Received 23 October 2024; revised 19 November 2024; accepted 25 November 2024. Date of publication 28 November 2024;
date of current version 20 December 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3507818

Optimization of Communication Quality for Energy-Limited Inspection AAV: A Hybrid Algorithm

WEI WANG[1], JIANGLING CAO[1], DINGCHENG YANG[1] (Member, IEEE), HAO HE[2], AND ZHIHAI XU[1] (Member, IEEE)

[1]School of Information Engineering, Nanchang University, Nanchang 330031, China
[2]School of Oceanography, Jiangsu University of Science and Technology, Zhenjiang 212003, China

CORRESPONDING AUTHORS: D. YANG AND Z. XU (e-mail: yangdingcheng@ncu.edu.cn; xuzhihai@ncu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62261035, Grant 62461040, Grant 62401241, and Grant 62061027; and in part by Jiangxi Science Technology Foundation under Grant 20223BCJ25016, Grant 20224BBC3100, Grant 20242BCC32016, and Grant 20223BBE51035.

ABSTRACT In this paper, we study a autonomous aerial vehicle (AAV) inspection system. In this system, the AAV flies to all inspection points in a certain area for patrol inspection, and the energy of inspection AAV is limited. Our goal is to optimize the communication quality of the AAV by planning the inspection sequence and flight trajectory, so as to ensure that the AAV can complete the inspection task and minimize the outage time subject to limited energy of the AAV. To solve this problem, we propose a hybrid algorithm, which consists of simulated annealing (SA) algorithm and Dueling Double Deep Q Network (D3QN) algorithm. The SA algorithm is used to obtain the inspection sequence of the AAV with the most energy saving. On this basis, the D3QN algorithm is used to optimize the flight trajectory of the energy-limited inspection AAV. To prove the effectiveness of the sub-optimal solution obtained by our proposed algorithm, we use several algorithms as a comparison. Numerical results show that the proposed algorithm is effective in optimizing the communication quality of the inspection AAV with limited energy, and its performance is improved by about 15%-50% compared with other benchmarks.

INDEX TERMS Dueling double deep Q network (D3QN), energy-limited inspection AAV, simulated annealing (SA), AAV inspection system.

I. INTRODUCTION

With the advantages of low cost, high efficiency and high flexibility, autonomous aerial vehicles (AAVs) have shown great application value in many fields [1], such as agricultural plant protection [2], emergency rescue [3], data collection [4], logistics and distribution [5]. Traditional AAVs usually use autonomous navigation systems and satellite navigation systems for flight control. However, in some specific cases, the communication and data transmission efficiency of AAVs is often limited by the existing traditional communication infrastructure [6]. In order to overcome this limitation, cellular-connected AAVs emerge as the times require [7], which refer to AAVs that communicate and control through cellular networks. Cellular-connected AAV is a promising technology [8].

In recent years, the application of cellular-connected AAVs in inspection scenarios has also become a research hotspot. Since the AAV can quickly cover a large area and transmit relevant inspection data in real time, the inspection efficiency is greatly improved and the cost is reduced compared with the traditional manual inspection method. Especially in areas with complex terrain and inconvenient transportation, the advantages of inspection AAV are more obvious. Therefore, inspection AAV has been applied in many industries such as electric power, forestry and water conservancy and other industries [9], [10], [11], [12], which has brought many positive effects to these industries. This proves that the research on inspection AAV is of great significance.

However, there are still many challenges in the practical inspection applications of cellular-connected AAVs. It mainly focuses on communication quality and energy efficiency. On the one hand, it is obvious that cellular-connected AAVs depend on the signal coverage of the cellular base station (BS), and are also greatly affected

by the tilt of the BS's antennas [13]. In the cellular signal coverage hole area, the network signal may be weak or unstable, affecting the data transmission and control command reception of the AAV. Many researchers have worked to solve the problem of communication signal connection [14], [15], [16], [17], [18], [19], [20], [21]. On the other hand, energy consumption is also a important factor that must be considered. The flight time of AAV is usually limited by the battery energy it carries. Generally speaking, limited energy will affect the relevant performance of the AAV [22]. At present, there are also many studies dedicated to optimizing the performance of AAV with limited energy [23], [24], [25], [26], [27], [28].

A. RELATED WORKS

In this part, we review the latest researches on cellular-connected AAV relevant to this study. These related works can be classified into two parts: 1) communication quality optimization of cellular-connected AAV, 2) performance optimization of limited energy AAV.

1) *Communication Quality Optimization of Cellular-Connected AAV*: Zhang et al. [14] optimized the trajectory of AAVs, ensuring their continuous wireless connectivity with at least one BS at every time instant. Bulut and Guevenc [15] considered the connection quality of the BS-AAV link. Their goal is to design the AAV's flight path to minimize mission completion time. Zeng et al. [16] investigated a cellular-connected AAVs system and proposed a simultaneous navigation and radio mapping (SNARM) approach to minimize the weighted sum of flight time and outage duration. Im et al. [17] utilized an effective algorithm to address the path planning problem for cellular-connected AAVs with communication connectivity and battery constraints. Xie et al. [18] employed the Dueling Double Deep Q Network (D3QN) algorithm to optimize the three-dimensional (3D) trajectory of cellular-connected AAV, aiming to minimize the weighted sum of outage time and flight time. Chen and Huang [19] investigated the trajectory planning and BS association for cellular-connected AAVs. A key requirement of AAVs is to maintain a reliable cellular connection throughout their flight missions in order to support safety functions. Zhang et al. [20] optimized 3D path of cellular-connected AAVs to minimize their flight distance from origin to destination, while concurrently ensuring target link quality based on the expected Signal-to-Noise Ratio (SINR) for each AAV. Cherif et al. [21] proposed a novel joint cargo-AAV trajectory planning and cell association method to efficiently utilize available energy while ensuring a reliable communication link for command and control purposes.

2) *Performance Optimization of Limited Energy AAV*: Zhang et al. [23] proposed a trajectory optimization scheme for AAV-assisted Internet of Things (IoT) networks that is energy-efficient. Under the constraints of the average data rate and total energy consumption, the trajectory of the AAV is optimized. Cao et al. [24] proposed a multi-user-oriented AAV cargo distribution system. Considering the total energy consumption factor, the total outage time is reduced by jointly optimizing communication scheduling, allocation scheduling, and AAV's flight trajectory. Hong et al. [25] studied the access planning problem of AAV-assisted wireless-power sensor networks (WPSNs) under the premise of limited AAV energy, and introduced a hybrid data collection strategy to reduce the age information (AoI) of the data. Based on the further optimization of deep reinforcement learning (DRL), Yu et al. [26] proposed a multi-objective deep deterministic policy gradient (DDPG) algorithm. Under the condition of limited energy of AAV, the data rate and total harvest energy are maximized. Ye et al. [27] studied an energy-constrained AAV-supported Wireless Charging IoT network. An energy constrained flight and hovering energy harvesting (EFH-EH) scheme is proposed, and the optimal time allocation and hovering time ratio are deeply studied to maximize the total throughput. Do et al. [28] investigated a wireless communication system consisting of multiple AAVs, in which energy consumption is particularly considered. By applying deep reinforcement learning methods, the service time and throughput of the drone are maximized.

B. MOTIVATION AND CONTRIBUTIONS

Although there have been many studies on the communication quality optimization of AAV and the performance optimization of AAV with limited energy, as far as we know, there are no studies on the communication quality optimization of inspection AAV with limited energy. For example, although the author studied the inspection system of AAV and minimized the total flight distance of AAV by planning the inspection sequence of AAV in literature [29], this paper did not consider the communication quality problem during AAV flight. Although the literature [30] optimized the communication quality by optimizing the flight trajectory of the inspection AAV, it did not take into account the reality that the AAV energy is limited. In addition, although literature [31] considered the communication quality optimization problem of AAV and the battery energy of AAV is limited, the author studied only one target point in the system, which cannot be directly extended to the multi-target point inspection system of AAV. In view of the shortcomings of the existing work, we studied the communication quality optimization problem of the inspection AAV with limited energy in the multi-target point scenario. The main contributions of this paper are summarized as follows:

- We study a AAV inspection system with multiple inspection points, and consider the realistic factor that the battery energy of the AAV is limited. To optimize the communication quality of the inspection AAV with limited energy, we propose a hybrid algorithm combining SA algorithm and D3QN algorithm (SA-D3QN), in which SA algorithm is used to optimize the inspection sequence of the AAV, and D3QN algorithm is used to optimize the 2D trajectory of the AAV.
- To overcome the problem of AAV communication quality optimization under unknown channel environment, we approximate the outage probability in real

environment based on the channel data sampled by AAV, which could be achieved by leveraging the existing soft handover mechanisms with frequent reference signal received power (RSRP) and reference signal received quality (RSRQ) reports. Moreover, to demonstrate that training with AAV sampled data can solve the situation that the channel environment or radio map is not known in advance, we show the radio map trained with AAV sampled data.

- To demonstrate the superiority of the state-of-the-art DRL algorithm when the action space is discrete, i.e., D3QN, we use the DDQN algorithm as a comparison. Numerical results show that regardless of the on-board energy of the AAV, the total outage time of the AAV optimized by the D3QN algorithm is reduced by 15%-25% compared with that optimized by the DDQN algorithm.

The rest of this paper is organized as follows. The system model is introduced in Section II. The optimization problem is proposed and the radio map is constructed in Section III. In Section IV, we introduce our proposed algorithm. In Section V, the simulation parameter and results are presented. Finally, we summarize the conclusion in Section VI.

Notations: In this paper, scalars are denoted by italic letters, vectors are represented by bold-face lowercase letters, and matrices are denoted by bold-face uppercase letters. Moreover, let $\mathbb{R}^{N\times 1}$ represent the space of N-dimensional real-valued vector. For a vector $\mathbf{a}$, $\|\mathbf{a}\|$ represents its Euclidean norm. For a time-dependent function $y(t)$, $\dot{y}(t)$ represents the first-order derivative with respect to time t. For any real number m, $\lfloor m \rfloor$ means rounding m to the nearest integer no more than it.

II. SYSTEM MODEL

As shown in Figure 1, we consider a AAV inspection system, in which M inspection points are randomly distributed, and the AAV needs to fly to these inspection points for patrol inspection. The energy of the inspection AAV is limited, and the battery energy storage is denoted by $E_{\max}$. The M inspection points are denoted as $\mathcal{M} = \{1, 2, \ldots, M\}$, $M \geq 1$. The position of M inspection points is represented by the set $\mathcal{U} = \{U_1, U_2, \ldots, U_M\}$, $M \geq 1$, where for any $U_m \in \mathcal{U}$, it represents the two-dimensional coordinates (x_m, y_m) of the m-th inspection point. In addition, the position of the flight starting point of the inspection AAV is represented by $U_0 = (x_0, y_0)$.

The time consumed by the AAV to complete all the inspection tasks is denoted by T_{total}. The position of the inspection AAV is represented by $\mathbf{S}(t) = (x(t), y(t), H_i), \forall t \in [0, T_{\text{total}}]$, where $x(t)$, $y(t)$, and H_i represent the horizontal coordinate, vertical coordinate, and flight height of the AAV at time t, respectively. In particular, to reduce the complexity of the system, we assume that the flying altitude H_i of the AAV is constant and greater than the height of all buildings in the flying area. In addition, the velocity and weight of the inspection AAV are denoted by $\mathbf{V}_i$ and W_i, respectively. In particular, both $\mathbf{V}_i$ and W_i are constant in this paper.

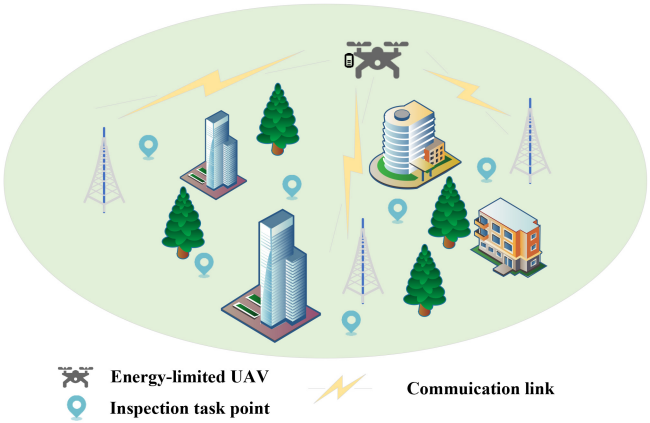


FIGURE 1. Energy-limited AAV inspection system.

A. ENERGY CONSUMPTION MODEL

In the AAV inspection system considered in this paper, we consider the inspection AAV is a rotary-wing AAV. In the AAV inspection system, the AAV battery energy storage is mainly used in two parts: 1) the energy consumed to maintain the communication between the AAV and the BS; 2) the energy consumed to maintain the AAV's flight. According to the literature [32], the communication energy consumption of AAV is very small compared with the flight energy consumption. Consequently, in this paper, we only consider the flight energy consumption of the inspection AAV. The flight energy consumption model of rotary-wing AAV can be expressed as [32]:

$$P(\mathbf{V}_i) = \underbrace{P_0\left(1+\frac{3\mathbf{V}_i^2}{U_{\text{tip}}^2}\right)}_{\text{blade profile}} + \underbrace{\frac{1}{2}d_0\rho s A\mathbf{V}_i^3}_{\text{parasite}} + \underbrace{P_i\left(\sqrt{1+\frac{\mathbf{V}_i^4}{4v_0^4}}-\frac{\mathbf{V}_i^2}{2v_0^2}\right)^{1/2}}_{\text{induced}}, \tag{1}$$

where P_0 and P_i denote the AAV blade profile and induced poer in hover condition. U_{tip} denotes the tip speed of the rotor blade, ρ denotes the air density, A denotes rotor disc area, d_0 denotes fuselage drag ratio, and s denotes rotor solidity. Note that in (1), the two parameters $P_i = (1+k)\frac{W_i^{3/2}}{\sqrt{2\rho A}}$ and $v_0 = \sqrt{\frac{W_i}{2\rho A}}$ are related to the weight W_i of the AAV, where k denotes incremental correction factor to induced power. In addition, we set the AAV at time $t = 0$ at the beginning of the inspection mission, then the energy consumed by the inspection AAV at time t can be calculated as follows:

$$E_{\text{con}}(t) = \int_0^t P(\mathbf{V}_i)dt, \forall t \in [0, T_{\text{total}}]. \tag{2}$$

Therefore, the remaining battery energy $E_{\text{remain}}(t)$ of the inspection AAV at time t can be expressed as:

$$\begin{aligned} E_{\text{remain}}(t) &= E_{\max} - E_{\text{con}}(t) \\ &= E_{\max} - \int_0^t P(\mathbf{V}_i)dt, \forall t \in [0, T_{\text{total}}]. \end{aligned} \tag{3}$$

Note that due to the limited energy of the AAV, the remaining energy of the AAV at any time needs to meet the following constraints:

$$0 \leq E_{\text{remain}}(t) \leq E_{\max}, \forall t \in [0, T_{\text{total}}]. \tag{4}$$

B. COMMUNICATION MODEL

In the urban environment constructed in this paper, we consider an air-to-ground (A2G) channel model. In this model, we adopt the log-distance path loss model and combine the 3rd Generation Partnership Project (3GPP) specification to simulate the communication link between AAV and BSs [33], [34], [35]. In this work, there are a total of N BSs in the flight area of the AAV. In order to be closer to the actual deployment, we use standard sector technology, where each BS includes three different cells. Assuming that there are a total of N_b cellular cells in the region, then $N_b = 3 \times N$. At time t, $h_n(t)$ is expressed as the channel gain from cell n-th to AAV, i.e.,

$$h_n(t) = \sqrt{\mathrm{G}_n(t)\beta_n(t)}\tilde{h}_n(t), \tag{5}$$

where $G_n(t)$ represents the antenna gain, $\beta_n(t)$ and $\tilde{h}_n(t)$ represent large-scale fading and small-scale fading, respectively. The small-scale fading $\tilde{h}_n(t)$ is a random variable. The large-scale fading $\beta_n(t)$ is influenced by the position of the buildings between the AAV and BSs. It is worth noting that there are Line-of-Sight (LoS) links and non Line-of-Sight (NLoS) links. The two path losses between AAV and cell n are expressed as [18]:

$$\begin{aligned} \beta_n^{LoS}(t) = \max\Big\{ & \beta_n^{FSPL}, 28 \\ & + \left(22 + 20\log_{10} h_t\right)\log_{10} d_n(t) + 20\log_{10} f_c \Big\}, \end{aligned} \tag{6}$$

$$\begin{aligned} \beta_n^{NLoS}(t) = \max\Big\{ & \beta_n^{LoS}(t), -17.5 \\ & + \left(46 - 7\log_{10} h_t\right)\log_{10} d_n(t) + 20\log_{10} f_c \Big\}, \end{aligned} \tag{7}$$

where β_n^{FSPL} represents the free-space path loss, h_t is the AAV's height at time t, $d_n(t)$ is the 3D distance between AAV and cell n, and f_c is the carrier frequency. Due to the presence of static and substantial obstacles blocking the signal, it can be assumed that this fading remains time-invariant. We can get

$$\beta_n(t) = \begin{cases} \beta_n^{LoS}(t), & \text{if LoS link,} \\ \beta_n^{NLoS}(t), & \text{if NLoS link.} \end{cases} \tag{8}$$

Consequently, the signal power that AAV receives from n-th cell can be expressed as:

$$p_n(\mathbf{S}(t)) = \bar{P}_n(h_n(t))^2. \tag{9}$$

where $\bar{P}_n$ represents the transmit power of the n-th cell, which is assumed to remain constant over time.

For the communication scheme established between AAV and the BSs, some reasons are considered. The short time slot interval of time division multiple access (TDMA) can makes low response times, which facilitates high-quality communication. Additionally, adopting this mature TDMA technology can reduce the risks associated with the design and implementation of AAV communication systems, while also contributing to lower overall costs [36]. Therefore, TDMA is employed as the access method. We introduce communication scheduling $a(t)$ to designate the BS that contacts with the AAV at time t. This variable $a(t)$ satisfies the following condition:

$$a(t) \in \{1, 2, \ldots, N_b\}. \tag{10}$$

Then, in addition to the BS currently communicating with the AAV, signals sent by other BSs are considered as the noise received by the AAV at time t, which are denoted as $N(\mathbf{S}(t))$, i.e., $N(\mathbf{S}(t)) = \sum_{m \notin a(t)} p_m(\mathbf{S}(t))$. The signal-to-interference ratio (SIR) received by the inspection AAV can be written as:

$$\mathrm{SIR}(t) \triangleq \frac{p_{a(t)}(\mathbf{S}(t))}{N(\mathbf{S}(t))}. \tag{11}$$

Furthermore, we define a signal quality threshold, denoted as γ_{th}. When the SIR of the AAV at time t is lower than this threshold, which indicates that the AAV has been disconnected from the BS. The mathematical expression corresponding to this condition is:

$$P_{\text{out}}(\mathbf{S}(t), a(t)) = \Pr\{\mathrm{SIR}(t) < \gamma_{th}\}, \tag{12}$$

where $\Pr\{\cdot\}$ represents the mathematical expectation, which quantifies the probability of a particular event occurring. At time t, the larger $P_{\text{out}}(\mathbf{S}(t), a(t))$ is, the more likely there will be a communication outage at the AAV's position $\mathbf{S}(t)$, i.e., the worse the communication quality of the AAV at this position [37]. We use the total communication outage time to evaluate the communication quality of the AAV, and the total expected outage time T_{out} can be calculated as follow:

$$T_{\text{out}}(\mathbf{S}(t), a(t)) = \int_0^{T_{\text{total}}} P_{\text{out}}(\mathbf{S}(t), a(t))dt. \tag{13}$$

C. AAV INSPECTION MODEL

We use $\Pi = \{\pi_0, \pi_1, \ldots, \pi_M\}$ to represent the inspection sequence of the AAV, where $\pi_m \in \Pi$ indicates that the destination of the AAV's m-th inspection is the π_m-th inspection point [38]. The inspection sequence Π meets the following conditions:

$$\pi_0 = 0, \tag{14}$$

$$\pi_m = k, \forall k \in \mathcal{M}, \exists m \in \{1, 2, \ldots, M\}, \tag{15}$$

$$\pi_i \neq \pi_j, i \neq j, \forall i, j \in \{1, 2, \ldots, M\}. \tag{16}$$

We assume that the inspection operation is completed instantly after the AAV arrives at the inspection point.

Meanwhile, we consider that when the AAV is located directly above the inspection point, i.e., when the projection position $\mathbf{s}(t) = (x(t), y(t))$ of the AAV coincides with the location (x_m, y_m) of the m-th inspection point, the AAV has performed the patrol inspection task at the m-th inspection point. We assume that the AAV arrives at the π_m-th inspection point at time t_{π_m}. Then the trajectory of the AAV should meet the following conditions:

$$\mathbf{s}(t_{\pi_m}) = (x_m, y_m), \exists t_{\pi_m} \in [0, T_{\text{total}}], \forall m \in \mathcal{M}. \quad (17)$$

III. PROBLEM FORMULATION

In the AAV inspection scenario studied in this paper, the AAV must complete all inspection tasks under the premise of limited energy and maintain good communication quality as much as possible. Our goal is to minimize the total communication outage time of AAV by optimizing the inspection sequence Π, communication schedule $a(t)$, and flight trajectory $\mathbf{S}(t)$ of AAV. The optimization problem can be formulated as follows:

$$\begin{aligned}
&(\text{P0}): \min_{\mathbf{S}(t), a(t), \Pi} T_{\text{out}} && (18)\\
\text{s.t.}\quad &\|\mathbf{S}(t)\| = \mathbf{V}_i, && (18a)\\
&\mathbf{s}(0) = (x_0, y_0), && (18b)\\
&0 \le x(t) \le L, \forall t \in [0, T_{\text{total}}], && (18c)\\
&0 \le y(t) \le L, \forall t \in [0, T_{\text{total}}], && (18d)\\
&(4), (10), (14)-(17).
\end{aligned}$$

In problem (P0), (18) is the objective function, which is set to minimize the total communication outage time of the AAV. In this paper, we achieve the purpose of improving the communication quality by minimizing the communication outage time of the AAV. Eq. (18a) is the AAV dynamics constraint. It indicates that the trajectory and velocity of the AAV must satisfy a strict mathematical relationship. Eq. (18b) indicates that the trajectory projection of the AAV at the beginning of the task must be located at the starting point U_0. Eq. (18c) and (18e) indicate that the AAV cannot exceed the specified area when flying, where the size of the specified area is $L \times L$. In addition, (4), (10), (14) - (16) represent the energy constraint, the communication scheduling constraint, and the inspection constraint of the AAV, respectively.

IV. PROPOSED ALGORITHM

Obviously, the problem (P0) is a mixed-integer nonlinear programming problem. It is difficult for traditional optimization methods (e.g., convex optimization) to solve this problem. At the same time, directly using DRL to address this problem will also lead to the problem of too high decision dimension. To solve the problem (P0), we propose a hybrid algorithm in this paper, which consists of simulated annealing (SA) and D3QN. Specifically, SA is used to optimize the inspection sequence of the AAV. The D3QN algorithm is used to optimize the trajectory, subject to the battery energy of the AAV and the unknown channel environment.

A. OPTIMIZATION OF AAV INSPECTION SEQUENCE

When planning the inspection sequence of the AAV, we assume that the AAV flies in a straight line. Since the AAV's energy is limited, we expect that the AAV can complete the inspection task with as little energy as possible when flying in a straight line, so that the AAV has more energy to adjust its flight trajectory to improve the communication quality to a greater extent. The optimization problem can be formulated as follows:

$$\begin{aligned}
&(\text{P1}): \min_{\Pi} E_{\text{con}}(T_{\text{total}})\\
&\text{s.t.}\quad (18a)-(18e), (4), (14)-(17), && (19)
\end{aligned}$$

where $E_{\text{con}}(T_{\text{total}})$ in (19) represents the total energy consumption of the AAV. We use $E_{\text{con}}^{\pi_i,\pi_j}$ to denote the energy consumed by the AAV to fly from π_i-th inspection point to π_j-th inspection point, then the problem (P2) can be reformulated as:

$$\begin{aligned}
&(\text{P2}): \min_{\Pi} \sum_{k=0}^{M-1} E_{\text{con}}^{\pi_k,\pi_{k+1}}\\
&\text{s.t.}\quad (18a)-(18e), (4), (14)-(17). && (20)
\end{aligned}$$

To reduce the complexity of the system, we have assumed that the AAV flies straight line when optimizing the inspection sequence Π, then $E_{\text{con}}^{\pi_k,\pi_{k+1}}$ can be calculated by the following formula:

$$\begin{aligned}
E_{\text{con}}^{\pi_k,\pi_{k+1}} &= \int_{t_{\pi_k}}^{t_{\pi_{k+1}}} P(\mathbf{V}_i) dt\\
&= P(\mathbf{V}_i) \times (t_{\pi_{k+1}} - t_{\pi_k})\\
&= P(\mathbf{V}_i) \times \frac{\|U_{\pi_{k+1}} - U_{\pi_k}\|}{\mathbf{V}_i}, && (21)
\end{aligned}$$

where $\|U_{\pi_{k+1}} - U_{\pi_k}\|$ indicates the Euclidean distance between π_{k+1}-th inspection point and π_k-th inspection point, i.e.,

$$\|U_{\pi_{k+1}} - U_{\pi_k}\| = \sqrt{(x_{\pi_{k+1}} - x_{\pi_k})^2 + (y_{\pi_{k+1}} - y_{\pi_k})^2}. \quad (22)$$

In this paper, we assume that the speed of the AAV is a constant, then problem (P2) can be simplified as follows:

$$\begin{aligned}
&(\text{P3}): \min_{\Pi} \sum_{k=0}^{M-1} \|U_{\pi_{k+1}} - U_{\pi_k}\|\\
&\text{s.t.}\quad (18a)-(18e), (4), (14)-(17). && (23)
\end{aligned}$$

It can be seen that problem (P3) is a travel salesman problem (TSP). In this paper, we use SA algorithm to solve the problem (P3), and refer to Algorithm 1 for the specific process.

B. OPTIMIZATION OF AAV TRAJECTORY AND COMMUNICATION SCHEDULING

We further optimize the flight trajectory and communication scheduling of AAV to improve the communication quality of AAV. The optimization problem can be expressed as:

Algorithm 1 Optimization of AAV Inspection Sequence Based on SA Algorithm

Initialize: initial temperature T_0, termination temperature T_t, cooling coefficient c, random initialize inspection sequence Π^*. Set the maximum number of iterations to N_{SA}. Set the objective function (23) to $f()$. Set current task iteration number $k \leftarrow 0$.
while $k < N_{\mathrm{SA}}$ and $T > T_t$ **do**
  Randomly select a neighborhood solution Π_k at the k-th iteration, calculate the increment of the objective function (23) as $\Delta_f = f(\Pi_k) - f(\Pi^*)$;
  if $\Delta_f < 0$ **then**
    $\Pi^* = \Pi_k$
  else
    Generate random numbers $\xi = U(0, 1)$
    if $\xi < \exp(-\Delta_f / T)$ **then**
      $\Pi^* = \Pi_k$
    end if
  end if
  Reduce temperature $T = T \times c$
  $k = k + 1$
end while

$$\text{(P4)}: \min_{\mathbf{S}(t), a(t)} \quad T_{\mathrm{out}} \qquad \text{s.t.} \quad \text{(18a)} - \text{(18e)}, \text{(4)}, \text{(17)}. \tag{24}$$

According to (13), it is very difficult to derive the mathematical expression T_{out} of the total communication interruption time of the AAV. To solve problem (P4), we first optimize the communication scheduling of the AAV, and then further optimize the trajectory of the AAV based on this. The optimal communication schedule for the AAV is defined as $a^*(t)$, which is defined as follows:

$$\mathrm{SIR}(a^*(t)) \geq \mathrm{SIR}(a(t)), \forall a(t) \in \{1, 2, \ldots, N_b\}, \tag{25}$$

$$a^*(t) \in \{1, 2, \ldots, N_b\}. \tag{26}$$

After the establishment of (25) and (26), there will be derived as follows:

$$\Pr\{\mathrm{SIR}(a^*(t)) < \gamma_{th}\} < \Pr\{\mathrm{SIR}(a(t)) < \gamma_{th}\}. \tag{27}$$

According to (12) and (13), we can derive the following formula:

$$T_{\mathrm{out}}(\{\mathbf{S}(t), a^*(t)\}) < T_{\mathrm{out}}(\{\mathbf{S}(t), a(t)\}). \tag{28}$$

Therefore, after optimizing the communication scheduling of the AAV, the outage probability of the AAV is only related to the position of the AAV at the current time, i.e.,

$$P_{\mathrm{out}}(\mathbf{S}(t)) = P_{\mathrm{out}}(\mathbf{S}(t), a^*(t)). \tag{29}$$

$$\mathrm{SIR}(t) \triangleq \frac{p_{a^*(t)}(\mathbf{S}(t))}{N(\mathbf{S}(t))} \tag{30}$$

$$T_{\mathrm{out}}(\mathbf{S}(t)) = \int_0^{T_{\mathrm{total}}} P_{\mathrm{out}}(\mathbf{S}(t))dt = \int_0^{T_{\mathrm{total}}} P_{\mathrm{out}}(\mathbf{S}(t), a^*(t))dt. \tag{31}$$

After optimizing the communication scheduling of AAV, we need to optimize the trajectory of AAV. The optimization problem can be expressed as:

$$\text{(P5)}: \min_{\mathbf{S}(t)} \sum_{k=0}^{M-1} T_{\mathrm{out}}^{\pi_k, \pi_{k+1}} \qquad \text{s.t.} \quad \text{(18a)} - \text{(18e)}, \text{(4)}, \text{(17)}, \tag{32}$$

where $T_{\mathrm{out}}^{\pi_k, \pi_{k+1}}$ represents the total communication outage time of the AAV during the journey from π_k-th inspection point to π_{k+1}-th inspection point, and $T_{\mathrm{out}} = \sum_{k=0}^{M-1} T_{\mathrm{out}}^{\pi_k, \pi_{k+1}}$. $T_{\mathrm{out}}^{\pi_k, \pi_{k+1}}$ is only related to the trajectory $\mathbf{S}_{\pi_k}^{\pi_{k+1}}(t)$ of the AAV between π_k-th inspection point and π_{k+1}-th inspection point, and is not affected by the trajectory between any other two inspection points. Therefore, we can optimize the sub-trajectory of the AAV in turn to optimize the whole trajectory. The trajectory optimization problem between π_k-th inspection point and π_{k+1}-th inspection point can be expressed as follows:

$$\text{(P6)}: \min_{\mathbf{S}_{\pi_k}^{\pi_{k+1}}(t)} T_{\mathrm{out}}^{\pi_k, \pi_{k+1}} \tag{33}$$

$$\text{s.t.} \quad \|\dot{\mathbf{S}}_{\pi_k}^{\pi_{k+1}}(t)\| = \mathbf{V}_i, \tag{33a}$$

$$\mathbf{S}(t_{\pi_k}) = (x_{\pi_k}, y_{\pi_k}, H_i), \tag{33b}$$

$$\mathbf{S}(t_{\pi_{k+1}}) = (x_{\pi_{k+1}}, y_{\pi_{k+1}}, H_i), \tag{33c}$$

$$0 \leq x(t) \leq L, \forall t \in [t_{\pi_k}, t_{\pi_{k+1}}], \tag{33d}$$

$$0 \leq y(t) \leq L, \forall t \in [t_{\pi_k}, t_{\pi_{k+1}}],$$

$$E_{\mathrm{con}}^{\pi_k, \pi_{k+1}} \leq E_{\mathrm{max}}^{\pi_k, \pi_{k+1}}, \tag{33e}$$

where $E_{\mathrm{max}}^{\pi_k, \pi_{k+1}}$ represents the maximum energy that the AAV is allowed to consume from the π_k-th inspection point to the π_{k+1}-th inspection point. It is defined as follow:

$$E_{\mathrm{max}}^{\pi_k, \pi_{k+1}} = \frac{\|U_{\pi_{k+1}} - U_{\pi_k}\|}{\sum_{m=0}^{M-1} \|U_{\pi_{m+1}} - U_{\pi_m}\|} \times E_{\mathrm{max}}. \tag{34}$$

In problem (P6), the objective function is to minimize the total communication outage time $T_{\mathrm{out}}^{\pi_k, \pi_{k+1}}$ of the AAV, and the optimization variable is the flight trajectory $\mathbf{S}_{\pi_k}^{\pi_{k+1}}(t)$ of the AAV. It can be seen from (13) that $T_{\mathrm{out}}^{\pi_k, \pi_{k+1}}$ is a function of $\mathbf{S}_{\pi_k}^{\pi_{k+1}}(t)$, and the mathematical expression of the relationship between $T_{\mathrm{out}}^{\pi_k, \pi_{k+1}}$ and $\mathbf{S}_{\pi_k}^{\pi_{k+1}}(t)$ is difficult to derive, so the optimization problem is highly non-convex. In this paper, we use D3QN algorithm to solve the non-convex problem. The architecture of D3QN algorithm is shown in Figure 2.

The first step to apply D3QN algorithms for solving a problem is to reformulate it as an Markov decision process (MDP) [39]. Since MDP is defined on discrete time steps, for problem (P6), our first task is to discretize the time interval $[0, T_{\mathrm{total}}]$ into N_{total} time steps with fixed intervals Δ_t, where $T_{\mathrm{total}} = N_{\mathrm{total}} \Delta_t$. Note that Δ_t is the length of the discrete time slot, and its value is so small that the outage probability of the AAV within that time slot can be considered to be approximately constant. Then, the AAV's trajectory $\mathbf{S}_{\pi_k}^{\pi_{k+1}}(t)$

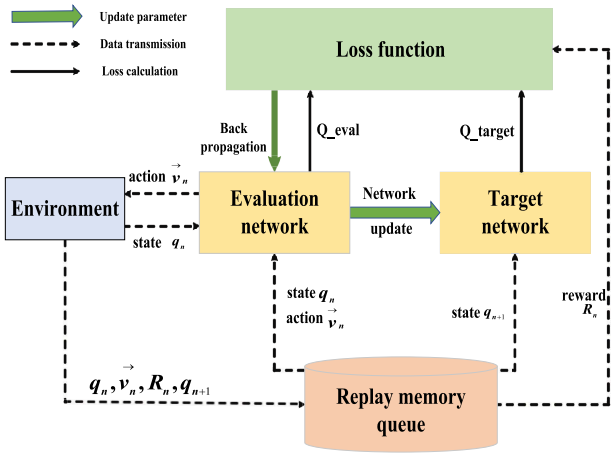


FIGURE 2. D3QN architecture.

can be reformulated by the sequence $\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}_{n=1}^{N_k}$, and $\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}_{n=1}^{N_k}$ meets the following conditions [16]:

$$\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n+1] = \mathbf{S}_{\pi_k}^{\pi_{k+1}}[n] + \Delta_t \mathbf{V}_i \vec{\mathbf{v}}_{\pi_k,\pi_{k+1}}^{n}, \forall k, n, \quad (35)$$

$$\|\vec{\mathbf{v}}_{\pi_k,\pi_{k+1}}^{n}\| = 1, \forall k, n, \quad (36)$$

$$N_{\text{total}} = \sum_{k=0}^{M-1} N_k, \quad (37)$$

where $\vec{\mathbf{v}}_{\pi_k,\pi_{k+1}}^{n}$ represent the flight direction of the AAV in the n-th time slot, respectively. N_k represents the total number of time slots for the AAV to fly from the π_k-th inspection point to π_{k+1}-th inspection point. Then, the outage time $T_{\text{out}}^{\pi_k,\pi_{k+1}}$ generated by the AAV during the journey from π_k-th inspection point to π_{k+1}-th inspection point can be calculated as follows:

$$T_{\text{out}}^{\pi_k,\pi_{k+1}} = \Delta_t \sum_{n=1}^{N_k} P_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}). \quad (38)$$

Thus, problem (P6) can be reformulated as:

$$\text{(P7)}: \min_{\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}} \sum_{n=1}^{N_k} P_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}) \quad (39)$$

$$\text{s.t.} \quad \mathbf{S}_{\pi_k}^{\pi_{k+1}}[1] = (x_{\pi_k}, y_{\pi_k}, H_i), \quad (39a)$$

$$\mathbf{S}_{\pi_k}^{\pi_{k+1}}[N_k] = (x_{\pi_{k+1}}, y_{\pi_{k+1}}, H_i), \quad (39b)$$

$$0 \le x[n] \le L, \forall n, \quad (39c)$$

$$0 \le y[n] \le L, \forall n, \quad (39d)$$

$$N_k \le N_k^{\max}, \forall k, \quad (39e)$$

where $N_k^{\max}$ represents the maximum step that the AAV is allowed to fly from π_k-th inspection point to π_{k+1}-th inspection point. It can be calculated by the following equation:

$$N_k^{\max} = \lfloor \frac{E_{\max}^{\pi_k,\pi_{k+1}}}{P(\|\mathbf{V}_{\pi_k}^{\pi_{k+1}}\|, W_i)} \rfloor. \quad (40)$$

When the channel environment of the AAV flight area is unknown, the outage probability $P_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\})$ of the AAV at position $\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]$ is still difficult to calculate. According to literature [16], in order to solve this problem, we define an indicator function $I(\cdot)$, which is defined as follows:

$$I(\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]) = \begin{cases} 1, & \text{if SIR}[n] < \gamma_{\text{th}}, \\ 0, & \text{otherwise.} \end{cases} \quad (41)$$

Then we have

$$P_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}) = \Pr(\text{SIR}[n] < \gamma_{\text{th}}) = \mathbb{E}_{\tilde{h}}[I(\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n])] \quad (42)$$

Assume that within each time step n, the AAV performs J SIR measurements for each of the N_b cell associations, which could be achieved by leveraging the existing soft handover mechanisms with frequent RSRP and RSRQ reports. The corresponding outage indicator value $I(\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n])$ can be obtained based on (41). Then the empirical outage probability can be obtained as:

$$\hat{P}_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}) \triangleq \frac{1}{J} \sum_{j=1}^{J} I(\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]). \quad (43)$$

By applying the Law of Large Numbers to (44), we have $\lim_{J\to\infty} \hat{P}_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}) = P_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\})$. Therefore, as long as the AAV performs signal measurements sufficiently frequently so that $J \gg 1$, $P_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\})$ in (42) can be evaluated by its empirical value $\hat{P}_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\})$.

Then, (P7) can be reformulated as follows:

$$\text{(P8)}: \min_{\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}} \sum_{n=1}^{N_k} \hat{P}_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\}) \quad (44)$$

$$\text{s.t.} \quad (39a)-(39e).$$

Mathematically, an MDP is represented by $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R} \rangle$, where $\mathcal{S}$ and $\mathcal{A}$ denote the state space and action space, respectively. $\mathcal{P}$ indicates the state transition probability, $\mathcal{R}$ indicates reward [39]. Corresponding to the optimization problem in this paper, state, action, probability, and reward are defined as follows:

- *State* $\mathcal{S}$: The state space is composed of all its legitimate positions in the airspace, denoted as $s_n = (x[n], y[n], H_i)$.
- *Action* $\mathcal{A}$: In this paper, the AAV have eight flight directions, which are: front, back, left, right, left front, left back, right front, and right back.
- *Probability* $\mathcal{P}$: the state transition probability is deterministic governed by (35).
- *Reward* $\mathcal{R}$: The reward function is set to $r_n = -\hat{P}_{\text{out}}(\{\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n]\})$, which corresponds to the contribution to the objective function of (P8) at one time step. In this paper, we used multi-step learning algorithms to accelerate convergence speed, where N_m-step return $R_{n:n+N}$ is defined as $R_{n:n+N} = \sum_{k=0}^{N-1} \gamma^k r_{n+k+1}$.

As is shown in Figure 2, different from classical RL methods, e.g., Q-learning, D3QN algorithm introduces two neural networks, i.e., Q network and target network. And ζ and ζ' are parameters of Q network and target network respectively. The distinction between the Q network and

the target network lies in their update mechanisms. The Q network undergoes updates during each training iteration, whereas the target network involves a hard copy process, where the entire set of Q network parameters is periodically copied to the target network to facilitate its update. The DDQN algorithm addresses the constraints of table-based approaches by employing Deep Neural Networks (DNNs) to estimate the Q-value function, i.e., $Q_(s, a; \boldsymbol{\zeta}) \approx Q_(s, a)$, and

$$Q(s, a) = V(s) + A(s, a), \tag{45}$$

where $V(s)$ is state value, which is related to state and independent of action. $A(s, a)$ is the advantage function, which measures the advantage of each action over other actions. In policy gradient, this method can reduce the variance of error during learning and make learning more stable.

Then D3QN algorithm is to minimize the loss given by [39]:

$$\left(R_{n:n+N} + \gamma^N \max_{a_n^* \in \mathcal{A}} Q\big(s_{n+N}, a_n^*; \boldsymbol{\zeta}'\big) - Q(s_n, a_n; \boldsymbol{\zeta})\right)^2, \tag{46}$$

where

$$a_n^* = \arg\max_{a' \in \hat{\mathcal{A}}} Q\big(s_{n+N}, a'; \boldsymbol{\zeta}\big). \tag{47}$$

The specific algorithm process refers to Algorithm 2.

Algorithm 2 Trajectory Optimization of AAV From π_k-th Inspection Point to π_{k+1}-th Inspection Point

Initialize: maximum flight steps of the AAV $N_k^{\max}$, maximum training episodes N_{ep}, tolerable distance to reach flight endpoint D_t, initial exploration θ_0, decaying rate α, discount factor γ, reward of reach flight endpoint R_{end}, punishment for flying out of the boundary R_p, and set $\theta \leftarrow \theta_0$.
for $n = 1$ to N_{ep} **do**
  if n is a multiple of 10 **then**
    Initialize state of the AAV as $\mathbf{S}_{\pi_k}^{\pi_{k+1}}[1] = (x_{\pi_k}, y_{\pi_k}, H_i)$ and set the time step $n \leftarrow 1$.
  else
    Randomly initialize the state of the AAV $\mathbf{S}_{\pi_k}^{\pi_{k+1}}[1] \in \mathcal{S}$ and set the time step $n \leftarrow 1$.
  end if
  repeat
    Select an action a_n^* from the action space $\mathcal{A}$ based on θ-greedy policy, where there is a probability of θ to choose a random action, and there is a probability of $1-\theta$ to choose the action that maximizes the current Q value.
    The AAV perform flight actions a_n^* and obtain a new state $\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n+1]$. Obtain the empirical outage probability $P_{\text{out}}(\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n+1])$ and set the reward as $r_n = -P_{\text{out}}(\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n+1])$.
    Store transition $(\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n], a_n^*, r_n, \mathbf{S}_{\pi_k}^{\pi_{k+1}}[n+1])$ in the sliding window.
    Randomly sampled N-step transition from the experience replay buffer are used to train the neural network.
    Set

$$y_n = \begin{cases} R_{n:n+N} + R_{\text{end}}, \text{if} \|\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n+N] - (U_{\pi_{k+1}}, H_i)\| \le D_t, \\ R_{n:n+N} + R_p, \text{ if } \mathbf{S}_{\pi_k}^{\pi_{k+1}}[n+N] \notin \mathcal{S}, \\ R_{n:n+N} + \gamma^N Q\big(s_{n+N}, a'; \boldsymbol{\zeta}\big), \text{otherwise.} \end{cases}$$

    The parameters of neural network $\boldsymbol{\zeta}$ are updated by gradient descent method and $n \leftarrow n+1$, $\theta \leftarrow \theta\alpha$.
  until $\|\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n] - (U_{\pi_{k+1}}, H_i)\| \le D_t$ or $\mathbf{S}_{\pi_k}^{\pi_{k+1}}[n] \notin \mathcal{S}$ or n $\ge N_k^{\max}$
  Update the target network coefficient $\zeta^- \leftarrow \zeta$ every B episodes
end for

V. SIMULATION RESULT AND PARAMETERS

In this section, we show simulation results to demonstrate the effectiveness of our proposed algorithm, and present the values of the specific simulation parameters.

A. PARAMETERS OF SIMULATION

In the AAV inspection system, there exists 6 inspection points and 7 BSs, which are randomly distributed in a rectangular area with a size of 2 km × 2 km, i.e., $M = 6$, $N = 7$, and $L = 2$. The maximum height of the building is 80 meters. The proportion of building area to inspection area is 0.28, and the number of buildings per square kilometer is 250. The flight altitude of the AAV is 100 meters, i.e., $H_i = 100$. The height of BSs is 25 meters, and the inclination angle of the antenna is 12.5°.

Note that we use D3QN algorithm in Algorithm 2, so there are two feedforward fully connected DNNs, and the structure of these two DNNs is the same. Specifically, the numbers of neurons of the first 4 hidden layers are 512, 256, 128, and 128, respectively. The last hidden layer is the dueling layer with $K+1$ neurons, with one neuron corresponding to the estimation of the state-value and the other K neurons corresponding to the action advantages for the K actions, i.e., the difference between the action values and the state value of each state. The outputs of these $K+1$ neurons are then aggregated in the output layer to obtain the estimation of the K action values. In addition, the DNN for radio mapping also contains 5 hidden layers, and the number of neurons in each layer is 512, 256, 128, 64, and 32. The DNN is trained using the Adam optimizer and the Python programming language, combined with TensorFlow and Keras frameworks, to achieve the goal of minimizing mean square error (MSE) losses. Other parameters used in the simulation are summarized in Table 1.

B. SIMULATION RESULT

Figure 3 shows the radio map when the height is 100 meters in the AAV inspection area. It can be observed from Figure 3 that there are many areas in the air with poor signal coverage, especially the area between two BSs, whose signal coverage is almost 0. This indicates that optimizing the AAV's trajectory is necessary to improve the communication quality of the AAV.

To prove the effectiveness of the method proposed in this paper to learn a radio map by using the information collected by the AAV, we show the radio map obtained by learning in Figure 4. It can be seen that the radio map obtained by training and learning of DNN is almost exactly the same as the radio map obtained by channel model (5)-(13), with only minor differences. This shows that the method of learning

TABLE 1. Partial simulation parameters.

Simulation parameter	Value
The weight of the UAV W_i	50N
The size of the inspection area L	2
cooling coefficient c	0.98
Initial temperature T_0	100
Termination temperature T_t	0
Maximum number of iterations of SA N_{SA}	300
Maximum number of training N_{ep}	8000
Initial exploration factor θ_0	0.6
Exploration factor decaying rate α	0.998
Time step length Δ_t	0.5 s
Step coefficient of multi-step learning N	30
Tolerable distance to reach flight endpoint D_t	10
Reward of reach flight endpoint R_{end}	12
Punishment for flying out of the boundary R_p	-100
Neural coefficient updating frequency of target network B	6

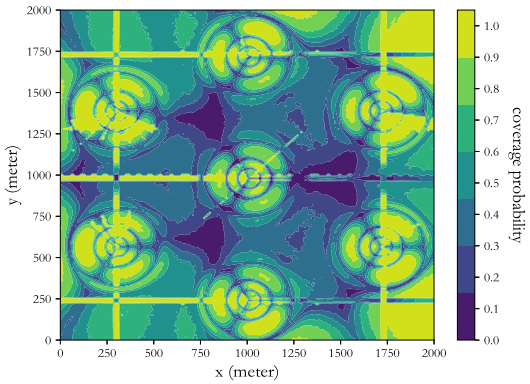


FIGURE 3. Radio map.

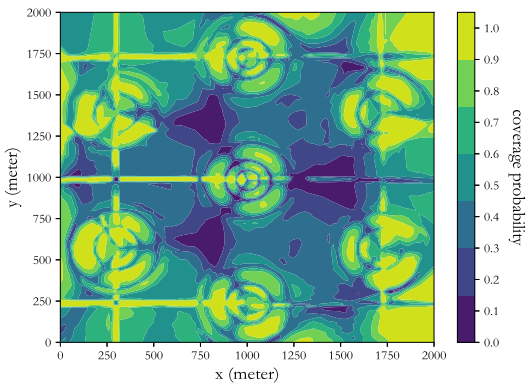
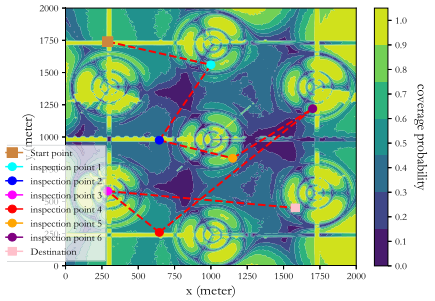


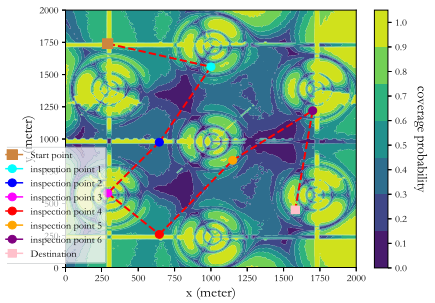
FIGURE 4. Radio map obtained by learning.

a radio map of an unknown environment through DNN is very efficient and reliable.

Similar to [40] and [41], to demonstrate the superiority of the proposed algorithm, we used a scheme as a comparison, the specific details of which are as follows:



(a) Based on greedy algorithm



(b) Based on SA algorithm

FIGURE 5. AAV inspection sequence based on different algorithms.

- greedy-D3QN: The inspection sequence optimization of AAV is greedy algorithm, and the trajectory optimization is D3QN algorithm.
- SA-DDQN: The inspection sequence optimization of AAV is SA algorithm, and the trajectory optimization is DDQN algorithm.

To prove the necessity of optimizing the AAV's inspection sequence, we first show the inspection sequence obtained based on different algorithms and its corresponding performance parameters, as shown in Figure 4 and Table 2.

From Figure 4 and Table 2, it can be seen that the flying distance of the AAV obtained by the SA algorithm is smaller than that obtained by the greedy algorithm, and the energy consumption of the AAV is positively correlated with the flying distance, so the energy consumption obtained by the SA algorithm is smaller than that obtained by the greedy algorithm. Therefore, the AAV has more residual energy that can be used to adjust its flight trajectory based on SA algorithm, resulting in higher communication quality. In addition, it can be seen that when the trajectory is not optimized, i.e., when the AAV flies in a straight line, the outage time and outage probability of the AAV during the entire task execution are large, which indicates that the communication quality of the AAV is poor when the trajectory is not optimized, and proves the importance of optimizing the flight trajectory of the AAV to improve the communication quality of the AAV.

In addition, we show the energy consumed by the AAV and the remaining battery energy based on different algorithms in Figure 6. As can be seen from Figure 6,

TABLE 2. Comparison of different algorithms.

Algorithm	Delivery sequence	Total flight distance (meter)	Outage time (second)	Outage ratio
Greedy	Start point → 1 → 2 → 5 → 6 → 4 → 3 → Destination	5813.03	143.3	59.16%
SA	Start point → 1 → 2 → 3 → 4 →5 → 6 → Destination	4641.70	105.3	54.45%

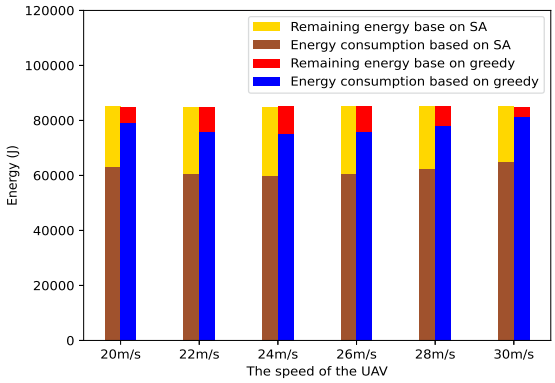


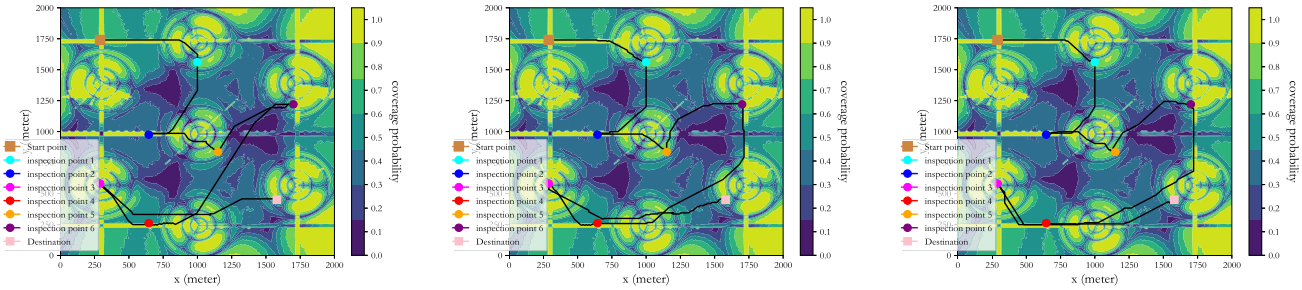
FIGURE 6. Energy consumption and remaining battery energy based on different algorithms when E_{max}=85Kj.

regardless of the speed of the AAV, the energy consumed by the SA-D3QN algorithm is always lower than that consumed by the greedy-D3QN algorithm, and when the battery energy of the AAV is constant, the remaining battery energy based on the SA-D3QN algorithm is always higher than that based on the greedy-D3QN algorithm. The numerical results show that when the battery energy of the AAV is limited, the SA-D3QN algorithm can save 20% energy consumption compared with the greedy-D3QN algorithm, which makes the AAV have more residual energy to optimize its trajectory and improve the communication quality.

After optimizing the inspection sequence of the AAV, we next show the flight trajectories under different AAV battery energy storage $E_{\max}$, as shown in Figure 7 and Figure 8. As can be seen from Figure 7, when the battery energy of the AAV $E_{\max}$=85Kj, a large part of the trajectory obtained based on the greedy-D3QN algorithm still passes through the area with poor signal coverage. When the battery energy of the AAV increases to 90Kj, the communication quality of the AAV is improved compared with that when the battery energy $E_{\max}$=85Kj, but there is still a part of the trajectory through the area with poor signal coverage. When the battery energy of the AAV is increased to 95Kj, it can be observed that the communication quality of the AAV has been greatly improved, but it can also be observed that the trajectory of the AAV is more circuitous than that of the battery energy of 85Kj or 90Kj, and the energy consumption is obviously higher at this time. Different from the greedy-D3QN algorithm, the flight trajectory of the AAV does not change significantly with the increase of battery energy $E_{\max}$ when based on SA-D3QN. This is because it can be observed from Figure 8(a) that when the battery energy $E_{\max}$=85Kj of the AAV, the trajectory of the AAV will hardly pass through the area where the cellular signal coverage rate is less than 50%. At this time, the AAV has obtained a good communication quality and does not need to further consume energy to optimize its flight trajectory.

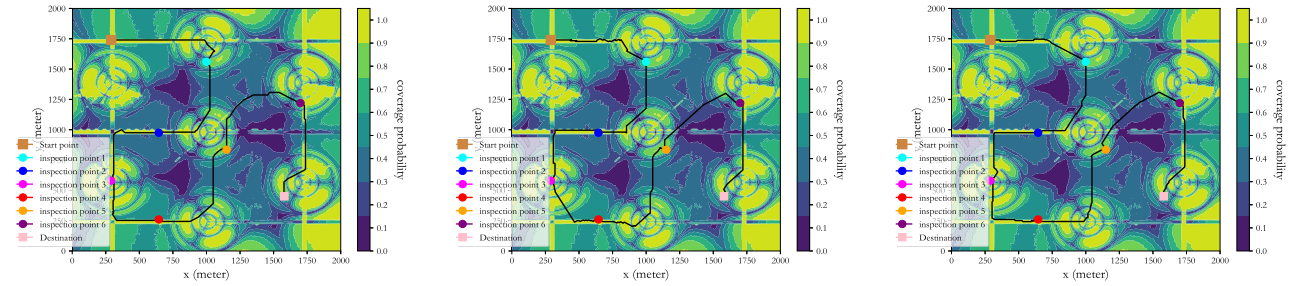
We show the communication outage time of AAV obtained based on different algorithms in Figure 9. In particular, to prove that D3QN is the state of the art reinforcement learning algorithm when the action space is discrete, we additionally show in Figure 9 the communication outage time of the AAV obtained based on SA-DDQN. As can be seen from Figure 9, when greedy-D3QN algorithm is used, the total communication outage time of AAV decreases with the increase of AAV battery energy $E_{\max}$, which corresponds to the trajectory variation trend of AAV based on greedy-D3QN algorithm in Figure 7. In addition, it can be found that regardless of the battery energy of the AAV, the total outage time based on SA-DDQN or SA-D3QN algorithm does not show a decreasing trend with the increase of the battery energy. This is consistent with the phenomenon that the AAV trajectory in Figure 8 shows almost no significant change. Finally, it can be found from Figure 9 that regardless of the battery energy of the AAV, the total communication outage time of the AAV based on the SA-D3QN algorithm is the smallest, that is, the value of the objective function in the problem (P8) is the smallest. Moreover, it can be seen that in optimizing the communication quality in the multi-inspection point system, the SA-D3QN algorithm is about 15%-50% better than the greedy-D3QN algorithm. This numerical result proves that optimizing the inspection sequence of AAV is very necessary to improve the communication quality of AAV. At the same time, the SA-D3QN algorithm improves by about 15%-25% compared with the SA-DDQN algorithm. This numerical result proves that D3QN is superior to DDQN algorithm.

To more intuitively illustrate the necessity of optimizing the inspection sequence for improving the communication quality of the energy limited AAV, we show the energy consumption of the AAV with different battery energies in Figure 10. It can be seen from Figure 10 that regardless of the battery energy of the AAV, the AAV energy obtained based on SA-DDQN algorithm or SA-D3QN algorithm is always lower than 80Kj, while the AAV energy consumption obtained based on greedy-D3QN algorithm is very close to the maximum battery energy $E_{\max}$ of the AAV. In addition, it can be seen from Figure 9 that the total communication outage time based on SA-DDQN algorithm or SA-D3QN algorithm is lower than the greedy-D3QN algorithm, which means that better communication quality can be obtained with less energy consumption. This is very meaningful for



(a) $E_{\max}$ = 85Kj (b) $E_{\max}$ = 90Kj (c) $E_{\max}$ = 95Kj

FIGURE 7. Based on greedy algorithm.



(a) $E_{\max}$ = 85Kj (b) $E_{\max}$ = 90Kj (c) $E_{\max}$ = 95Kj

FIGURE 8. Based on SA algorithm.

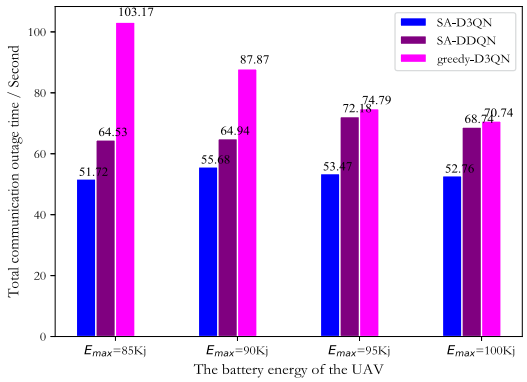


FIGURE 9. The total communication outage time of the AAV.

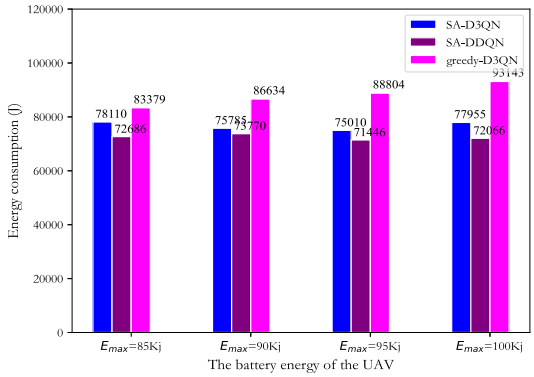


FIGURE 10. Energy consumption of AAV.

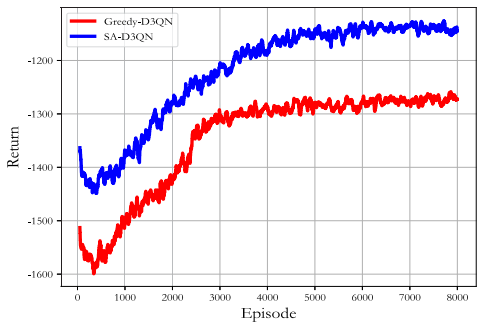


FIGURE 11. The moving average return obtained under different algorithm.

the AAV inspection system, and proves the necessity of optimizing the inspection sequence in the AAV inspection system. Moreover, it can be seen from Figure 10 that regardless of the battery energy of the AAV, the energy consumption based on SA-DDQN is always less than that based on SA-D3QN. This means that compared to the SA-DDQN, the AAV based on the SA-D3QN have more energy to adjust their flight trajectory, thereby achieving better communication quality. This also explains why the communication quality based on the SA-D3QN is superior to that of the SA-DDQN in Figure 9.

To prove the convergence and effectiveness of the algorithm, we show the moving average returns in Figure 11 - Figure 13. Figure 11 shows the moving average return obtained based on SA-D3QN algorithm and greedy-D3QN algorithm. As can be seen from Figure 11, both algorithms are convergent, but the moving average return obtained based

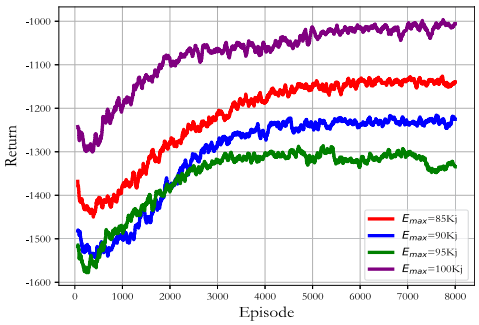


FIGURE 12. The moving average return obtained under different AAV battery energy storage.

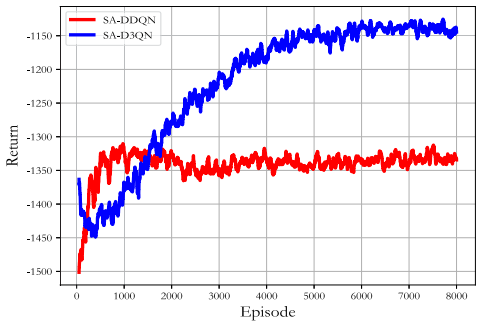


FIGURE 13. The moving average return obtained under different AAV battery energy storage when E_{max} = 85kj.

on SA-D3QN algorithm is greater than that obtained based on greedy-D3QN algorithm. Figure 12 shows the moving average return obtained under different AAV battery energy storage. As can be seen from Figure 12, regardless of the AAV's battery energy storage, the moving average return obtained by the proposed algorithm always show an upward trend with the increase of episodes, and eventually tend to be stable and converge. Finally, Figure 13 shows the moving average return based on different DRL algorithms. As can be seen from Figure 13, regression based on D3QN algorithm gradually increases when episode is greater than 800 and tends to converge when episode is greater than 6000, while regression based on DDQN algorithm always fluctuates when episode=2000, and the return obtained is smaller than that obtained by D3QN algorithm. This result shows that D3QN algorithm not only converges more stably, but also has better performance.

VI. CONCLUSION

In this paper, we studied a AAV inspection system in which the AAV has limited energy. To optimize the communication quality of the inspection AAV, we optimized the inspection sequence, and flight trajectory of the AAV. Numerical results proved that the algorithm proposed in this paper was effective and superior in solving the communication quality optimization problem of inspection AAV with limited energy, and the performance was improved by about 15%-50% compared with the comparison algorithm. In addition, we verified the performance of different battery energy storage, which proved the robustness of our proposed algorithm.

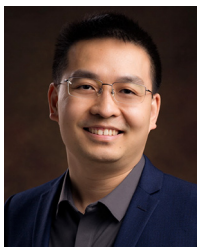
REFERENCES

[1] M. Hu et al., "Integrated robotics networks with co-optimization of drone placement and air-ground communications," in *Proc. IEEE Veh. Technol. Conf.*, 2023, pp. 1–5.
[2] H. Chen, Y. Lan, B. Fritz, C. Hoffmann, and S. Liu, "Review of agricultural spraying technologies for plant protection using unmanned aerial vehicle (UAV)," *Int. J. Agric. Biol. Eng.*, vol. 14, pp. 38–49, Feb. 2021.
[3] Y. Huang, H. Han, B. Zhang, X. Su, and Z. Gong, "Supply distribution center planning in UAV-based logistics networks for post-disaster supply delivery," in *Proc. IEEE Int. Conf. E-Health. Netw. Appl. Services (HEALTHCOM)*, 2021, pp. 1–6.
[4] J. Xiong, Z. Li, H. Li, L. Tang, and S. Zhong, "Energy-constrained UAV data acquisition in wireless sensor networks with the age of information," *Electronics*, vol. 12, no. 7, p. 1739, 2023.
[5] F. Hong, G. Wu, Q. Luo, H. Liu, X. Fang, and W. Pedrycz, "Logistics in the sky: A two-phase optimization approach for the drone package pickup and delivery system," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 9175–9190, May 2023.
[6] N. Hossein Motlagh, T. Taleb, and O. Arouk, "Low-altitude unmanned aerial vehicles-based Internet of Things services: Comprehensive survey and future perspectives," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 899–922, Dec. 2016.
[7] J. Li et al., "UAV-RIS-Aided space-air-ground integrated network: Interference alignment design and DoF analysis," *IEEE Trans. Wireless. Commun.*, vol. 23, no. 9, pp. 11678–11692, Sep. 2024.
[8] Q. Hao, H. Zhao, H. Huang, G. Gui, T. Ohtsuki, and F. Adachi, "Deep reinforcement learning aided online trajectory optimization of cellular-connected UAVs with offline map reconstruction," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, 2023, pp. 1–5.
[9] S. Zhang, Y. Zeng, and R. Zhang, "Cellular-enabled UAV communication: A connectivity-constrained trajectory optimization perspective," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2580–2604, Mar. 2019.
[10] Y.-J. Chen and D.-Y. Huang, "Trajectory optimization for cellular-enabled UAV with connectivity outage constraint," *IEEE Access*, vol. 8, pp. 29205–29218, 2020.
[11] M. R. Dileep, A. V. Navaneeth, S. Ullagaddi, and A. Danti, "A study and analysis on various types of agricultural drones and its applications," in *Proc. Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN)*, 2020, pp. 181–185.
[12] P. Agarwal and M. K. Singh, "A multipurpose drone for water sampling & video surveillance," in *Proc. Int. Conf. Adv. Comput. Commun. Paradigms (ICACCP)*, 2019, pp. 1–5.
[13] M. Matracia, M. A. Kishk, and M.-S. Alouini, "Aerial base stations for global connectivity: Is it a feasible and reliable solution?" *IEEE Veh. Technol. Mag.*, vol. 18, no. 4, pp. 94–101, Dec. 2023.
[14] S. Zhang, Y. Zeng, and R. Zhang, "Cellular-enabled UAV communication: Trajectory optimization under connectivity constraint," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2018, pp. 1–6.
[15] E. Bulut and I. Guevenc, "Trajectory optimization for cellular-connected UAVs with disconnectivity constraint," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2018, pp. 1–6.
[16] Y. Zeng, X. Xu, S. Jin, and R. Zhang, "Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4205–4220, Jul. 2021.
[17] H.-S. Im, K.-Y. Kim, and S.-H. Lee, "Trajectory optimization for cellular-enabled UAV with connectivity and battery constraints," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, 2023, pp. 1–7.
[18] H. Xie, D. Yang, L. Xiao, and J. Lyu, "Connectivity-aware 3D UAV path design with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 13022–13034, Dec. 2021.
[19] Y.-J. Chen and D.-Y. Huang, "Joint trajectory design and BS association for cellular-connected UAV: An imitation-augmented deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2843–2858, Feb. 2022.
[20] S. Zhang and R. Zhang, "Radio map-based 3D path planning for cellular-connected UAV," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1975–1989, Mar. 2021.

[21] N. Cherif, W. Jaafar, H. Yanikomeroglu, and A. Yongacoglu, "RL-based cargo-UAV trajectory planning and cell association for minimum handoffs, disconnectivity, and energy consumption," *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 7304–7309, May 2024.
[22] M.-A. Lahmeri, M. A. Kishk, and M.-S. Alouini, "Charging techniques for UAV-assisted data collection: Is laser power beaming the answer?" *IEEE Commun. Mag.*, vol. 60, no. 5, pp. 50–56, May 2022.
[23] L. Zhang, A. Celik, S. Dang, and B. Shihada, "Energy-efficient trajectory optimization for UAV-assisted IoT networks," *IEEE Trans. Mobile Comput.*, vol. 21, no. 12, pp. 4323–4337, Apr. 2022.
[24] J. Cao, L. Xiao, D. Yang, and F. Wu, "Energy consumption and communication quality tradeoff for logistics UAVs: A hybrid deep reinforcement learning approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2023, pp. 1–6.
[25] H. Hong, Y. Zhang, and Y. Xie, "Energy-limited UAV visiting planning for age-aware wireless-powered sensor networks," in *Proc. IEEE Veh.Technol. Conf.*, 2023, pp. 1–6.
[26] Y. Yu, J. Tang, J. Huang, X. Zhang, D. K. C. So, and K.-K. Wong, "Multi-objective optimization for UAV-assisted wireless powered IoT networks based on extended DDPG algorithm," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6361–6374, Sep. 2021.
[27] H.-T. Ye, X. Kang, J. Joung, and Y.-C. Liang, "Optimization for wireless-powered IoT networks enabled by an energy-limited UAV under practical energy consumption model," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 567–571, Mar. 2021.
[28] Q. T. Do, D. T. Hua, A. T. Tran, and S. Cho, "Energy efficient multi-UAV communication using DDPG," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, 2022, pp. 1071–1075.
[29] J. Liu, Y. Zhang, X. Wang, C. Xu, and X. Ma, "Min-max path planning of multiple UAVs for autonomous inspection," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2020, pp. 1058–1063.
[30] W. Zhang, D. Yang, F. Wu, and L. Xiao, "Trajectory design for UAV-based inspection system: A deep reinforcement learning approach," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2023, pp. 1654–1659.
[31] Y. Gao, L. Xiao, F. Wu, D. Yang, and Z. Sun, "Cellular-connected UAV trajectory design with connectivity constraint: A deep reinforcement learning approach," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 3, pp. 1369–1380, Sep. 2021.
[32] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329–2345, Apr. 2019.
[33] V. V. Díaz and D. Marcano Aviles, "A path loss simulator for the 3GPP 5G channel models," in *Proc. IEEE Int. Conf. Electron., Electr. Eng. Comput. (INTERCON)*, Nov. 2018, pp. 1–4.
[34] "Study on 3D channel model for LTE," 3GPP, Sophia Antipolis, France, Rep. 36.873, Dec. 2017.
[35] "Technical specification group radio access network: Study on enhanced LTE support for aerial vehicles," 3GPP, Sophia Antipolis, France, Rep. 36.777, Dec. 2017.
[36] A. Samandari and A. Willig, "TDMA slot allocation for UAV formations: Minimum superframe lengths for two-dimensional equidistant deployments," in *Proc. Int. Telecommun. Netw. Appl. Conf. (ITNAC)*, 2022, pp. 56–63.
[37] H. Xiao, H. Jiang, L.-P. Deng, Y. Luo, and Q.-Y. Zhang, "Outage energy efficiency maximization for UAV-assisted energy harvesting cognitive radio networks," *IEEE Sensors J.*, vol. 22, no. 7, pp. 7094–7105, Apr. 2022.
[38] D. Yang, J. Wang, F. Wu, L. Xiao, Y. Xu, and T. Zhang, "Energy efficient transmission strategy for mobile edge computing network in UAV-based patrol inspection system," *IEEE. Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5984–5998, May 2024.
[39] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 1998.
[40] B. Duo, M. He, Q. Wu, and Z. Zhang, "Joint dual-UAV trajectory and RIS design for ARIS-assisted aerial computing in IoT," *IEEE Internet Things J.*, vol. 10, no. 22, pp. 19584–19594, Nov. 2023.
[41] X. Guo, S. Zhang, and L. Liu, "Trajectory optimization of cellular-connected UAV for information collection and transmission," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*. Dec. 2022, pp. 5977–5982.

WEI WANG received the B.S. degree in electronic science and technology from the School of Mechanical and Electrical Engineering, East China University of Science and Technology, Nangchang, China, in 2022. She is currently pursuing the master's degree with the Information Engineering School, Nanchang University. Her research interests include unmanned aerial vehicle communications.

JIANGLING CAO received the B.S. degree in electronic science and technology from the School of Electrical and Information Engineering, Hunan Institute of Engineering in 2021, and the M.S. degree from the School of Information Engineering, Nanchang University in 2024. His research interests include AAV communications and machine learning.



DINGCHENG YANG (Member, IEEE) received the B.S. degree in electronic engineering and the Ph.D. degree in space physics from Wuhan University, Wuhan, China, in 2006 and 2012, respectively. He is currently a Professor with the Information Engineering School, Nanchang University, Nanchang, China. He had published more than 50 papers, including journal papers on IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and conference papers such as IEEE Globecom. His research interests include cooperation communications, IoT/cyber–physical systems, AAV communications, and wireless resource management.

HAO HE received the B.E. degree in electronic information engineering and the Ph.D. degree in communication and information systems from Wuhan University, China, in 2006 and 2012 respectively. He is currently an Associate Professor with the Jiangsu University of Science and Technology, China. His research interests include MIMO, interference coordination, cooperative communication, unmanned aerial vehicles, unmanned surface vehicles, satellites for search and rescue, MF/HF communication, and machine learning.

ZHIHAI XU (Member, IEEE) was born in Nanchang, Jiangxi, China, in 1975. He received the master's degree from the Huazhong University of Science and Technology, China. He currently works with the School of Information Engineering, Nanchang University, Nanchang, China. His research interests include computer network, database system, and big data analysis.