Available Online: 27 May 2026 DOI: 10.54254/2977-3903/2026.33855

Multi-UAVs collaborative inspection methods based on coarse-grained filtering and fine-grained image acquisition

Jiayu Qiu*, Minghao Li

School of Computer Science, Guangdong University of Science and Technology, Dongguan, China

*Corresponding Author. Email: qiuchaocaho@gmail.com

Abstract. In existing multi-Unmanned Aerial Vehicle (UAV) inspection methods, UAVs are typically limited to image acquisition tasks and lack real-time understanding of captured content. As a result, they are unable to identify suspected defect areas on-site or perform immediate fine-grained image acquisition of suspicious targets, which may lead to the omission of critical information and inefficient resource utilization. To address these issues, this paper proposes a multi-UAV cooperative inspection method based on coarse-grained global screening and fine-grained local image acquisition. The method endows UAVs with preliminary image understanding capabilities, enabling them to assess captured content in real time during flight and dynamically adjust image acquisition strategies accordingly. By leveraging reinforcement learning, an intelligent inspection loop of "coarse screening first, fine-grained capture later" is achieved. Experimental results demonstrate that the proposed method outperforms traditional approaches and other reinforcement learning-based methods in key metrics such as task duration, resource utilization efficiency, and inspection quality. It provides a new perspective for multi-UAV cooperative inspection in dynamic environments.

Keywords: multi-UAV inspection, coarse-grained global screening, fine-grained local image acquisition, reinforcement learning

1. Introduction

In recent years, with the accelerating transformation toward industrial intelligence and the maturation of autonomous unmanned system technologies, Unmanned Aerial Vehicles (UAVs) have been increasingly deployed in industrial inspection applications. A single UAV, benefiting from its flexibility and high mobility, has already been successfully applied in many standardized scenarios. However, when confronted with increasingly complex and large-scale modern industrial infrastructure, the limitations of single-UAV operation in terms of coverage capability, operational efficiency, and task robustness have become increasingly evident. Against this backdrop, multi-UAV cooperative inspection systems have emerged as a prominent research direction in both academia and industry, as they can significantly improve inspection efficiency through parallel operation [1]. The core objective of multi-UAV cooperative inspection is to achieve efficient, complete, and high-quality visual data acquisition of target areas under constrained time and resources. Existing studies primarily focus on optimizing coverage efficiency and system energy consumption. From a technical perspective, research in this field has gradually evolved from traditional optimization methods

toward more flexible and adaptive reinforcement learning-based approaches [2]. In terms of traditional optimization methods, researchers have mainly employed swarm intelligence algorithms such as genetic algorithms, particle swarm optimization, and ant colony optimization for solution search and path planning [3, 4]. These methods can achieve relatively high coverage efficiency and path quality in scenarios with complete environmental information and fixed task constraints. Their advantage lies in clear physical interpretability and well-defined modeling structures. However, traditional optimization approaches typically rely on accurate environmental modeling and globally synchronized information. In dynamically changing inspection environments, they suffer from model mismatch and high computational overhead for replanning, making it difficult to meet real-time responsiveness and adaptive adjustment requirements.

With the breakthrough of deep reinforcement learning in sequential decision-making problems, data-driven multi-UAV cooperative inspection methods have become an emerging research direction. In this paradigm, each UAV is modeled as an intelligent agent that learns cooperative strategies through continuous interaction with the environment, without requiring an explicitly defined mathematical model. Researchers have explored deep Q-networks, actor-critic frameworks, and their multi-agent extensions (e.g., MADDPG, QMIX) to address cooperative path planning, dynamic task allocation, and collision avoidance [5, 6]. Reinforcement learning-based methods demonstrate strong robustness and generalization ability in partially observable and communication-constrained environments, making them particularly suitable for complex inspection tasks where environmental models are unknown or difficult to construct. However, these methods still face challenges such as unstable multi-agent training, low sample efficiency, and weak interpretability of learned policies. In addition, a significant gap remains between simulation and real-world deployment [7-9].

Despite the substantial progress achieved in improving operational efficiency, a closer analysis reveals that existing methods still suffer from two unresolved critical issues:

1) UAVs are typically assigned only simple image acquisition and transmission tasks, while all data analysis and defect detection are performed offline at a centralized backend server after the inspection mission. This approach not only causes a large amount of redundant non-defective images to continuously occupy limited wireless communication bandwidth, but also imposes heavy computational and storage burdens on the server when processing massive image streams from dozens of UAVs simultaneously. In some cases, it even leads to delays in defect analysis results, significantly affecting the timeliness of subsequent maintenance decisions and failing to meet the real-time and high-precision requirements of industrial inspection.

2) Most existing methods configure uniform imaging parameters for UAVs without differentiating acquisition strategies according to the importance of inspection regions. In critical defect-prone areas such as power line insulators, oil and gas pipeline welds, and bridge bearings, insufficient imaging resolution may lead to incomplete or unclear defect details. Conversely, in non-critical regions such as regular pole sections, straight pipeline segments, or non-joint areas of bridge girders, indiscriminate over-sampling results in unnecessary energy and bandwidth consumption. This creates a fundamental contradiction between insufficient data acquisition in critical regions and redundant data acquisition in non-critical regions.

To address the above issues, this paper integrates computer vision and multi-agent reinforcement learning techniques and proposes a solution from two perspectives: onboard real-time screening and multi-UAV cooperative fine-grained inspection. The main contributions are as follows:

1) To address the lack of real-time image understanding capability in UAVs and their inability to make on-site judgments about captured content, a lightweight YOLO-based object detection network is deployed on each UAV. Considering the limited computational resources of onboard platforms, depthwise separable convolutions are used to replace standard convolutions, and channel pruning and INT8 quantization are applied for model compression. These optimizations ensure that defect detection accuracy is maintained while

inference time is reduced to within 40 ms, satisfying real-time onboard execution requirements. During routine inspection, UAVs perform fast coarse-grained global screening of captured images and output a defect confidence score. When the confidence exceeds an adaptive threshold, the region is identified as a suspected defect area, triggering a fine-grained local acquisition mechanism. This enables a transition from passive indiscriminate imaging to active intelligent screening, fundamentally reducing redundant image generation and transmission.

2) To achieve high-precision and multi-angle acquisition of suspected defect regions and address the challenges of multi-UAV coordination and resource allocation in fine-grained inspection, a MADDPG-based multi-agent reinforcement learning framework is constructed. This framework enables UAV swarms to make autonomous collaborative decisions during fine-grained inspection tasks. A composite action space is designed, including imaging actions, coordination actions, and motion actions. In addition, a multi-objective reward function is formulated to jointly consider image acquisition quality, multi-UAV cooperation effectiveness, and energy/bandwidth consumption, guiding UAVs toward both high-quality imaging and efficient resource utilization. Furthermore, a closed-loop feedback mechanism driven by coarse screening results is introduced. Real-time screening information and fine inspection states are fed back to a central controller, which dynamically updates the importance weights of task points, generates fine inspection tasks, and inserts them into the inspection queue. Priority is then given to suspected defect regions in subsequent task allocation, concentrating limited energy, bandwidth, and computational resources on regions of higher value, thereby forming an intelligent inspection loop of "coarse screening first, fine-grained inspection later."

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 presents the overall system modeling; Section 4 details the proposed coarse screening result-driven closed-loop feedback mechanism; Section 5 introduces the reinforcement learning-based joint optimization algorithm; Section 6 presents numerical simulation experiments and analysis; and Section 7 concludes the paper.

2. Related work

2.1. Multi-UAV cooperative inspection based on traditional methods

Multi-UAV cooperative inspection is an important research direction in the field of unmanned systems. In the early development of cooperative inspection methods, approaches based on traditional optimization models and rule-based strategies constituted the primary technical pathway [10]. These methods typically formulate the problem as classical combinatorial optimization problems such as the multiple Traveling Salesman Problem (mTSP) or coverage path planning, and solve them using centralized or distributed strategies [11].

In centralized optimization approaches, it is commonly assumed that a global central node exists, and solutions are obtained using swarm intelligence-based heuristic algorithms. For instance, Xu et al. proposed a particle swarm optimization algorithm considering communication constraints for multi-UAV cooperative path planning [12]. The method achieved path cost optimization under static constraints; however, its computational efficiency suffers significantly in dynamic scenarios due to frequent replanning, leading to considerable overhead. Such methods can guarantee theoretical solution quality when accurate global information is available, but their computational complexity increases exponentially with problem scale, and they rely heavily on stable and reliable global communication.

In distributed cooperative approaches, research focuses on achieving self-organization through local rules. With the emergence of models such as Graph Neural Networks (GNNs), some studies have attempted to integrate them with traditional optimization methods to improve distributed coordination performance. For example, Wang et al. proposed leveraging graph neural networks to extract topological features and combining

them with Proximal Policy Optimization (PPO) for cooperative path planning, thereby enhancing coordination capability in partially observable environments [13].

Although such methods improve system scalability and robustness, they still depend on accurate environmental modeling. When task patterns change abruptly or strong disturbances occur, performance degradation or coordination failure may still arise. Overall, although traditional methods offer advantages in model rigor and interpretability, their performance heavily depends on the accuracy of predefined models and the degree of environmental structure. In dynamic and uncertain industrial inspection scenarios, these methods exhibit clear limitations in real-time adaptability and cooperative flexibility, particularly in handling variations in communication constraints and unexpected task insertions [14, 15].

2.2. Multi-uav cooperative inspection based on reinforcement learning

With the advancement of Deep Reinforcement Learning (DRL), researchers have begun exploring multi-agent DRL to achieve adaptive cooperative control of UAVs. For example, Westheider et al. proposed a multi-agent DRL-based adaptive path planning method [16], which significantly improved multi-UAV task planning capabilities in 3D environments through cooperative learning mechanisms and effectively addressed coordination among different UAVs. Similarly, Shen et al. introduced the DNQMIX method for multi-UAV target search problems [14], which enhances coverage rate and target capture efficiency through a centralized training and decentralized execution framework, demonstrating the effectiveness of DRL in complex UAV cooperative tasks. In addition, Xu et al. proposed a cooperative pursuit planning method for multi-UAV systems based on a distributed actor-critic architecture [17], achieving higher success rates and faster capture times, further illustrating the potential of multi-agent reinforcement learning in complex dynamic decision-making.

Despite these advances, existing studies still exhibit insufficient attention to two key issues. First, most methods do not consider real-time onboard image processing capabilities, requiring all captured images to be transmitted to ground stations for analysis, which prevents on-site identification of suspected defect regions and timely adjustment of imaging strategies. Second, existing work largely ignores the heterogeneous requirements of different inspection regions regarding image quality, applying uniform acquisition parameters to both critical and non-critical areas. This results in a coexistence of insufficient information capture in key regions and redundant data acquisition in ordinary regions. Therefore, under bandwidth-constrained multi-UAV cooperative inspection scenarios, it is of significant research value to design a joint decision-making framework that integrates onboard real-time screening with differentiated image acquisition strategies.

3. System model

3.1. Task scenario description

This section establishes a formal model of the multi-UAV cooperative inspection system and clearly defines the core optimization problem to be addressed. The system consists of a heterogeneous UAV cluster and a set of inspection tasks. The UAV cluster includes two types of roles:

Inspection UAVs responsible for image acquisition, denoted as i UAVs in total, as shown in Equation (1):

$$U = \{U_1, U_2, U_3 \dots U_i\} \tag{1}$$

A data collection UAV U_0, which receives data and serves as the central coordinator.

UAVs operate in a three-dimensional space, and their trajectories are represented as discrete temporal position sequences (where k denotes the maximum execution step), as shown in Equation (2). The position of

UAV U_i at time t, denoted as $p_i(t)$, is determined by its spatial coordinates $(x_i(t), y_i(t), z_i(t))$:

$$p_i(t) = (x_i(t), y_i(t), z_i(t)), t = 1, 2, \ldots, k \tag{2}$$

The energy consumption of image-acquisition UAVs $Ui(i > 0)$ is proportional to flight distance, hovering time, and the number of imaging operations. As shown in Equation (3), $E_i(t)$ denotes the remaining energy of UAV U_i at time t, and $n_i^{cap}(t)$ denotes the imaging action performed by UAV U_i:

$$E_i(t + \Delta t) = E_i(t) - (||p_i(t + \Delta t) - p_i(t)|| + \Delta t + n_i^{cap}(t)) \tag{3}$$

Inspection tasks are discretized into a set of target points requiring high-quality imaging (e.g., industrial equipment and building surfaces), as shown in Equation (4):

$$T = \{T_1, T_2, T_3 \ldots T_j\} \tag{4}$$

Each task point T_j represents an object requiring inspection and is characterized by a center position $O_j = (x_j, y_j, z_j)$, an inspection radius r_j, volumetric information v_j, and multiple inspection regions.

3.2. Spherical mapping and task representation

In multi-UAV cooperative inspection tasks, target objects are typically complex three-dimensional structures, such as equipment surfaces, architectural facades, or power transmission towers.

To unify the representation of inspection targets and camera viewpoints across different UAVs, this paper introduces a spherical task representation model. By performing spherical segmentation and spatial projection, inspection targets in Euclidean 3D space are mapped onto a spherical coordinate domain and further projected onto a two-dimensional plane, enabling unified representation and hierarchical allocation of imaging tasks. Specifically, a spherical surface is constructed with the center of the target object as the origin. The surface of the object is then projected onto the sphere, followed by labeling of key regions and normal regions. Finally, the labeled spherical surface is unfolded into a 2D plane. Importance weights are assigned to different surface elements according to local inspection priority, as illustrated in Figure 1.

3.2.1. Spherical Segmentation Definition

Let the geometric center of task object T_j be O_j. A local spherical coordinate system is established with O_j as the origin. To enable discretization, the spherical surface is divided into $m * n$ surface elements. All surface elements under task T_j form a set C_j. The position $p_i(t)$ of UAV U_i at time t and its camera orientation d_i define a projection function:

$$F(p_i(t), d_i) = c_{d,b} \tag{5}$$

That is, when UAV U_i is located at $p_i(t)$ and captures images in direction d_i, the corresponding spherical surface elements within its field of view are $c_{d,b}(0 < d \leq m, 0 < b \leq n)$. Each surface element $c_{d,b}$ has a binary state: if $c_{d,b} = 1$, it has been captured; if $c_{d,b} = 0$, it has not yet been captured.

To emphasize inspection priority across different regions, the importance weight of each surface element $c_{d,b}$ is defined as:

$$\omega_{d,b} = \begin{cases} \omega_{\text{major}}, & \text{if } c_{d,b} \text{ are located in key areas} \\ \omega_{\text{normal}}, & \text{if } c_{d,b} \text{ are located in normal areas} \\ 0, & \text{if } c_{d,b} \text{ needs no inspection} \end{cases} \tag{6}$$

In the proposed inspection framework, task importance is modeled as a spatial weighting attribute over surface elements. As shown in Equation (6), each discretized element $c_{d,b}$ is assigned a weight $\omega_{i,j}$ according to structural criticality, where key regions ω_{major} are assigned significantly higher weights than normal

regions ω_{normal}, i.e., $\omega_{major} > \omega_{normal} > 0$. Under this weighting scheme, only elements with $\omega_{d,b} > 0$ are considered in task allocation and reinforcement learning decision-making, enabling prioritized coverage under limited resources and ensuring that critical regions receive higher probability and higher-quality observations.

The advantage of spherical task representation lies in transforming arbitrary complex 3D objects into a continuous angular domain, thereby enabling task allocation and reinforcement learning decisions within a bounded state space.

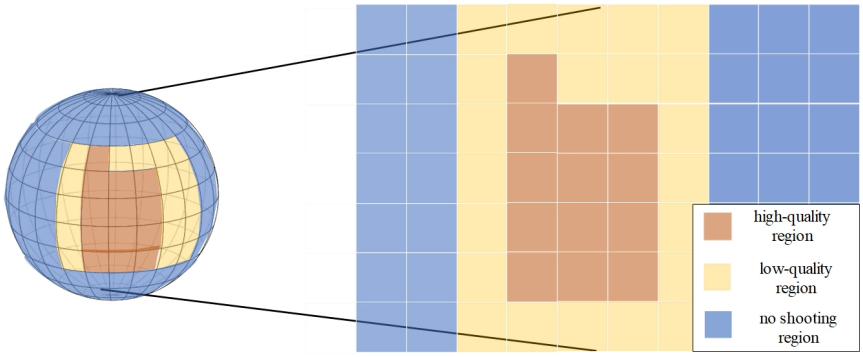

Figure 1. Spherical segmentation model

3.2.2. UAV imaging model

The imaging decision $n_i^{cap}(t)$ of UAV U_i at time t updates the state of surface elements under the current viewpoint. The update rule is given in Equation (7):

$$c_{d,b}(t+1) = \begin{cases} 1, & \text{if } c_{d,b} = F(p_i(t), d_i) \text{ and } c_{d,b}(t) = 0 \\ c_{d,b}, & \text{otherwise} \end{cases} \tag{7}$$

This update rule enables dynamic state evolution of inspection targets. Once a surface element is completed or does not require imaging, it is excluded from repeated allocation, effectively reducing task redundancy and energy consumption.

3.2.3. Relationship between image quality and bandwidth

Let the raw image quality $Q_i^{shot}(t)$ captured by inspection UAV $U_i (i > 0)$ at time t be:

$$Q_i^{shot}(t) = f(p_i(t), \theta_i(t), \phi_i(t)) \tag{8}$$

where $\theta_i(t), \phi_i(t)$ represent the elevation and azimuth angles of UAV U_i relative to the target surface element $c_{d,b}$. Thus, image clarity depends on both the relative distance and the shooting angle between the UAV and the target object.

In practical scenarios, image quality degradation may occur due to limited communication bandwidth. Under bandwidth constraints, images must be compressed, and the resulting effective image quality $Q_i^{eff}(t)$ is given by:

$$Q_i^{eff}(t) = Q_i^{shot}(t) \cdot g(\frac{band_i(t)}{band_{req}}) \tag{9}$$

where $band_i(t)$ denotes the uplink transmission rate allocated to UAV U_i, and $band_{req}$ denotes the minimum transmission rate required for high-quality image transmission. The function $g(\cdot)$ represents the bandwidth factor, whose input is the ratio of the actual bandwidth currently allocated to the UAV $band_i(t)$ to the minimum bandwidth required for high-definition images $band_{req}$. When $band_i(t) \geq band_{req}$, $g(\cdot) = 1$; otherwise, $g(\cdot) < 1$, to quantify the quality degradation caused by insufficient bandwidth.

3.3. Coarse-grained global screening model based on YOLO

Each UAV is equipped with a lightweight YOLO detection network as a coarse screening module to perform real-time analysis of aerial images. YOLO is a one-stage object detection algorithm that reformulates the detection task as a regression problem. It performs detection through a single forward pass, simultaneously outputting object class labels and bounding box coordinates. Owing to its high inference speed and strong detection accuracy, it is well-suited for real-time onboard deployment.

As shown in Figure 2, for images captured by UAV U_i at time t, the YOLO network outputs a detection result set $P_i(t) = [P_i^0(t), P_i^1(t), ..., P_i^K(t)]$, where $P_i^K(t)$ denotes the confidence score for class K, and K is the total number of defect categories defined in the model, including cracks, corrosion, loosening, deformation, etc.

The coarse screening confidence $c_{d,b}^{coarse}(t)$ is defined as the maximum probability over all defect categories:

$$c_{d,b}^{coarse}(t) = max_{k=1,2,\ldots,K} P_i^K(t) \tag{10}$$

This confidence value represents the likelihood that defects exist in the current image. If no defect is detected, the value is set to 0, and the maximum value is bounded by 1. The triggering condition for fine-grained image acquisition is defined as $c_{d,b}^{coarse}(t) > \vartheta$, where ϑ is a screening threshold. The threshold adopts an adaptive strategy: for key regions ($\omega_{d,b} = \omega_{major}$), a lower threshold ϑ_{low} is used to ensure a higher detection guarantee rate; for normal regions, a higher threshold ϑ_{high} is applied to avoid excessive fine-grained inspection.

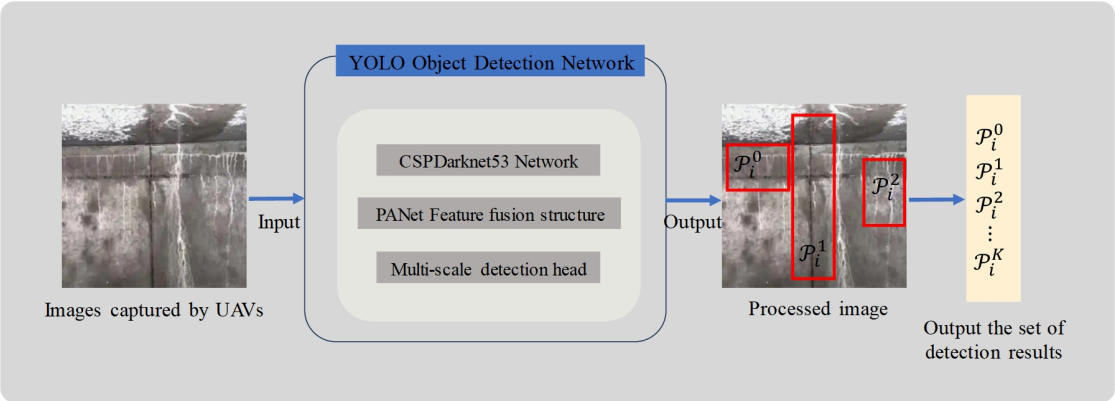

Figure 2. YOLO-based confidence detection model

When the coarse-grained screening confidence exceeds the predefined threshold, the system activates the fine-grained image acquisition mechanism to perform multi-angle, high-resolution imaging of suspected defect regions. The purpose of fine-grained local acquisition is to obtain sufficiently clear defect evidence, providing high-quality visual data for subsequent defect recognition and assessment. Fine-grained actions include two categories: single-UAV adaptive acquisition and cooperative acquisition, corresponding to scenarios where a single UAV is sufficient or multi-UAV coordination is required.

When a single UAV can complete the fine-grained task by adjusting its own pose and camera parameters, an onboard adaptive acquisition strategy is adopted. Let UAV U_i perform fine imaging on target surface element $c_{d,b}$. The fine-grained image quality $Q_{d,b}^{fine}$ is related to the shooting distance $d_{i,c}$ and camera focal length f. According to optical imaging principles, under other conditions being equal, a shorter distance leads to a larger number of pixels occupied by the target on the image sensor, resulting in higher resolution. A longer

focal length increases optical magnification, thereby improving detail clarity. Therefore, the fine-grained image quality is positively correlated with focal length and inversely proportional to the shooting distance, as expressed in Equation (11):

$$Q_{d,b}^{fine} = \frac{f}{d_{i,c}} \tag{11}$$

In practical inspection scenarios, once the UAV reaches the vicinity of a suspected defect region, it can reduce flight altitude or move closer to shorten the imaging distance, while adjusting the focal length to obtain higher-resolution images. If the predicted image quality under the current position and focal length reaches or exceeds the quality threshold Q_{th} , the UAV performs image capture; otherwise, it continues adjusting its position or attempts alternative viewing angles.

When a single UAV cannot fully cover the target region from multiple perspectives, neighboring UAVs can be requested for assistance, enabling cooperative acquisition. Cooperative acquisition allows multiple UAVs to capture the same surface element from different viewpoints, enabling more complete defect information through multi-view fusion. Let the set of UAVs participating in cooperative imaging be denoted as U_{coop}. The final fine-grained quality of surface element $c_{d,b}$ is obtained by fusing image quality from multiple viewpoints, as shown in Equation (12):

$$Q_{d,b}^{final} = \frac{1}{|U_{coop}|} \sum_{U_i \in U_{coop}} Q_{d,b}^{fine} \tag{12}$$

$Q_{d,b}^{final}$ represents the arithmetic mean of image quality across UAVs and reflects the baseline fused image quality. This design encourages multi-UAV collaboration. In practical applications, when defects exhibit directional characteristics (e.g., cracks propagating along specific orientations), multi-angle imaging can capture more complete morphological information, and the fused result often provides superior quality compared to single-view observations.

The objective of this study is to enable adaptive cooperation between coarse-grained global screening and fine-grained local image acquisition under joint optimization of inspection efficiency and image quality. This differs from existing approaches that rely on fixed thresholds or post hoc analysis-based passive inspection strategies. By endowing UAVs with preliminary image understanding capability, they can assess captured content in real time during flight and dynamically adjust imaging strategies accordingly. This allows limited energy, bandwidth, and computational resources to be concentrated on truly critical suspected defect regions. While ensuring defect detection performance, unnecessary resource consumption is significantly reduced, ultimately forming an intelligent inspection loop of "coarse screening first, fine-grained inspection later."

4. Fine-grained local image acquisition decision method based on multi-agent reinforcement learning

4.1. MADDPG model definition

This section formulates the fine-grained image acquisition decision problem as a Multi-Agent Markov Decision Process (MMDP) and provides a detailed definition based on the MADDPG framework. MADDPG adopts a centralized training and decentralized execution paradigm. During training, the critic network has access to the states and actions of all agents and learns a global value function; during execution, the actor network makes decisions based only on local observations, thereby adapting to communication-constrained deployment environments.

The multi-UAV cooperative fine-grained imaging decision problem is modeled as an MMDP defined as the tuple $\langle U, S, A, P, R, \gamma \rangle$, where U is the set of UAVs, S is the joint state space, A is the joint action space, P is the state transition probability, R is the reward function, and γ is the discount factor. Unlike the single-agent reinforcement learning setting, this problem involves multiple interacting agents. The policy optimization of each agent depends not only on its own behavior but also on the decisions of other agents, which makes the environment non-stationary from the perspective of any individual agent.

The core optimization objective of the fine-grained image acquisition problem is to maximize the overall benefit of multi-UAV cooperative fine inspection under constraints such as energy, bandwidth, and time. This objective can be formulated as:

$$max_{\pi_1,\pi_2\ldots,\pi_n} \sum_{t=0}^{T} \gamma^t \sum_{i=1}^{N} r_1(t) \tag{13}$$

where π_i denotes the policy of UAV U_i , $r_1(t)$ is the instantaneous reward at time t, $\gamma \in [0,1]$ is the discount factor, and T is the task horizon. This objective represents the maximization of the expected cumulative discounted reward from the initial state to the end of the task. By maximizing this quantity, the UAV swarm can learn cooperative strategies that balance multiple objectives under resource constraints.

The local state of each UAV U_i at time t, denoted as $s_i(t) \in S$, includes the following components:

(1) Fine-grained task information: This includes the target surface element $c_{d,b}^{target}$, task priority p, deadline $t_{deadline}$, and the set of completed imaging angles $\Phi_i^{done}(t)$. The completed angle set records previously captured viewpoints to avoid redundant imaging. When the coverage of completed angles reaches a predefined requirement, the UAV may terminate the fine-grained task early.

(2) Target region information: This includes the position of the target surface element and the set of UAVs currently participating in cooperative imaging U_{coop}. When multiple UAVs collaborate, their imaging angles must be non-overlapping to maximize viewpoint diversity.

(3) Self-state information: This includes current position $p_i(t)$, remaining energy B_i, and available bandwidth $band_i$. These variables reflect the UAV's operational capability and resource constraints. When energy is low, the UAV should prioritize returning for recharging rather than continuing fine-grained inspection; when bandwidth is limited, high-definition image uploading should be postponed.

(4) Neighbor UAV information: This includes the positions $p_j(t)$ of nearby UAVs within communication range, their current task types, whether they are performing fine-grained inspection, and their completed imaging angles. This information is obtained via local communication and supports cooperative decision-making. Upon receiving assistance requests, a UAV decides whether to respond based on its own state.

(5) Environmental state information: This includes global task completion progress, overall system bandwidth utilization, and total energy consumption. These variables allow UAVs to perceive the system-level operational status and make decisions aligned with global objectives.

The joint state $S(t) = \{s_1(t), s_2(t), \ldots, s_N(t)\}$ forms the global system state used for critic network training.

The action of each UAV U_i at time t, denoted as $a_i(t)$, is a composite action consisting of three dimensions, fully describing its behavior during fine-grained inspection:

(1) Imaging action: $a_i^{shoot}(t) \in \{0,1,2\}$, where 0 indicates no imaging, 1 indicates imaging, and 2 indicates adjusting camera resolution. The decision depends on whether the current viewpoint is optimal and whether redundant imaging has occurred.

(2) Cooperation action: $a_i^{coop}(t) \in \{0,1,2\}$, where 0 indicates independent execution, 1 indicates requesting assistance, and 2 indicates accepting assistance. This mechanism enables flexible collaborative behavior among UAVs.

(3) Flight action: $a_i^{move}(t)$ represents the desired movement direction in 3D space, enabling UAVs to adjust trajectories to reach optimal imaging positions while maintaining safe distances from neighboring UAVs.

The composite action is defined as: $a_i(t) = \{a_i^{shoot}(t), a_i^{coop}(t), a_i^{move}(t)\}$, serves as the output of the Actor network. The joint action $A(t) = \{a_1(t), a_2(t), \ldots, a_i(t)\}$ collectively influences the next system state.

The state transition probability $P : S \times A \to [0, 1]$ describes the probability of transitioning from the current state to the next state under joint actions. The transition process is influenced by the following factors:

(1) Position update: After the drone performs a flight maneuver, its position is updated according to the kinematic model:

$$p_i(t+1) = p_i(t) + a_i^{move}(t) \cdot \Delta t \tag{14}$$

(2) Energy update: Performing an action consumes energy, and the remaining energy is updated to

$$B_i(t+1) = B_i(t) - e_{fly} \cdot |a_i^{move}(t)| - e_{shoot} \cdot 1_{\{a_i^{shoot}=1\}} \tag{15}$$

(3) Imaging state update: If a photo is taken, record the current shooting angle and image quality:

$$\Phi_i^{done}(t+1) = \Phi_i^{done}(t) \cup \{\Phi_i(t)\} \tag{16}$$

$$Q_{d,b}^{fine}(t+1) = max(Q_{d,b}^{fine}(t), Q_{new}) \tag{17}$$

(4) Cooperation state update: If a drone sends or receives a request for assistance, update the collaboration set:

$$U_{coop}(t+1) = U_{coop}(t) \cup \{U_i\} \vee U_{coop}(t) - \{U_i\} \tag{18}$$

(5) Bandwidth update: Bandwidth used when uploading high-resolution images:

$$band_i(t+1) = band_i(t) - band_{fine} \cdot 1_{\{upload\}} \tag{19}$$

The reward function serves as the key bridge between the optimization objective and the reinforcement learning framework, transforming the global objective into instantaneous feedback signals. For fine-grained image acquisition, three reward components are designed:

Fine-grained imaging quality reward: This term corresponds to maximizing image quality in the optimization objective. A positive reward is assigned when the UAV successfully captures high-quality images, and the reward is proportional to both image quality and defect importance:

$$r_i^{quaility}(t) = \omega_{d,b} \cdot Q_{d,b}^{fine}(i) \cdot 1_{\{a_i^{shoot}(t)=1\}} \tag{20}$$

Cooperative completion reward: This term corresponds to maximizing viewpoint coverage and ensuring task timeliness. When multiple UAVs successfully cooperate to complete multi-view coverage of the same target, all participating UAVs receive an additional reward. This encourages UAVs to actively collaborate when necessary, leveraging multi-agent cooperation:

$$r_i^{collab}(t) = \frac{U_{coop}}{N_{max}} \cdot 1_{\{Q_{d,b}^{final} \geq Q_{th}\}} \tag{21}$$

Resource consumption penalty: This term corresponds to minimizing resource consumption in the optimization objective, penalizing excessive energy usage:

$$r_i^{energy}(t) = -\Delta B_i(t) \tag{22}$$

The total reward is defined as a weighted sum of the above components, directly corresponding to the instantaneous reward $r_i(t)$ in the optimization objective:

$$r_i(t) = \varphi_1 r_i^{quaility}(t) + \varphi_2 r_i^{collab}(t) + \varphi_3 r_i^{energy}(t) \tag{23}$$

4.2. Maddpg-based reinforcement learning model training

MADDPG is a multi-agent reinforcement learning method developed from the Actor–Critic framework and specifically designed to address multi-agent cooperative decision-making problems. By introducing a centralized training and decentralized execution paradigm, the algorithm effectively mitigates the non-stationarity problem inherent in multi-agent environments.

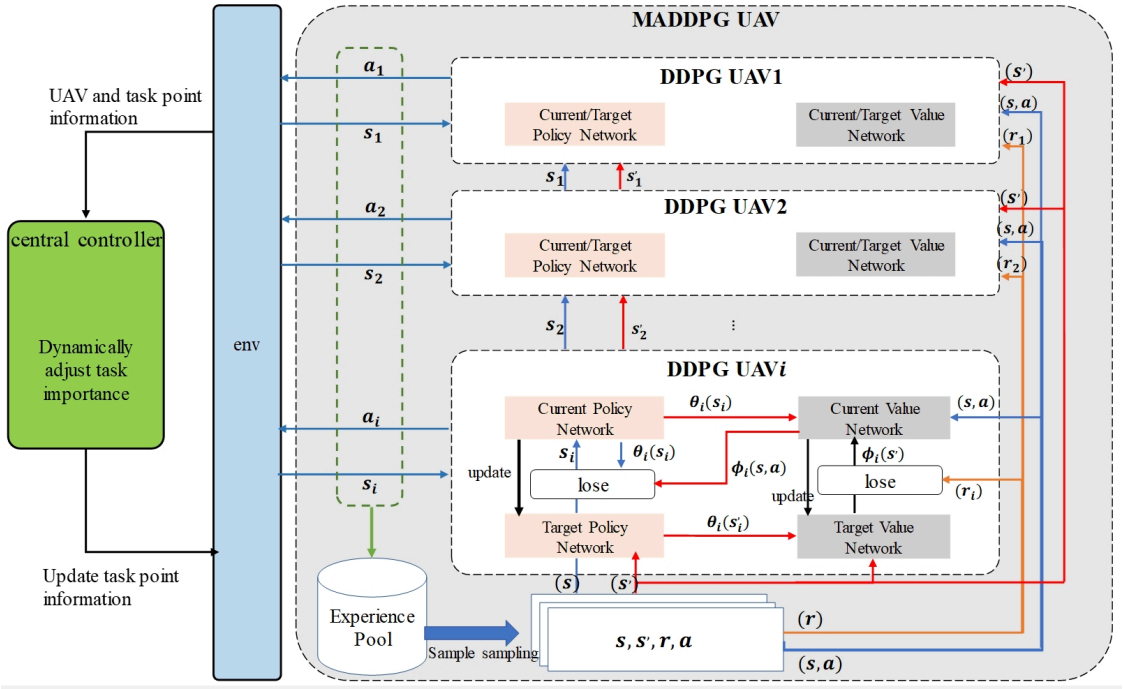

Figure 3. MADDPG architecture diagram

In a multi-agent system, the policy of each agent is continuously updated, which makes the environment appear non-stationary from the perspective of any individual agent. Traditional single-agent reinforcement learning methods struggle to handle such dynamic variations. MADDPG addresses this issue through a centralized critic network. Specifically, during training, the critic network of each agent can access the states and actions of all agents to learn a global value function. During execution, however, the actor network relies only on local observations to make decisions, enabling fully decentralized execution. The core advantage of centralized training with decentralized execution lies in the following: global information during training enables the learning of more effective cooperative policies, while decentralized execution allows real-time decision-making based solely on local observations, thereby satisfying communication constraints in real-world inspection scenarios. More specifically, the critic network takes the states and actions of all agents as input and captures global environmental dynamics, providing more accurate gradient signals for the actor network. In contrast, the actor network only receives local observations and outputs actions independently of other agents, which enhances system scalability and robustness. Based on these characteristics, this paper adopts the MADDPG algorithm for policy optimization to improve the cooperative decision-making performance of UAVs, as shown in Figure 3.

Table 1. MADDPG model training algorithm

1	Randomly initialize actor network parameters θ_i and critic network parameters ϕ_i for all UAVs; initialize target network parameters $\theta_i' \leftarrow \theta_i, \phi_i' \leftarrow \phi_i$; initialize replay buffer $H = \emptyset$.
2	for l = 1, 2, ... , L do
3	Reset the simulation environment and obtain the initial joint state S_l ;
4	for each UAV U_i in U do
5	Select action $a_i(t)$ according to the current policy and exploration rate.
6	end for
7	Execute the joint action $A(t) = \{a_1(t), a_2(t), \ldots, a_i(t)\}$;
8	Compute the immediate reward $r_i(t)$ for each UAV and observe the next state S_{l+1} ;
9	Store the transition tuple $(S(t), A(t), R(t), S(t+1))$ into the replay buffer H
10	if the capacity of buffer H exceeds the predefined threshold then
11	Sample complete trajectory data from buffer H
12	Update the actor network using the MADDPG objective function:
13	Estimate the state-action value under the current policy using the critic network.
14	Compute the policy objective based on the policy ratio and advantage function.
15	Update actor network parameters θ_i using an optimizer.
16	Update the critic network using mean squared error loss:
17	Compute the value function loss.
18	Update critic network parameters ϕ_i using an optimizer.
19	end if
20	Update UAV states and task point information.

Table 1 presents the pseudocode of the MADDPG-based fine-grained image acquisition decision algorithm. The algorithm aims to enable autonomous cooperation among UAVs in fine inspection tasks through multi-agent reinforcement learning, optimizing joint decisions over imaging parameters, cooperation strategies, and flight trajectories to maximize image quality and resource utilization efficiency. The algorithm begins by randomly initializing the parameters of all UAV actor networks and critic networks, initializing target networks, and setting the experience replay buffer to empty (Line 1). Lines 2–21 describe the training loop, which runs for L episodes. At the beginning of each episode, the simulation environment is reset and the initial joint state S_l is obtained (Line 3). Each UAV then selects an action $a_i(t)$ based on its current policy and exploration rate, including imaging actions, cooperation actions, and motion actions (Lines 4–6). After executing the joint action $A(t)$, the immediate reward $r_i(t)$ for each UAV is computed and the next state S_{l+1} is observed (Lines 7–8). The experience tuple is stored in the replay buffer (Line 9). When the buffer size exceeds a predefined threshold (Line 10), complete trajectory data are sampled from the buffer (Line 11). The actor network is updated using the MADDPG objective: the critic network first estimates the value of the current state–action pairs under the current policy, and then the actor network is optimized based on the policy gradient derived from the advantage function (Lines 12–15). Meanwhile, the critic network is updated using the mean squared error loss. The value function loss is computed and optimized accordingly (Lines 16–18). At the end of each episode, UAV and task state information is updated (Line 20). Through this training process, the UAV swarm learns an optimal cooperative policy for fine-grained image acquisition. In the early stage of training, UAV behaviors are largely random, exploring diverse action combinations. As training progresses,

the policy gradually converges toward efficient cooperative strategies that balance image quality and resource consumption. After training, the actor networks are deployed onboard UAVs, enabling decentralized fine-grained decision-making during real-world inspection tasks and effectively adapting to communication constraints and dynamic environmental changes.

Through the above training process, the UAV swarm is able to learn an optimal cooperative policy for fine-grained image acquisition. In the early stage of training, UAV behaviors are largely stochastic, exploring a wide range of possible action combinations. As training progresses, the policy gradually converges toward an efficient cooperative pattern, which maintains high imaging quality while effectively controlling resource consumption. After training, the actor networks are deployed onboard each UAV. During real-world inspection tasks, UAVs can then perform decentralized fine-grained image acquisition decision-making based solely on local observations. This enables effective adaptation to practical scenarios characterized by communication constraints and dynamic environmental changes in inspection missions.

5. Experimental setup and results analysis

5.1. Experimental setup

This section validates the effectiveness of the proposed multi-UAV cooperative inspection method based on coarse-grained global screening and fine-grained image acquisition through simulation experiments. The experimental design is divided into three parts: first, evaluating the detection performance of the coarse-grained global screening module; second, assessing the convergence behavior and cooperative efficiency of the MADDPG-based fine-grained local image acquisition decision method; and third, conducting ablation studies to analyze the impact of the closed-loop feedback mechanism on overall system performance.

During the simulation environment construction, a multi-UAV inspection scenario consisting of 40 inspection task points and seven UAVs is established. Each task point is discretized into multiple surface elements based on the spherical mapping model, where key elements account for 30% of the total, with importance weights set as $\omega_{major} = 0.8$ and $\omega_{major} = 0.3$. The coarse-grained screening module adopts YOLOv5 as the backbone network. After lightweight optimization for onboard deployment, the model size is compressed to 2.1M parameters, and the inference time per frame is reduced to less than 40 ms. The training dataset consists of 6,800 annotated images of equipment surface defects, split into training and validation sets in an 8:2 ratio.

The inspection space is defined as a three-dimensional region of 1,000×1,000×200 units. Initial UAV positions and task point locations are randomly generated within the region, subject to a minimum safety distance constraint $d_{min} > 10$ units to ensure spatial separability and task schedulability. Each UAV is initialized with an energy budget of 100 units, where the energy consumption is set to 0.5 units per unit flight distance and 0.2 units per imaging operation. The total system bandwidth is set to 100 bps, and the minimum bandwidth required for fine-grained image upload is 25 bps. The reinforcement learning hyperparameters are configured as follows: discount factor $\gamma = 0.95$, actor learning rate 10^{-4}, critic learning rate 10^{-3}, number of training episodes $E = 2000$, and maximum steps per episode $T = 500$. The exploration strategy adopts an ε-greedy policy with an initial exploration rate of 0.2.

5.2. Performance evaluation of the reinforcement learning model

Figure 4 to 6 jointly illustrate the convergence behavior of the MADDPG-based fine-grained image acquisition decision model under three different scenario scales. From the average reward curve in Figure 4, it can be observed that in all three scenarios, the reward increases steadily with the number of training episodes

and eventually stabilizes. Among them, the small-scale scenario converges the fastest and achieves the highest final reward value. The medium-scale scenario ranks second, while the default-scale scenario converges more slowly and stabilizes at approximately 152. These results indicate that the proposed model exhibits strong scalability. The Actor loss curve in Figure 5 and the Critic loss curve in Figure 6 further validate the training stability. In all scenarios, both loss values decrease rapidly within the first 500 episodes and then stabilize without significant oscillations. Moreover, the convergence points of the loss curves are consistent with those of the reward curves, demonstrating that the model can effectively learn cooperative fine-grained inspection policies and maintain stable convergence even as the scenario scale increases.

From Figure 4–6, it can be observed that the reinforcement learning algorithm demonstrates clear advantages in both convergence performance and model stability. The proposed approach not only achieves faster convergence and approaches optimal solutions more rapidly, but also exhibits superior control over model loss, indicating higher stability and accuracy.

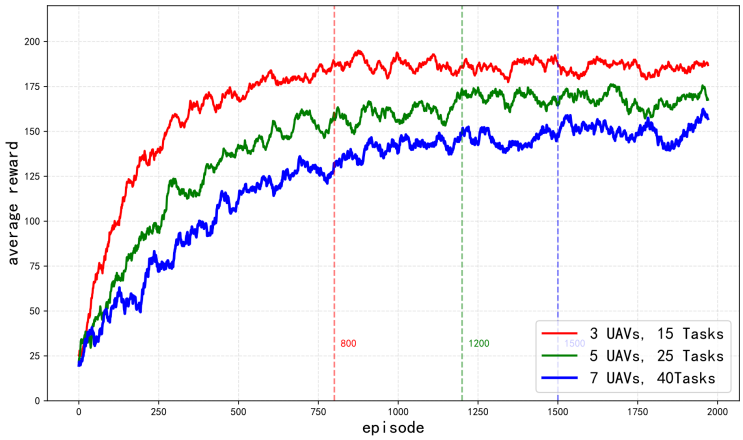

Figure 4. Average reward convergence curve

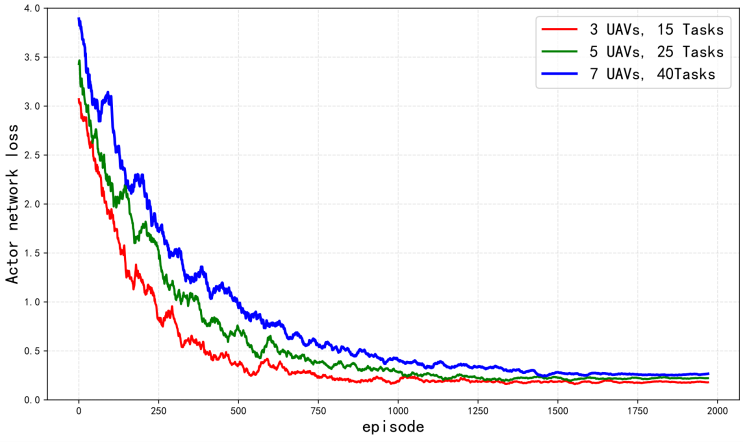

Figure 5. Actor network loss curve

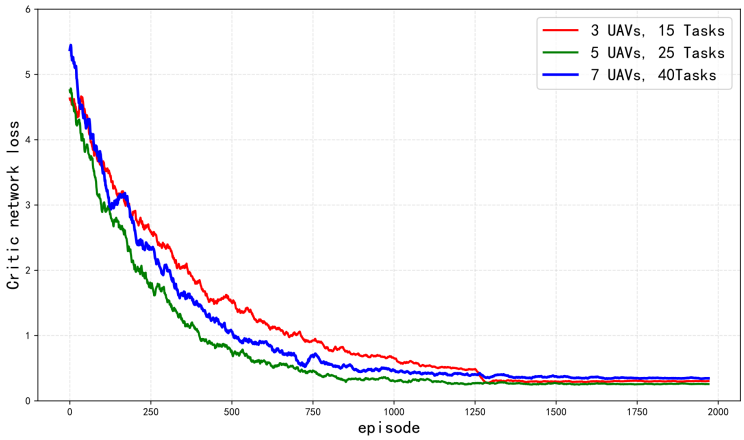

Figure 6. Critic network loss curve

5.3. Task execution performance comparison analysis

To verify the effectiveness of the proposed algorithm, three representative baseline methods are selected for comparative experiments: Greedy-Fine, IQL, and Centralized-DDPG. Greedy-Fine is a greedy strategy-based method [18], where the UAV immediately flies to the nearest optimal imaging position upon detecting a suspected region. If the image quality is insufficient, it simply adjusts its position and repeats imaging without any cooperative planning. Independent Q-Learning (IQL) allows each UAV to independently learn its own policy while treating other UAVs as part of the environment without explicit coordination modeling [19]. Centralized-DDPG serves as a centralized control baseline, where a central controller makes unified decisions for all UAVs. It aims at global optimality but incurs high communication overhead. The proposed MADDPG-Fine method adopts a centralized training and decentralized execution paradigm. During training, global information is used to learn cooperative strategies, while during execution, only local observations are required for real-time decision-making. This enables multi-UAV autonomous cooperative fine-grained inspection in dynamic environments. A comprehensive comparison is conducted in terms of task completion time, image quality score, and communication overhead to evaluate the trade-off between cooperative efficiency and imaging quality.

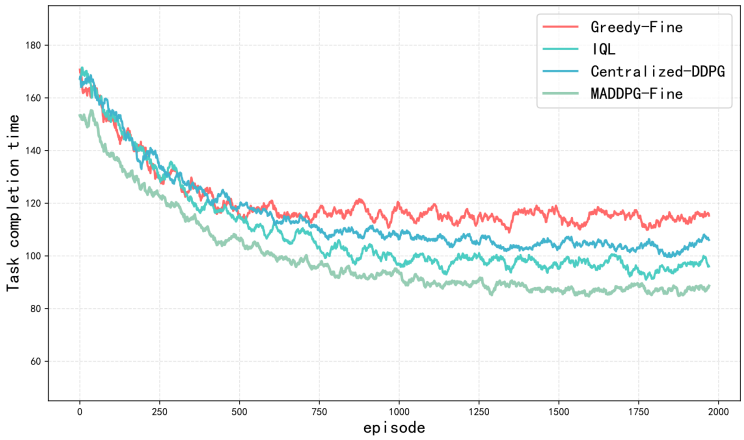

Figure 7. Task completion time comparison

As shown in Figure 7, Greedy-Fine exhibits the longest task completion time. This is because its greedy strategy lacks global planning, causing UAVs to repeatedly adjust positions without effectively reaching optimal imaging angles. IQL shows improvement over Greedy-Fine; however, since each UAV learns independently and treats others as environmental noise, cooperative efficiency remains limited. Centralized-DDPG further reduces task completion time, as centralized control enables global optimization of trajectories. However, real-time decision-making introduces significant communication overhead. In contrast, the proposed MADDPG-Fine achieves the shortest completion time. This is attributed to its centralized training and decentralized execution mechanism. The model learns cooperative strategies during training and relies only on local observations during execution, effectively reducing path detours and waiting time among UAVs during fine-grained inspection.

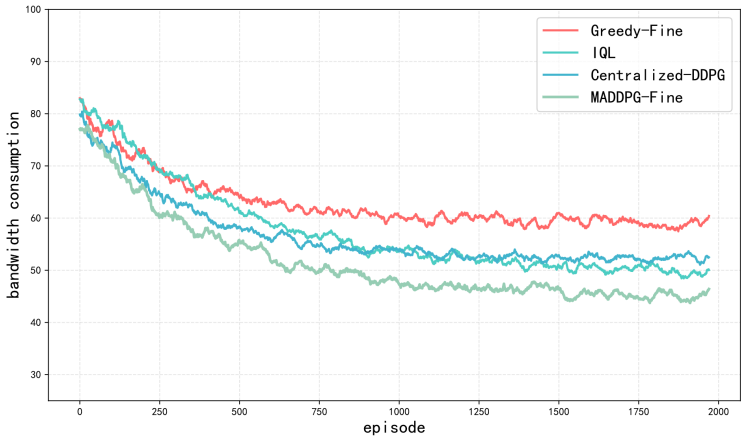

Figure 8. Total bandwidth consumption comparison

Figure 8 illustrates the variation of total bandwidth consumption across the four algorithms. Greedy-Fine consistently exhibits the highest bandwidth usage due to repeated position adjustments and multiple image uploads, caused by its inability to evaluate imaging quality effectively, resulting in excessive redundant transmissions. IQL performs slightly better than Greedy-Fine, but still suffers from redundant coverage due to the absence of cooperative mechanisms. Centralized-DDPG achieves lower bandwidth consumption by globally scheduling imaging tasks and reducing duplicate transmissions. The proposed MADDPG-Fine achieves the lowest bandwidth consumption. This improvement is mainly due to two design factors: (1) the coarse-grained global screening mechanism filters out a large number of defect-free images locally, transmitting only fine-grained images that require further analysis; (2) the multi-agent cooperative mechanism enables UAVs to coordinate imaging angles through local communication, fundamentally avoiding redundant image acquisition and associated bandwidth waste.

Figure 9 shows the variation of average UAV movement distance over training episodes for the four algorithms. Greedy-Fine exhibits only limited improvement. Due to its lack of global planning, UAVs directly fly toward the nearest imaging point upon detecting suspected regions. If image quality is insufficient, they repeatedly adjust positions, resulting in disordered trajectories and longer travel distances. Moreover, the absence of cooperation forces individual UAVs to cover multiple angles independently, further increasing movement cost. IQL performs significantly better than Greedy-Fine. Although independent learning allows UAVs to develop improved policies through trial and error, the lack of coordination leads to path conflicts and redundant movements when multiple UAVs converge to the same region, limiting further optimization. Centralized-DDPG outperforms Greedy-Fine but is slightly inferior to IQL. While centralized control enables

global path planning, communication latency and limited responsiveness to dynamic environmental changes reduce real-time performance, preventing optimal trajectory execution. The proposed MADDPG-Fine achieves the best performance among all methods. This is due to its centralized training and decentralized execution framework: the critic network learns global cooperative strategies during training, while each UAV makes real-time decisions based solely on local observations during execution, effectively eliminating communication delays. In addition, multi-agent cooperation enables UAVs to coordinate flight paths and imaging angles autonomously, avoiding detours and waiting delays, thereby achieving the shortest average movement distance.

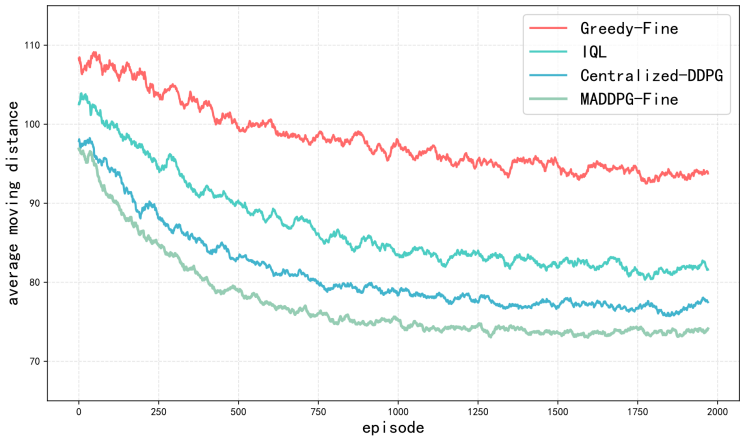

Figure 9. Average movement distance comparison

As shown in Figure 10, Greedy-Fine converges quickly but achieves the lowest image quality score due to its fixed rule-based strategy, which cannot adaptively adjust imaging strategies based on defect characteristics. IQL achieves slight improvement, as trial-and-error learning enables better imaging angle selection; however, the lack of multi-UAV cooperation limits viewpoint coverage. Centralized-DDPG enables global coordination and multi-UAV cooperative planning, but communication delays and limited real-time responsiveness hinder performance under dynamic imaging conditions. MADDPG-Fine achieves the highest image quality score. This is attributed to its multi-agent cooperative mechanism, which dynamically adjusts imaging resolution based on real-time observations. Multiple UAVs capture the target region from different viewpoints, and multi-view fusion provides more complete and clearer defect evidence, significantly improving image quality.

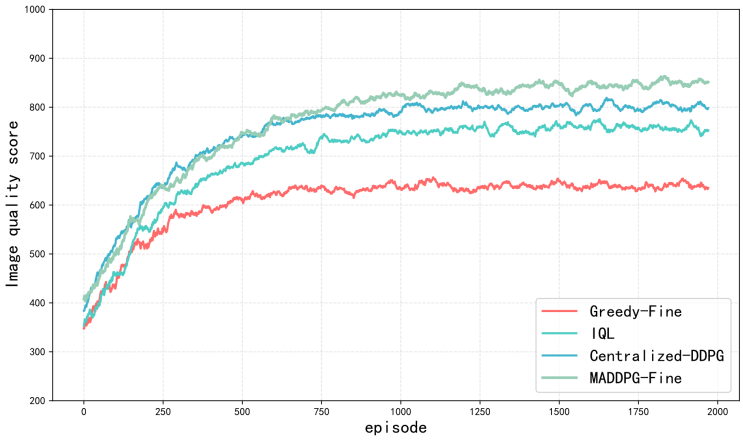

Figure 10. Image quality score comparison

Across all four figures, the proposed MADDPG-Fine algorithm demonstrates consistent superiority over baseline methods. By integrating coarse-grained global screening with fine-grained adaptive image acquisition, it achieves significant improvements in cooperative robustness, path efficiency, resource utilization, and final inspection quality, providing a reliable technical solution for practical multi-UAV cooperative inspection applications.

6. Conclusion

This paper addresses two key challenges in multi-UAV cooperative inspection systems: the lack of real-time image understanding capability on UAVs and the homogeneous processing of inspection regions. To tackle these issues, an intelligent inspection framework based on coarse-grained global screening and fine-grained image acquisition is proposed. Experimental results demonstrate that the proposed MADDPG-Fine framework significantly outperforms both traditional methods and existing reinforcement learning approaches across multiple performance metrics. In terms of real-time screening capability, the coarse-grained global screening module filters out a large number of defect-free images locally and transmits only fine-grained inspection results that require further analysis, thereby effectively reducing communication bandwidth consumption. In terms of fine-grained inspection efficiency, the average task completion time of UAVs is reduced by more than 25% compared with baseline methods. In terms of image acquisition quality, the system achieves detection accuracy exceeding 0.85 across different defect categories, significantly outperforming comparative algorithms. These results comprehensively demonstrate the advantages of the proposed method in improving cooperative efficiency, data acquisition quality, and resource utilization. The main contribution of this study lies in the integration of onboard real-time visual screening and multi-agent cooperative fine-grained inspection within a unified optimization framework. By introducing a two-stage cooperative mechanism of "coarse screening followed by fine-grained inspection," the proposed method enables precise acquisition of suspected defect regions under resource-constrained conditions. This provides a solution that is both theoretically rigorous and practically applicable for real-world deployment of multi-UAV intelligent inspection systems. Future work may further improve the adaptability, robustness, and generalization capability of the proposed approach by incorporating more advanced attention mechanisms, extending experiments to real-world physical platforms, and integrating continuous dynamic task scenarios. These directions will further promote the application of UAV swarm intelligence in complex scenarios such as industrial inspection, disaster response, and urban governance.

References

[1] Rahman, M., Sarkar, N. I., & Lutui, R. (2025). A survey on multi-UAV path planning: Classification, algorithms, open research problems, and future directions. *Drones, 9*(4), Article 263.

[2] Zhang, H., Zhao, J., Yu, J., Yu, R., Liu, W., Wang, D., & Liu, D. (2023). Application of UAV in intelligent patrol inspection of transmission line. In *International Conference on Mechatronics Engineering and Artificial Intelligence (MEAI 2022)* (Vol. 12596, Article 1259626). SPIE. https: //doi.org/10.1117/12.2672208

[3] Liu, Y., Li, X., Wang, J., Wei, F., & Yang, J. (2024). Reinforcement-learning-based multi-UAV cooperative search for moving targets in 3D scenarios. *Drones, 8*(8), Article 378. https: //doi.org/10.3390/drones8080378

[4] Huo, L., Wang, Z., Xu, M., & Song, Y. (2023). A task-agnostic regularizer for diverse subpolicy discovery in hierarchical reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53*(3), 1932–1944. https: //doi.org/10.1109/TSMC.2022.3209070

[5] Hou, Z., Fei, J., Deng, Y., & Xu, J. (2021). Data-efficient hierarchical reinforcement learning for robotic assembly control applications. *IEEE Transactions on Industrial Electronics, 68*(11), 11565–11575. https: //doi.org/10.1109/TIE.2020.3038072

[6] Ekechi, C. C., Elfouly, T., Alouani, A., & Khattab, T. (2025). A survey on UAV control with multi-agent reinforcement learning. *Drones, 9*(7), Article 484. https: //doi.org/10.3390/drones9070484

[7] Bayerlein, H., Gafvert, M., & Karaman, S. (2020). Multi-UAV path planning for wireless data harvesting with deep reinforcement learning. *IEEE Transactions on Robotics, 37*(4), 1346–1362.

[8] Li, X., Yao, L., Li, M., & Zhang, B. (2025). Reinforcement learning based collaborative path planning for UAVs and unmanned vehicles. In Proceedings of the *2025 2nd International Conference on Machine Learning and Intelligent Computing* (Vol. 278, pp. 595–603). PMLR.

[9] Liu, J., Zhang, Q., & Wang, W. (2025). Dual-timescale hierarchical MADDPG for multi-UAV cooperative path planning. *Applied Intelligence, 55*(3), 123–145.

[10] Wang, X., Yi, M., Liu, J., Zhang, Y., Wang, M., & Bai, B. (2023). Cooperative data collection with multiple UAVs for information freshness in the Internet of Things. *IEEE Transactions on Communications, 71*(5), 2740–2755. https: //doi.org/10.1109/TCOMM.2023.3255240

[11] Kim, M., Lee, H., Hwang, S., Debbah, M., & Lee, I. (2024). Cooperative multiagent deep reinforcement learning methods for UAV-aided mobile edge computing networks. *IEEE Internet of Things Journal, 11*(23), 38040–38053. https: //doi.org/10.1109/JIOT.2024.3447090

[12] Xu, L., Wang, H., & Ming, J. (2023). Cooperative path planning optimization for multiple UAVs considering communication constraints with PSO. *Knowledge-Based Systems, 249*, Article 110164.

[13] Wang, J., Tian, G., & Wu, Y. (2024). Multi-UAV cooperative path planning via graph neural networks and proximal policy optimization. *IEEE Transactions on Neural Networks and Learning Systems, 35*(2), 345–356.

[14] Shen, G., Lei, L., Zhang, X., Li, Z., Cai, S., & Zhang, L. (2023). Multi-UAV cooperative search based on reinforcement learning with a digital twin driven training framework. *IEEE Transactions on Vehicular Technology, 72*(7), 8354–8368. https: //doi.org/10.1109/TVT.2023.3245120

[15] Zhou, Y., Zhang, X., & Li, Z. (2024). Multi-UAV cooperative 3D path planning with communication constraints: A deep reinforcement learning approach. *IEEE Transactions on Vehicular Technology, 73*(5), 6821–6836.

[16] Westheider, J., Rückin, J., & Popović, M. (2023). Multi-UAV adaptive path planning using deep reinforcement learning. *IEEE Robotics and Automation Letters, 8*(4), 2120–2127.

[17] Xu, Y., Wei, Y., & Jiang, K. (2023). Multi-UAV dynamic path planning with dueling deep Q-network under communication denial. *Mathematics, 11*(2), Article 405.

[18] Cui, J., Liu, Y., & Nallanathan, A. (2019). Multi-agent reinforcement learning based resource allocation for UAV networks. *IEEE Transactions on Wireless Communications, 18*(4), 2045–2059.

[19] Ting-ting, T., & Chen, Y. (2025). Autonomous decision-making of UAV cluster with communication constraints based on multi-agent deep deterministic policy gradient. *Journal of Cloud Computing, 14*(1), Article 12.