large-scale problems, making it difficult to meet practical needs. To overcome these limitations, this paper proposes a UAV inspection path planning method based on multi-agent deep reinforcement learning. By introducing a multi-agent reinforcement learning model, an efficient network framework is established to achieve superior path planning.

This study introduces a multi-agent deep reinforcement learning model into UAV inspection path planning, leveraging the advantages of the Multi-Agent Actor-Attention-Critic algorithm to enable collaboration among multiple UAVs, thereby optimizing the inspection paths. The MAAC algorithm incorporates an attention mechanism, allowing each agent to fully consider the states and actions of other agents during decision-making, thus achieving more effective collaboration and superior path planning. To ensure the simulation environment reflects realistic inspection scenarios, we modeled key operational constraints — including UAV dynamics (speed, acceleration, battery consumption), sensor characteristics (GPS/IMU noise, camera field of view), and environmental variability (static/dynamic obstacles)—based on typical UAV inspection platforms and field-reported data. The AirSim platform provides high-fidelity physics and sensor simulations, which have been widely validated in prior UAV studies for realistic flight behavior and perception modeling.

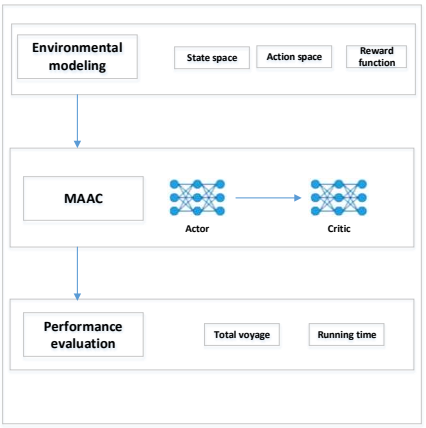

Figure 1: Algorithm framework

As shown in Figure 1, this study first defines the state space, action space, and reward function to construct a reinforcement learning model suitable for the UAV inspection task. Subsequently, leveraging the advantages of the MAAC algorithm, a multi-agent deep reinforcement learning network framework is built to achieve collaborative path planning among UAVs. Finally, simulation experiments are conducted to verify the significant advantages of the proposed method in terms of cumulative discounted reward and algorithm running time, demonstrating its potential in complex environments.

2.1 Environment modeling

We constructed a reinforcement learning model suitable for the UAV inspection task by defining the state space, action space, and reward function.

Firstly, the state space S represents the various states the UAV may encounter during inspection. To comprehensively describe the UAV's inspection task, this study assumes that The state s_t of a UAV is defined by four elements: (1) its current coordinates (x_t, y_t) on a 2D grid map; (2) the set C_t of equipment already inspected; (3) its remaining battery level E_t; and (4) environmental information ε_t, which includes obstacle locations and positions of uninspected equipment. That is, the state s_t is represented as:

$$s_t = \left((x_t, y_t), C_t, E_t, \mathrm{E_t}\right) \quad (1)$$

Where, (x_t, y_t) is the position information, C_t represents the set of inspected equipment, E_t represent the remaining battery level, and ε_t represents the environmental information.

Next, the action space A represents the actions the UAV can take in each state. Assuming the UAV can move in four directions (up, down, left, right) on a 2D grid map. Additionally, the UAV can choose to hover at its current position for inspection, thus the action space is expanded to:

$$a_t \in \{\text{up}, \text{down}, \text{left}, \text{right}, \text{stay}\} \quad (2)$$

Where, up denotes moving upward, down denotes moving downward, left denotes moving left, right denotes moving right, and stay denotes remaining at the current position.

Finally, a reward function is defined to evaluate the quality of the UAV's path during inspection. Designing a reasonable reward function is crucial for guiding the UAV to learn the optimal path. In this study, the reward function design considers factors such as energy consumption, inspection effectiveness, and obstacle avoidance mechanisms.

Energy consumption is the energy consumed by the UAV to execute an action. Assuming the UAV executes action a_t to move from position (x_t, y_t) to position (x_{t+1}, y_{t+1}), the energy consumption can be expressed as:

$$d(a_t) = \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2} \quad (3)$$

Inspection effectiveness determines whether the UAV successfully inspects new equipment through action a_t. If the UAV successfully inspects new equipment after executing the action, then, $I(a_t) = 1$, otherwise, $I(a_t) = 0$.

The obstacle avoidance mechanism aims to prevent the UAV from colliding with obstacles. Let $D(a_t)$ represent the distance to the nearest obstacle after the UAV executes action a_t. If this distance is less than a threshold, the obstacle avoidance reward is negative; otherwise, it is positive.

$$B(a_t) = \begin{cases} 1 \text{ if} D(a_t) < \text{threshold} \\ 0 \text{otherwise} \end{cases} \quad (4)$$

Specifically, the final reward function is represented as:

$$R(s_t, a_t) = \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2} + \lambda I(a_t) + \beta \begin{cases} 1 \text{ if} D(a_t) < \text{threshold} \\ 0 \text{otherwise} \end{cases} \quad (5)$$

Where λ and β are weight parameters used to balance energy consumption, inspection effectiveness, and the obstacle avoidance mechanism.

2.2 Multi-Agent deep reinforcement learning model

Multi-agent systems can improve task efficiency and effectiveness through collaboration. This study adopts a multi-agent deep reinforcement learning model to achieve collaborative path planning among UAVs. This model can fully utilize information sharing and collaboration among multiple agents, thereby achieving superior path planning.

The MAAC algorithm employs an attention mechanism within a multi-agent deep reinforcement learning framework. This mechanism allows each agent to dynamically weigh the states and actions of other agents during decision-making, thereby enhancing collaborative efficiency. The core idea of the MAAC algorithm is to use a global critic and multiple local actors, aggregating information from individual agents through the attention mechanism. The attention mechanism is particularly suited to UAV inspection, as it enables each agent to dynamically prioritize the most relevant information from other UAVs, such as their proximity to shared targets or low battery status, thereby directly enhancing cooperative behaviors like collision avoidance, inspection coverage optimization, and efficient task allocation in complex, dynamic environments.

The advantage of this algorithm is that through the attention mechanism, individual agents can share and aggregate information from each other, enhancing decision quality. Furthermore, the MAAC algorithm can adapt to different numbers and types of agents, offering high flexibility. Through the collaboration of the global critic and local actors, the MAAC algorithm can effectively handle multi-agent problems in large-scale and complex environments.

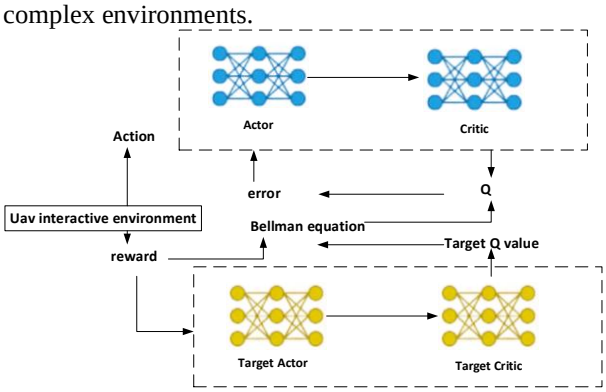

Figure 2: MAAC structure diagram

As shown in Figure 2, in the MAAC algorithm, each agent i contains an actor network π_i and a critic network Q_i. The actor network is responsible for generating the agent's actions, while the critic network is used to evaluate the value of the agent selecting a specific action in a given state. To achieve information sharing and collaboration among multiple agents, the MAAC algorithm introduces an attention mechanism to aggregate information from individual agents. In real-world deployments, such communication must also be secure and efficient. Recent batch authentication protocols from VANET research—such as L-CPPA and ECA-Vfog—demonstrate how lattice-based cryptography can enable efficient verification of multiple agents without compromising post-quantum security. While the current study assumes a trusted communication layer, integrating such lightweight batch authentication mechanisms in future work would enhance the practical resilience of multi-UAV inspection systems against spoofing and replay attacks, especially in open or adversarial network environments.

The input to the actor network is the current agent's state, and the output is an action. The actor network generates actions through a parameterized policy $\pi_{\theta_i}(a_i \mid s_i)$, where θ_i represents the parameters of the actor network. The actor network can be represented as a neural network with the structure:

$$\pi_{\theta_i}(a_i \mid s_i) = f_{\theta_i}(s_i) \quad (6)$$

Where, f_{θ_i} is a neural network mapping function.

The input to the critic network Q_i includes the state s_i and action a_i of the current agent i, as well as the state and action information of other agents. Through the attention mechanism, the critic network can aggregate information from other agents, thereby more accurately evaluating the action value of the current agent. The structure of the critic network is represented as:

$$Q_i(s_i, a_i, \{(s_j, a_j)\}_{j \neq i}) = \text{Attention}(\{(s_j, a_j)\}_{j \neq i}) f_i(s_i, a_i) \quad (7)$$

Where, Attention represents the attention mechanism, and f_i is the mapping function of the critic network.

The attention mechanism is a key component of the MAAC algorithm, used to aggregate information from multiple agents in the critic network. The input to the attention mechanism is the state-action pairs of other agents (s_j, a_j), and the output is an aggregated context vector that captures the influence of other agents on the decision-making of the current agent.

For each other agent j, an attention weight is calculated, representing the importance of agent j to agent i:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \neq i} \exp(e_{ik})} \quad (8)$$

Where, score e_{ij} is the attention score of agent j for agent i, calculated by a trainable scoring function.

Using the attention weights, a weighted sum of the information from other agents is performed to obtain the context vector:

$$c_i = \sum_{j \neq i} \alpha_{ij} (s_j, a_j) \quad (9)$$

This context vector c_i represents the aggregated influence information of other agents on the current agent i. The critic network uses this to compute the value function.

Through the above structural design, the network architecture of the MAAC algorithm can fully utilize information sharing and collaboration among multiple agents, achieving efficient path planning and decision-making.

2.3 Network training

In the MAAC algorithm, network training is performed by minimizing the Mean Squared Error loss function of the critic network. The critic network aims to estimate the state-action value function $Q_i(s_i, a_i)$ for the agent. To update the parameters of the critic network, gradient descent is performed using samples (s_i, a_i, r_i, s_i') from the experience replay buffer. The loss function for the critic network is expressed as:

$$L_i = E_{(s_i, a_i, r_i, s_i') \sim D} \left[\left(Q_i(s_i, a_i) - y_i \right)^2 \right] \quad (10)$$

Where, y_i is the target value, defined as:

$$y_i = r_i + \gamma E_{a_i' \sim \pi_i(s_i')} \left[Q_i(s_i', a_i') \right] \quad (11)$$

Where, γ is the discount factor, representing the decay rate of future rewards.

The training of the actor network is performed via the policy gradient method. The objective of the actor network is to maximize the expected state-action value function $Q_i(s_i, a_i)$. Specifically, the loss function for the actor network is expressed as:

$$J_i = E_{s_i \sim D} \left[Q_i(s_i, \pi_i(s_i)) \right] \quad (12)$$

To update the parameters of the actor network, gradient ascent is performed on:

$$\theta_i J_i = E_{s_i \sim D} \left[a_i Q_i(s_i, a_i)_{\theta_i} \pi_i(s_i) \right] \quad (13)$$

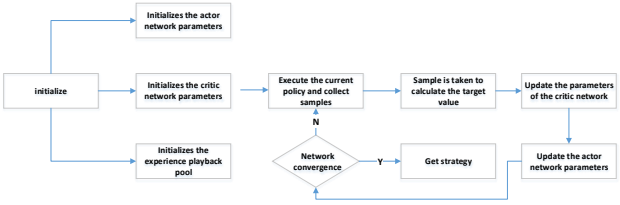

Figure 3: Algorithm Flowchart

The overall workflow of the MAAC algorithm is shown in Figure 3. First, the parameters of the actor and critic networks for all agents are initialized, and the experience replay buffer is initialized. Then, the current policy is executed in the environment to collect state transition samples, which are stored in the experience replay buffer. Next, mini-batches of samples are randomly sampled from the experience replay buffer, target values are computed, and the parameters of the critic network are updated by minimizing the MSE. Then, using the output of the critic network, the parameters of the actor network are updated via the policy gradient method. These steps are repeated until the algorithm converges.

Through the above process, the MAAC algorithm can effectively train the multi-agent deep reinforcement learning model, achieving collaborative planning among UAVs.

2.4 Practical implementation considerations

To transition from simulation to real-world deployment, several practical factors must be addressed. These include:

Communication constraints: In field environments, UAVs may operate under limited or intermittent connectivity. Our multi-agent attention mechanism can be extended to handle delayed or dropped messages using state prediction models.

Energy management: While our reward function includes battery terms, real UAVs require integration with battery-aware trajectory planners and possible recharge scheduling during long missions.

Weather and dynamic obstacles: The environmental model ε_t can be enriched with real-time weather data and moving obstacle forecasts to enhance safety during flight.

Regulatory compliance: Flight paths must respect no-fly zones and altitude restrictions, which can be incorporated as additional constraints in the state space.

To ensure the simulation environment reflects realistic inspection scenarios, we modeled key operational constraints — including UAV dynamics (speed, acceleration, battery consumption), sensor characteristics (GPS/IMU noise, camera field of view), and environmental variability (static/dynamic obstacles) — based on typical UAV inspection platforms and field-reported data. The AirSim platform provides high-fidelity physics and sensor simulations, which have been widely validated in prior UAV studies for realistic flight behavior and perception modeling.

3 Case study

3.1 Simulation environment setup

We evaluate our MAAC-based model in the Airsim simulation platform. Airsim offers high-fidelity UAV physics and sensor simulation. In this environment, we build a complex distribution network scenario with randomly placed power towers, transformers, and cables. This setup increases environmental variability and tests the robustness of our approach.

The simulation was built on the AirSim platform to replicate a realistic distribution network inspection scenario. The virtual environment contains 10 randomly placed power towers, 5 transformers, and connecting cables, creating a complex spatial layout. Each UAV was modeled with realistic flight dynamics and equipped with standard sensors (GPS, IMU, and a camera) for navigation and visual inspection. Key flight parameters (speed, acceleration, altitude) were configured to mirror real UAV operation. Additionally, static and dynamic obstacles such as trees and moving vehicles were introduced to further enhance environmental realism and test collision avoidance. Several static and dynamic obstacles, such as trees, buildings, and moving vehicles, are added to the simulation environment to simulate a realistic inspection scenario. The positions and movement trajectories of obstacles are also randomly generated. To achieve information sharing and collaboration among multiple agents, we implemented a communication mechanism between UAVs in Airsim. Each UAV can share its own state and action information through a wireless communication network, enabling collaborative path planning. The UAVs need to perform inspection tasks in this environment, planning optimal paths to cover all target equipment while avoiding collisions with obstacles and other UAVs.

Both the actor and critic networks in our implementation are three-layer fully connected neural networks. Each network contains two hidden layers with 128 and 64 neurons, respectively, and uses ReLU activation functions. This balanced architecture supports stable training while maintaining sufficient representational capacity for policy and value estimation. The scoring function for the attention mechanism employs a two-layer fully connected network with a hidden layer neuron count of 64. The discount factor γ is set to 0.95. The learning rates for both the actor and critic networks are set to 0.001, and the Adam optimizer is used for parameter updates. The capacity of the experience replay buffer is 100,000, from which mini-batch samples are randomly drawn for training with a batch size of 64. During each training episode, the parameters of the actor and critic networks are updated every 100 steps.

The hyperparameters for the MAAC algorithm were chosen based on established practices in related multi-agent deep reinforcement learning works and preliminary trials to ensure training stability. A comprehensive sensitivity analysis of hyperparameters is considered an important direction for future research.

3.2 Comparative experiments

To evaluate performance, we compared our MAAC-based model against two established algorithms: Enhanced Particle Swarm Optimization (EPSO) and Double Deep Q-Network (DDQN). EPSO improves upon standard PSO through deterministic initialization, random mutation, and edge exchange, which help avoid local optima in complex search spaces. DDQN extends the Deep Q-Network by using two separate networks to reduce overestimation bias, making it robust in dynamic decision-making tasks. Building upon traditional PSO, this algorithm increases particle diversity and optimizes the search strategy, enabling it to more effectively avoid local optima, making it particularly suitable for solving complex combinatorial optimization problems. Double Deep Q-Network is an improved algorithm based on Deep Q-Network, designed to address the overestimation problem in DQN. DDQN introduces two Q-networks—one for selecting actions (target network) and another for evaluating action values—effectively reducing Q-value bias. This algorithm performs well in many reinforcement learning tasks, especially in complex and dynamic environment decision-making problems. Table 2 presents the performance of the three models in terms of cumulative discounted reward and average algorithm running time.

Table 2: Comparative Experimental Results

	Cumulative discount Reward	**Average running time(s)**
MAAC	4.56	812.85
EPSO	2.21	833.45
DDQN	4.37	923.41

The comparative experimental results indicate that, considering both cumulative discounted reward and stable running time, the MAAC algorithm demonstrates significant advantages in the UAV inspection path planning task. Its highest cumulative discounted reward indicates that it can obtain more returns during task execution, reflecting the superiority of its path planning. Its shortest running time indicates that the algorithm possesses high real-time performance and efficiency, making it suitable for UAV inspection tasks requiring rapid responses. The cumulative discounted reward of the MAAC algorithm is 4.56, significantly higher than EPSO's 2.21 and slightly higher than DDQN's 4.37. By introducing the attention mechanism, MAAC can better integrate and utilize information among multiple agents, improving the quality of overall decision-making, thereby

achieving superior path planning and task execution in complex environments. In contrast, although EPSO has been optimized through the global search capability of the particle swarm, it has deficiencies in multi-agent collaboration, resulting in a relatively lower cumulative discounted reward. DDQN effectively reduces Q-value overestimation through its dual-network structure, performing close to MAAC, but still falls short of MAAC in multi-agent collaboration.

The stable running time of the MAAC algorithm is 812.85 seconds, outperforming EPSO's 833.45 seconds and DDQN's 923.41 seconds which demonstrating not only computational efficiency but also practical feasibility for time-sensitive missions such as post-storm grid assessment or routine maintenance windows. In a real deployment, this efficiency could translate into reduced flight time and extended battery life, allowing more equipment to be inspected per mission. The MAAC algorithm can complete tasks in the shortest time, benefiting from its efficient policy learning and information-sharing mechanism. The introduction of the attention mechanism allows the MAAC algorithm to adapt and respond to environmental changes more quickly, thereby improving overall operational efficiency. In contrast, although EPSO performs well in search and optimization, its global search characteristic leads to higher computational complexity and longer running times. DDQN, due to the additional computational overhead of its dual-network structure, has the longest running time, indicating room for improvement in its real-time performance.

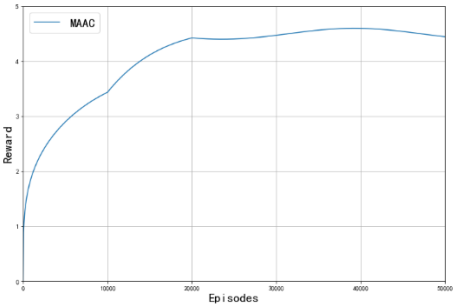

Figure 4: Model convergence curve

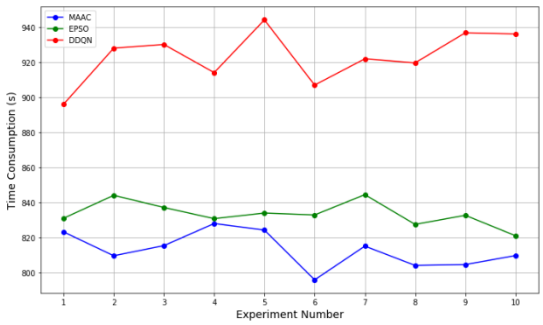

Figure 5: Model run time comparison

Figure 4 shows the reward convergence curve of the MAAC model. The reward gradually reaches its peak around 20,000 episodes, indicating that the model learns a superior policy and gradually converges. Figure 5 shows the average running times of the three models over 10 experiments, demonstrating that the MAAC algorithm's running time is almost always better than EPSO and DDQN in each trial.

3.3 Discussion

(1) Comparative Analysis with Existing Methods

Our experimental results demonstrate that the proposed MAAC-based model outperforms both EPSO and DDQN in terms of cumulative discounted reward and average running time. To contextualize these findings, we compare our approach with key methods reviewed in Section 1:

Compared with optimization-based methods (e.g., EPSO [11], PSO-based hybrids [18]): While such methods excel in static path planning, they lack adaptability in dynamic environments and do not support real-time collaborative decision-making. Our MAAC model, through its reinforcement learning foundation and attention mechanism, dynamically adapts to environmental changes and coordinates multiple UAVs without precomputed paths, leading to higher cumulative rewards and lower computational latency.

Compared with single-agent DRL methods (e.g., DDQN [20], DRL with greedy heuristics [16]): Although DDQN reduces Q-value overestimation and performs well in certain tasks, it does not explicitly model inter-agent interactions. In multi-UAV inspection scenarios, this leads to suboptimal coordination. Our integration of an attention mechanism enables each agent to weigh the states and actions of others, resulting in more efficient collective behavior and higher task rewards.

Compared with other multi-agent approaches (e.g., decentralized DDQN [20]): Prior multi-agent RL methods often treat agents independently or use simplistic communication protocols. MAAC's attention-based critic allows differentiated focus on relevant agents, improving the quality of shared decision-making and contributing to both performance and speed gains.

(2) Reasons for Performance Differences

The superior performance of our method can be attributed to the following design factors:

Attention-based collaboration: Unlike rule-based or fully decentralized coordination, MAAC allows UAVs to selectively focus on the most relevant peers, reducing redundant actions and improving path efficiency.

Adaptive reward shaping: Our composite reward function balances energy consumption, inspection progress, and obstacle avoidance, enabling sustained exploration without premature convergence.

Scalable multi-agent training: The use of a shared critic with attention reduces the observational partiality problem common in Dec-POMDPs, enabling stable learning even with varying numbers of UAVs.

(3) Novelty and Contributions

This work contributes to the field of UAV inspection path planning in the following ways:

First application of MAAC to distribution network inspection, demonstrating how attention-driven multi-agent coordination can improve both performance and efficiency in a complex, dynamic environment.

Integrated reward and state design tailored for multi-UAV inspection tasks, incorporating equipment coverage, energy constraints, and obstacle avoidance into a unified RL formulation.

Empirical validation in a high-fidelity simulation (AirSim) under realistic environmental variability, providing a reproducible benchmark for future multi-UAV inspection research.

Our approach thus advances the state-of-the-art by bridging multi-agent collaboration, attention mechanisms, and reinforcement learning into a scalable framework suitable for real-world adaptive inspection missions.

4 Conclusion

This study presents a multi-agent deep reinforcement learning approach for UAV inspection path planning in complex distribution networks. By leveraging the MAAC algorithm with an attention mechanism, the model enables coordinated multi-UAV decision-making, leading to more efficient and collision-free inspection routes. By introducing an attention mechanism, the MAAC algorithm enables effective information sharing and collaborative decision-making among multiple agents, thereby improving the quality and efficiency of path planning. The effectiveness of the proposed algorithm was verified through comparative experiments on the Airsim simulation platform. The experimental results show that the UAV inspection path planning model based on the MAAC algorithm exhibits significant advantages in both cumulative discounted reward and stable running time. Specifically, the MAAC algorithm achieved a cumulative discounted reward of 4.56, significantly higher than EPSO and DDQN. Simultaneously, the average running time of the MAAC algorithm was 812.85 seconds, outperforming both EPSO and DDQN.

Although the proposed multi-agent deep reinforcement learning approach has demonstrated promising performance in simulated inspection scenarios, several important research avenues remain open to advance the method toward real-world deployment and broaden its applicability.

(1) Algorithmic extensions for improved adaptability and efficiency

Future work can further integrate advanced reinforcement learning techniques, such as self-supervised learning and meta-learning, to improve the algorithm's ability to adapt quickly to new environments with limited interaction data. Combining model-free RL with model-based predictive control in a hybrid reinforcement learning architecture could also enhance robustness and safety in highly dynamic or uncertain settings, providing a more stable foundation for online re-planning.

(2) System-level integration for real-world operation

Practical deployment raises critical challenges that must be addressed. In real-world scenarios, communication between UAVs can be intermittent or delayed. Developing robust attention mechanisms that function under communication constraints is therefore essential. Furthermore, inspired by recent advances in VANET security, we plan to explore the integration of lattice-based batch authentication schemes—such as those in CLA-FC5G, D-BlockAuth, and L-CPPA—into the multi-UAV communication layer. This would enable the system to jointly optimize path planning and security verification, ensuring both operational efficiency and resilience against quantum-era threats in large-scale distributed inspections.

(3) Broader application across infrastructure domains

The core multi-agent coordination framework is not limited to power-grid inspection. Promising extensions include its adaptation to other critical infrastructure monitoring tasks, such as wind turbine blade inspection, photovoltaic farm monitoring, and oil-gas pipeline patrols. Each domain introduces unique constraints—e.g., flight regulations, target distributions, and risk levels—that would require tailored reformulations of the state space and reward function, offering rich opportunities for further research.

Through these research directions, we aim to bridge the gap between simulation-based validation and field-ready autonomous inspection systems, ultimately contributing to safer, smarter, and more resilient infrastructure maintenance.

References

[1] Leou, Rong-Ceng, Su, Chun-Lien & Lu, Chan-Nan. (2013). Stochastic analyses of electric vehicle charging impacts on distribution network. IEEE Transactions on Power Systems, 29(3), 1055-1063. DOI: 10.1109/TPWRS.2013.2291556.

[2] Roberge, Vincent, Tarbouchi, Mohammed & Labonte, Gilles. (2013). Comparison of Parallel Genetic Algorithm and Particle Swarm Optimization for Real-Time UAV Path Planning. IEEE Transactions on Industrial Informatics,9(1),132-141. DOI: 10.1109/TII.2012.2198665.

[3] Liu, Changan, Liu, Yang, Wu, Hua & Dong, Rui fang. (2015). A Safe Flight Approach of the UAV in the Electrical Line Inspection. International Journal of Emerging Electric Power Systems,16(5),503-515. DOI: 10.1515/ijeeps-2015-0021.

[4] Almazroi, Abdulwahab Ali, Alkinani, Monagi H.,Al-Shareeda, Mahmood A. & Manickam, Selvakumar. (2024). A Novel DDoS Mitigation Strategy in 5G-Based Vehicular Networks Using Chebyshev Polynomials. Arabian Journal for Science and Engineering,49(9),11991-12004. DOI:10.1007/s13369-023-08535-9.

[5] Abdulwahab Ali Almazroi,Mohammed A. Alqarni,Mahmood A. Al-Shareeda,Monagi H. Alkinani,Alaa Atallah Almazroey & Tarek Gaber. (2024). FCA-VBN: Fog computing-based authentication scheme for 5G-assisted vehicular blockchain network. Internet of Things,25. DOI: 10.1016/j.iot.2024.101096

[6] Almazroi, Abdulwahab Ali,Aldhahri, Eman A. A.,Al-Shareeda, Mahmood A. A. & Manickam, Selvakumar A. (2023). ECA-VFog: An efficient certificateless authentication scheme for 5G-assisted

vehicular fog computing. PLOS One,18(6). DOI: 10.1371/journal.pone.0287291.

[7] Almazroi, Abdulwahab Ali, Alqarni, Mohammed A., Al-Shareeda, Mahmood A. & Manickam, Selvakumar. (2023). L-CPPA: Lattice-based conditional privacy-preserving authentication scheme for fog computing with 5G-enabled vehicular system. PLOS One,18(10). DOI: 10.1371/journal.pone.0292690.

[8] Al-shareeda, Mahmood A., Anbar, Mohammed, Hasbullah, Iznan H., Manickam, Selvakumar, Abdullah, Nibras & Hamdi, Mustafa Maad. (2020). Review of Prevention schemes for Replay Attack in Vehicular Ad hoc Networks (VANETs),394-398. DOI: 10.1109/ICICSP50920.2020.9232047.

[9] Al-Shareeda, Mahmood A., Gaber, Tarek, Alqarni, Mohammed A., Alkinani, Monagi H.,Almazroey, Alaa Atallah & Almazroi, Abdulwahab Ali. (2025). Chebyshev Polynomial Based Emergency Conditions with Authentication Scheme for 5G-Assisted Vehicular Fog Computing. IEEE Transactions on Dependable and Secure Computing,22(5),4795-4812. DOI: 10.1109/TDSC.2025.3553868.

[10] Wang, Chuanyue,Zhang, Lei,Gao, Yifan,Zheng, Xiaoyuan,Wang, Qianling & Du, Xiaosong.(2023). A Cooperative Game Hybrid Optimization Algorithm Applied to UAV Inspection Path Planning in Urban Pipe Corridors. Mathematics,11(16),3620. DOI: 10.3390/math11163620.

[11] Manh Duong Phung, Cong Hoang Quach, Tran Hiep Dinh & Quang Ha. (2017). Enhanced discrete particle swarm optimization path planning for UAV vision-based surface inspection. Automation in Construction,81,25-33.

[12] Gonzalez de Santos, Luis Miguel, Frias Nores, Ernesto, Martinez Sanchez, Joaquin & Gonzalez Jorge, Higinio. (2021). (Indoor Path-Planning Algorithm for UAV-Based Contact Inspection). Sensors,21(2),642-642. DOI: 10.1016/j.autcon.2017.04.013.

[13] Guerrero, Jose Alfredo & Bestaoui, Yasmina. (2013). UAV Path Planning for Structure Inspection in Windy Environments. Journal of Intelligent & Robotic Systems,69(1-4),297-311. DOI: 10.1007/s10846-012-9778-2.

[14] Kun Li, Xinxin Yan & Ying Han. (2024). Multi-mechanism swarm optimization for multi-UAV task assignment and path planning in transmission line inspection under multi-wind field. Applied Soft Computing,150. DOI: 10.1016/j.asoc.2023.111033.

[15] Wang, Ju, Wang, Guoqiang, Hu, Xiaoxuan,Luo, He & Xu, Haiqing. (2020). Cooperative Transmission Tower Inspection with a Vehicle and a UAV in Urban Areas. Energies,13(2),326-326. DOI: 10.3390/en13020326.

[16] Yan, Chao, Xiang, Xiaojia & Wang, Chang. (2020). Towards Real-Time Path Planning through Deep Reinforcement Learning for a UAV in Dynamic Environments. Journal of Intelligent & Robotic Systems,98(2),297-309. DOI: 10.1007/s10846-019-01073-3.

[17] Lv, Hui, Chen, Yadong,Li, Shibo,Zhu, Baolong & Li, Min. (2024). Improve exploration in deep reinforcement learning for UAV path planning using state and action entropy. Measurement Science and Technology,35(5),056206-056206. DOI:10.1088/1361-6501/ad2663.

[18] Wang, Chuanyue, Zhang, Lei, Gao, Yifan,Zheng, Xiaoyuan,Wang, Qianling & Du, Xiaosong.(2023). A Cooperative Game Hybrid Optimization Algorithm Applied to UAV Inspection Path Planning in Urban Pipe Corridors. Mathematics,11(16),3620. DOI: 10.3390/math11163620.

[19] Yue, Longfei,Yang, Rennong,Zhang, Ying,Yu, Lixin & Wang, Zhuangzhuang.(2022). Deep Reinforcement Learning for UAV Intelligent Mission Planning. Complexity,2022(0). DOI: 10.1155/2022/3551508.

[20] Bayerlein, Harald, Theile, Mirco,Caccamo, Marco & Gesbert, David. (2021). Multi-UAV Path Planning for Wireless Data Harvesting with Deep Reinforcement Learning. IEEE Open Journal of the Communications Society,2,1171-1187. DOI: 10.1109/OJCOMS.2021.3081996.

[21] Zhang, Xin,Zou, Pingguo,Ma, Chi,Zhang, Zhentao,Guo, Hongbin,Chen, Yabin & Cheng, Zikun.(2022). Inspection and Classification System of Photovoltaic Module Defects Based on UAV and Thermal Imaging, 905-909. DOI: 10.1109/ICPRE55555.2022.9960506.

[22] Jianyang Li, Hui Wu, Chunhua Hu & Chenglie Yu. (2021). A Fault Diagnosis System Based on Case Decision Technology for UAV Inspection of Power Lines. IOP Conference Series: Earth and Environmental Science,632(4). DOI 10.1088/1755-1315/632/4/042077.

[23] Richard S. Sutton & Andrew G. Barto. (1998). Reinforcement Learning: An Introduction.

[24] Hua-Yun Xiao & Cong-Qiang Liu. (2004). Chemical characteristics of water-soluble components in TSP over Guiyang, SW China, 2003. Atmospheric Environment,38(37),6297-6306. DOI: 10.1016/j.atmosenv.2004.08.033.

[25] He, Ying, Zhang, Zheng,Yu, F. Richard,Zhao, Nan,Yin, Hongxi,Leung, Victor C. M. & Zhang, Yanhua. (2017). Deep-Reinforcement-Learning-Based Optimization for Cache-Enabled Opportunistic Interference Alignment Wireless Networks. IEEE Transactions on Vehicular Technology,66(11),10433-10445. DOI: 10.1109/TVT.2017.2751641.

Appendix

Table 1: Summary of related works on uav inspection path planning

Reference	Method / Approach	Key Contribution	Scenario / Test Setting	Key Metrics / Results	Main Limitations
[11] Phung et al.	Enhanced Discrete PSO	Deterministic initialization, random mutation, edge exchange for TSP	UAV vision-based surface inspection	Path length, convergence time	Single UAV only; no multi-agent coordination
[12] Luis et al.	Real-time path planning with point cloud processing	Reduced algorithm execution time for indoor contact inspection	Indoor UAV contact inspection	Execution time, path accuracy	Limited to indoor environments; not suitable for large-scale outdoor networks
[13] Guerrero et al.	Grid division + Zermelo-TSP	Automated inspection point generation and time-optimal path planning	Structure inspection in windy environments	Inspection time, path optimality	Requires predefined grid; not adaptive to real-time changes
[14] Li et al.	Improved Bidirectional ACO + Discrete Honey Badger Algorithm	Multi-UAV path planning under multi-windfield conditions	Transmission line inspection in wind fields	Path safety, energy consumption	High computational cost; limited scalability
[15] Wang et al.	Odd-Even Layer Genetic Algorithm	Effective UAV path planning with layered optimization	Cooperative inspection with UAV and vehicle	Coverage rate, path smoothness	Not designed for highly dynamic environments
[16] Yan et al.	DRL with global situational info + greedy heuristic	Path planning in dynamic environments with real-time adaptability	Dynamic obstacle avoidance	Success rate, path length	Sparse reward problem; no multi-agent support
[17] Lv et al.	Information-theoretic exploration based on state/action entropy	Improved exploration for UAV path planning with sparse rewards	UAV navigation in unknown environments	Exploration efficiency, reward consistency	Focused on single-agent; limited to simulation validation
[18] Wang et al.	Improved distributed DRL with LSTM networks	Decomposed navigation tasks to avoid non-convergence	UAV navigation with partial observability	Training stability, task completion rate	Complex architecture; high training cost
[19] Yue et al.	End-to-end DRL with domain randomization & policy entropy maximization	Enhanced generalization for UAV mission planning	Multi-task UAV planning	Generalization performance, mission success rate	Requires extensive simulation training
[20] Bayerlein et al.	Multi-Agent RL (DDQN) formulated as Dec-POMDP	Decentralized path planning for multi-UAV data harvesting	Wireless data harvesting with UAVs	Data collection rate, energy efficiency	No explicit attention mechanism; limited collaboration modeling