drones

MDPI

Article

Deep Reinforcement Learning-Based 3D Trajectory Planning for Cellular Connected UAV

Xiang Liu [1], Weizhi Zhong [1,*], Xin Wang [1], Hongtao Duan [2], Zhenxiong Fan [2], Haowen Jin [1], Yang Huang [1] and Zhipeng Lin [1]

[1] Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, Ministry of Industry and Information Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; liux6663@nuaa.edu.cn (X.L.); wangxin_dd@nuaa.edu.cn (X.W.); hwjin0126@nuaa.edu.cn (H.J.); yang.huang.ceie@nuaa.edu.cn (Y.H.); linlzp@nuaa.edu.cn (Z.L.)

[2] The State Radio Monitoring Center of China, Beijing 102609, China; duanht@srrc.org.cn (H.D.); fanzhenxiong@srrc.org.cn (Z.F.)

* Correspondence: zhongwz@nuaa.edu.cn

Abstract: To address the issue of limited application scenarios associated with connectivity assurance based on two-dimensional (2D) trajectory planning, this paper proposes an improved deep reinforcement learning (DRL) -based three-dimensional (3D) trajectory planning method for cellular unmanned aerial vehicles (UAVs) communication. By considering the 3D space environment and integrating factors such as UAV mission completion time and connectivity, we develop an objective function for path optimization and utilize the advanced dueling double deep Q network (D3QN) to optimize it. Additionally, we introduce the prioritized experience replay (PER) mechanism to enhance learning efficiency and expedite convergence. In order to further aid in trajectory planning, our method incorporates a simultaneous navigation and radio mapping (SNARM) framework that generates simulated 3D radio maps and simulates flight processes by utilizing measurement signals from the UAV during flight, thereby reducing actual flight costs. The simulation results demonstrate that the proposed approach effectively enable UAVs to avoid weak coverage regions in space, thereby reducing the weighted sum of flight time and expected interruption time.

Keywords: cellular connected UAV; trajectory planning; radio mapping; deep reinforcement learning

Citation: Liu, X.; Zhong, W.; Wang, X.; Duan, H.; Fan, Z.; Jin, H.; Huang, Y.; Lin, Z. Deep Reinforcement Learning-Based 3D Trajectory Planning for Cellular Connected UAV. *Drones* **2024**, *8*, 199. https://doi.org/10.3390/drones8050199

Academic Editor: Andrey V. Savkin

Received: 10 April 2024
Revised: 8 May 2024
Accepted: 8 May 2024
Published: 15 May 2024

1. Introduction

In recent years, with the rapid advancement of unmanned aerial vehicles (UAVs) technology and the increasing maturity of wireless communication technology, UAVs have demonstrated extensive potential applications in aerial photography, logistics distribution, search and rescue. Some researchers have conducted detailed studies in areas such as spectrum for air-to-ground (A2G) communication and wireless communication environments [1,2]. However, existing UAVs communications still inevitably be limited by factors such as controller or WIFI connection modes, resulting in restricted communication range, low data transmission rates and susceptibility to interference. The cellular network is a widely distributed mobile communication network with high capacity. Integrating UAVs into the cellular network can enhance communication distance, achieve higher data transmission rate and lower latency, as well as supplement positioning accuracy in adverse weather conditions or when obstacles affect GPS signals, thereby mitigating environmental impacts on communications. Consequently, cellular-connected UAVs communication emerges as a promising research area [3].

Despite the aforementioned advantages of cellular-connected UAVs communication, there are still several challenges that need to be addressed. Firstly, in order to cater to a larger number of ground users, the antenna orientation of the ground base station (GBS) is typically optimized for ground coverage, which may result in inadequate air

communication coverage [4]. Secondly, three-dimensional (3D) obstacles such as buildings may obstruct the communication link [5]. Additionally, building upon prior research on the A2G channel model [6–8], cellular-connected UAVs may encounter significant signal interference due to potential line-of-sight channel issues between the UAVs and non-associated base station (BS), as shown in Figure 1.

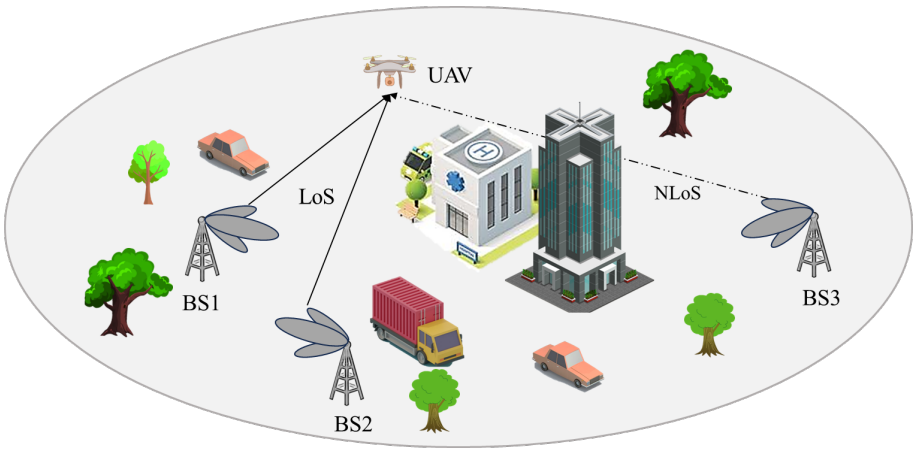


Figure 1. Schematic representation of UAV communication channel for cellular connectivity in an urban setting.

To address the aforementioned issues, the authors in [9] combined with the analysis of uplink/downlink 3D coverage performance, introduced the generalized Poisson multinomial distribution to simulate interference information and demonstrated the impact of different downdip angles of GBS antennas on 3D coverage. Additionally, Ref. [10] employed deep reinforcement learning to train aerial BS layout decision strategies, thereby enhancing coverage in complex environments. Ref. [11] investigated the performance of cellular connected UAVs under actual antenna configurations and revealed how the number of antenna units influences coverage probability and handover rate. Ref. [12] optimized the downdip angle of GBS antennas to maximize received signal quality for UAVs while ensuring throughput performance for ground users and reducing switching times. To mitigate strong ground-to-air interference, various anti-interference techniques were proposed [13–16]. For instance, Ref. [13] presented a novel cooperative interference elimination strategy for multi-beam UAVs uplink communication that aims to eliminate co-channel interference on each occupied GBS while maximizing the summation rate of available GBS.

Efficient path planning should ensures optimal air-ground communication conditions, high data transmission rates, and reliable connectivity while minimizing unnecessary movements of UAVs, thereby enhancing energy efficiency [17,18]. In [19], the problem of shortest path planning under the constraint of minimum reception SNR was investigated. In [20,21], the authors employed graph theory to design the shortest path under the minimum SINR constraint and deduced an optimal UAV path by solving an equivalent shortest path problem in graph theory. Ref. [22] proposed the constraction of a received signal intensity map using a distributed recursive Gaussian process regression framework. This approach achieves higher positioning accuracy with lower complexity and storage requirements, making it an efficient solution for positioning applications. Similar problems have been addressed in [23–25]. Some traditional trajectory optimization schemes simplified channel models in various environments for ease of solution. However, environmental models such as those assuming path loss of channels or isotropic radiation of antennas are not applicable to real-world scenarios [26]. Moreover, the trajectory optimization problem is non-convex, and its complexity increases dramatically with the number of optimization variables, which is difficult to solve effectively. Fortunately, machine learning techniques have emerged as another solution for non-convex optimization problems. For instance, Ref. [27] presents a

two-dimensional (2D) radio map-based approach for path planning in conjunction with machine learning techniques. Nevertheless, 2D path planning has limitations regarding its applicability and susceptibility to local optima; thus further research should focus on 3D path planning.

Some recent studies, such as [28], have proposed a Multi-Layer Trajectory Planning (MTTP) method, addressing the challenges of ensuring air-to-ground communication services and avoiding collisions in complex urban environments. The work referenced in [29] introduces a two-step centralized development system for 3D path planning of drone swarms. Additionally, Both articles [30,31] take into account energy consumption during the 3D trajectory planning process for UAVs. Ref. [32] proposed collaborative UAV trajectory optimization using federated learning to overcome challenges in ensuring reliable connectivity in 3D space. In this paper, we propose a novel synchronous path planning approach based on an improved deep reinforcement learning (DRL) algorithm, integrated with radio mapping techniques, to optimize the 3D trajectory of UAV. This method aims to efficiently navigate UAVs by avoiding areas with weak communication coverage and reaching the destination in minimal time. The major contributions and novelties of this paper are summarized as follows:

- We propose a 3D path optimization strategy that aims to minimize the weighted sum of task completion time and communication interruption time, thereby enhancing the efficiency and reliability of the system.
- We employ a multi-step dueling double deep Q network (D3QN) method incorporating with prioritized experience reply (PER) mechanism to efficiently optimize the proposed objective function and acquire the optimal path.
- We propose a simultaneous navigation and radio mapping (SNARM) framework that leverages 3D radio mapping and simulates flight processes to optimize the cost-effectiveness of real flights while enhancing learning accuracy.

The remainder of the article is organized as follows. Section 2 introduces the system model and problem formulation. Section 3 presents the improved DRL-based 3D path planning strategy. The simulation results and analysis are provided in Section 4. The conclusions are drawn in Section 5.

2. System Model and Problem Formulation

2.1. 3D Flight Environment Model

In this paper, the UAV operates in the airspace above a dense urban area measuring 2 km × 2 km. The height and locations of urban buildings are generated using the statistical model recommended by the International Telecommunication Union (ITU). This model involves three parameters: α_{bd}, which represents the ratio of land area covered by buildings to the total land area; β_{bd}, which denotes the average number of buildings per unit area; and γ_{bd}, a variable determining the distribution of building heights following a Rayleigh distribution, with a mean value of σ_{bd} ($\sigma_{bd} > 0$). Figure 2 shows the 2D views of one particular realization of the building locations and heights with $\alpha_{bd} = 0.3$, $\beta_{bd} = 300$ buildings/km^2, and $\sigma_{bd} = 50$ m. For convenience, the building height is clipped to not exceed 70 m.

The UAV's flight parameters include a flying height ranging from $h_{\min}$ to $h_{\max}$, a constant flight speed of V m/s, and the UAV's position at any given moment denoted as $q(t) = (x_t, y_t, h_t)$. The starting and ending points of the UAV's flight are represented as $q_s = (x_s, y_s, h_s)$ and $q_f = (x_f, y_f, h_f)$, respectively.

Within the target area, a total of 7 GBS are distributed in a honeycomb pattern, as indicated by black star markers in Figure 2. The GBS antenna stands at a height of h_{bs}, and each GBS site comprises 3 sectors, resulting in a total of M = 21 sectors. The GBS antenna is a vertically oriented 8-element uniform linear array (ULA) with a half-power beamwidth of 65° in both horizontal and vertical directions. The main lobe is tilted 10° to the ground, forming a directional antenna array.

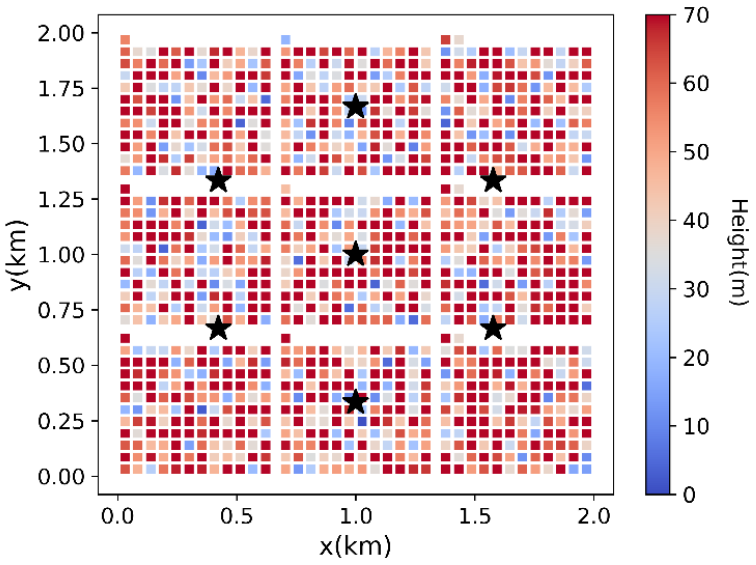


Figure 2. Top view of building and base station distribution.

2.2. Reception Signal Model

In the system model, we simulate path loss using the Urban Microcell (UMI) model specified by 3GPP. It is worth noting that the statistical building model has been widely used to estimate the line-of-sight (LoS) probability of ground-to-air links [33]. However, this model only reflects the average characteristics of large-scale geographic areas with similar types of terrain. For each local area with given building positions and heights, the presence/absence of LoS links with cellular base stations can be accurately determined by examining whether the communication path between the base stations and UAVs is obstructed by any buildings. The path loss for the LoS link between the UAV and sector m is represented as follows

$$h_m^{LoS}(t) = max\left\{h_m^{FSPL}, 30.9 + (22.25 - 0.5\log_{10}h_t)\log_{10}d_m(t) + 20\log_{10}f_c\right\} \quad (1)$$

where h_m^{FSPL} represents the free-space path loss, h_t represents the altitude of the UAV at time t, $d_m(t)$ represents the distance between the UAV and sector m, and f_c is the carrier frequency. When the communication path between the base station sector m and the UAV is obstructed by obstacles, a non-line-of-sight (NLoS) channel is formed, characterized by a path loss denoted as

$$h_m^{NLoS}(t) = max\left\{h_m^{LoS}(t), 32.4 + (43.2 - 7.6\log_{10}h_t)\log_{10}d_m(t) + 20\log_{10}f_c\right\}. \quad (2)$$

The channel gain between the UAV and sector m, denoted as $h_m(t)$, is primarily determined by three factors, GBS antenna gain, large-scale channel fading, and small-scale fading. According to [34], the received instantaneous signal power at the UAV from sector m can be mathematically expressed as

$$y_m(t) = P_m|h_m(t)|^2 = P_m\beta(q(t))\bar{h}_m(q(t))\tilde{h}_m(t), m \in M \quad (3)$$

where the constant P_m represents the transmit power of GBS in sector m, while $\beta(q(t))$ and $\bar{h}_m(q(t))$ respectively denote GBS antenna gain and large-scale channel fading. The variable $\tilde{h}_m(t)$ signifies the channel gain under small-scale fading, and $\bar{h}_m(q(t))$ can be determined by the building's location between the UAV and GBS

$$\bar{h}_m(q(t)) = \begin{cases} h_m^{LoS}(q(t)), LoS \\ h_m^{NLoS}(q(t)), NLoS. \end{cases} \quad (4)$$

The sector associated with the UAV at time t is denoted as $b(t) \in \{1, \cdots, M\}$. Consequently, the descending instantaneous SIR can be mathematically formulated as

$$\gamma(t) = \frac{y_{b(t)}(t)}{\sum_{m \neq b(t)} y_m(t)}. \tag{5}$$

The small-scale fading $\tilde{h}_m(t)$ introduces randomness to the variable $\gamma(t)$ at any given location $q(t)$ and its associated unit $b(t)$. To assess the reliability of the UAV-to-target cell link, we introduce the interrupt probability function as follows

$$P_{\text{out}}(q(t), b(t)) = \Pr\{\gamma(t) < \gamma_{th}\}. \tag{6}$$

The interruption of the connection to the GBS-UAV is considered when the SIR $\gamma(t)$ falls below the interruption threshold γ_{th}, where event probability $\Pr\{\cdot\}$ indicates its likelihood.

The direct solution of $P_{\text{out}}(q(t), b(t))$ being unattainable, we reformulate the instantaneous $\gamma(t)$ as a function of $q(t)$, $b(t)$, and small-scale fading $\tilde{h}_{b(t)}$. Subsequently, we define the interrupt indicator function as follows

$$c(q(t), b(t), \tilde{h}_{b(t)}) = \begin{cases} 1, \gamma(q(t), b(t), \tilde{h}_{b(t)}) < \gamma_{th} \\ 0, \text{otherwise.} \end{cases} \tag{7}$$

Then, the interrupt probability function in (6) can be expressed as the expectation of small-scale fading $\tilde{h}_{b(t)}$, i.e.,

$$P_{\text{out}}(q(t), b(t)) = E_{\tilde{h}_{b(t)}}\left[c(q(t), b(t), \tilde{h}_{b(t)})\right]. \tag{8}$$

The interruption probability of each time point t is obtained by conducting J-time signal measurements on M sectors within a short duration using the UAV. The j-th measurement of the small-scale fading is denoted as $\tilde{h}_{\mathrm{b}(t)}[\mathrm{t}, j]$, the corresponding SIR and the outage indication function are denoted as $\gamma(q(t), b(t), \tilde{h}_{\mathrm{b}(t)}[\mathrm{t}, j])$ and $\mathrm{c}(q(t), b(t), \tilde{h}_{b(t)}[t, j])$, therefore the corresponding communication interruption probability can be expressed as

$$\hat{P}_{out}(q(t), b(t)) \triangleq \frac{1}{J} \sum_{j=1}^{J} \mathrm{c}(q(t), b(t), \tilde{h}_{b(t)}[t, j]). \tag{9}$$

According to the large number theorem, $\hat{P}_{out}(q(t), b(t))$ can provide an accurate approximation of the actual interruption probability at $q(t)$ when J is sufficiently large. The optimal associated cell can be denoted as

$$b^*(t) = \underset{b(t) \in \{1, \cdots, M\}}{\arg\min} \hat{P}_{out}(q(t), b(t)) \tag{10}$$

where arg min signifies the argument or input value that minimizes the corresponding functionand and the estimation of the interruption probability at any given location can be calculated by

$$\hat{P}_{out}(q(t)) = \hat{P}_{out}(q(t), b^*(t)). \tag{11}$$

According to the aforementioned analysis, the anticipated interruption probability of UAV at any given location can be derived, enabling the construction of a 3D coverage probability graph (CPG). The constructed coverage probability map will be shown in Section 4, where coverage probability $\hat{P}_{\text{coverage}}(q(t)) = 1 - \hat{P}_{out}(q(t))$.

2.3. UAV Motion Model

The rotor UAV utilized in this experiment primarily consumes energy in two main aspects. The first aspect pertains to communication, encompassing signal processing, radiation, and circuitry. The second aspect involves propulsion energy, which is essential for sustaining the UAV's flight and movement. It is noted that the communication-related energy consumption of UAVs is considered negligible due to its typically smaller magnitude compared to the propulsion energy of UAVs [35]. According to [36], the instantaneous propulsion energy of a rotor UAV with a velocity of V can be expressed as

$$P(V) = P_0(1 + \frac{3V^3}{U_{tip}^2}) + P_i(\sqrt{1 + \frac{V^4}{2v_0^2}})^{1/2} + \frac{1}{2}d_0\rho sAV^3 \tag{12}$$

where P_0 and P_i are constants, representing the UAV's blade profile power and induced power in hovering states, respectively. v_0 represents the mean rotor induced velocity in hover, U_{tip} signifies the tip speed of the rotor blade, and d_0 and s denote the fuselage drag ratio and rotor solidity, respectively. ρ and A denote air density and rotor disc area, respectively. In a given environment, with all environmental parameters and UAV settings held constant, the power required for UAV flight remains constant for a given speed. Therefore, the consumed energy of the rotary-wing UAV during time T can be expressed as $E = \int_0^T P(V)dt = P(V)T$. It can be deduced that the energy consumption of UAVs is directly proportional to their flight time, indicating that longer flight durations result in higher energy consumption.

In this study, we prioritize flight time over energy consumption as our research metric. By imposing a maximum flight time constraint, we ensure the safe operation of UAVs. Additionally, we introduce the concept of communication interruption time, denoted as $\int_0^T \hat{P}_{out}(q(t))dt$, to represent the communication quality of UAVs within a given time period. The main objective of our study is to train UAVs to acquire optimal flight strategies. If UAVs solely focus on energy consumption, they would instinctively choose the shortest path from the starting point to the destination, inevitably compromising the communication quality between UAVs and associated ground stations. Similarly, if only communication quality is prioritized, it would significantly increase the energy consumption of UAVs. To address this trade-off, we introduce a weighting coefficient, denoted as μ, which combines the flight time and estimated interruption time of the UAV. By minimizing the weighted sum of both factors, we aim to achieve a balanced optimization between energy consumption and communication quality between UAVs and ground stations.

Based on the obtained CPG of 3D space, the optimization objective equation can be formulated as follows

$$\max_{T,\{q(t)\}} \quad -T - \mu \int_0^T \hat{P}_{out}(q(t))dt \tag{13}$$

$$s.t.\ q(0) = q_s, \tag{14}$$

$$q(T) = q_f, \tag{15}$$

$$\|\dot{q}(t)\| = V, \quad \forall t \in [0, T] \tag{16}$$

$$0 \le x_t \le D, \quad \forall t \in [0, T] \tag{17}$$

$$0 \le y_t \le D, \quad \forall t \in [0, T] \tag{18}$$

$$h_{\min} \le h_t \le h_{\max}, \quad \forall t \in [0, T] \tag{19}$$

where μ represents the trade-off between the flight time and the expected outage time. A higher value of μ indicates a greater emphasis on maintaining connectivity between the UAV and the GBS, but at the cost of potentially increased travel distance for the UAV. The constraints on the starting and ending positions are represented by (14) and (15), while

the limitation of the UAV velocity is denoted by (16). Additionally, (17)–(19) specify the constraints on the 3D motion space of the UAV.

The path planning problem can be formulated as a markov decision process (MDP) that is amenable to solution using DRL. However, addressing the continuous optimization aspect of (13) introduces challenges due to the inherent complexity arising from continuous state and action spaces. This often leads to instability or non-convergence during DRL training. To mitigate these issues, we convert problem (13) into a discrete-time formulation by discretizing the time period, which can be expressed as

$$\max_{N,\{\mathrm{q}(n)\}_{n=0}^{N}} \quad -N - \mu \sum_{n=1}^{N} \hat{P}_{out}(q(n)) \tag{20}$$

$$s.t.\ q(0) = q_s, \tag{21}$$

$$q(N) = q_f, \tag{22}$$

$$q_{n+1} = q_n + \Delta s \vec{a}_n, \quad \forall n \tag{23}$$

$$\|\vec{a}_n\| = 1, \quad \forall n \tag{24}$$

$$0 \le x_n \le D, \quad \forall n \tag{25}$$

$$0 \le y_n \le D, \quad \forall n \tag{26}$$

$$h_{\min} \le h_n \le h_{\max}, \quad \forall n \tag{27}$$

where $T = N\Delta t$, $\Delta s = a\Delta t$, and the time interval should be sufficiently small so that within each time step, the distance between the UAV and any GBS in the target area remains approximately constant, while ensuring that both the antenna gain and channel state parameters between the UAV-GBS remain nearly constant.

3. 3D Path Planning Based on Improved DRL

To address problem (20), we employ the multi-step D3QN model in DRL to optimize the objective function, and use PER mechanism instead of the conventional random experience replay (RER) approach to enhance learning efficiency and expedite convergence. Moreover, for assisting path planning, a radio mapping network is incorporated to generate simulated 3D radio maps and simulate flight processes. This incorporation not only mitigates flight costs, but can also enhances the accuracy of the D3QN network model. The improved framework based on DRL is then applied to UAV path planning, enabling efficient identification of an optimal route that satisfies all constraints from any given starting point.

3.1. Multi-Step D3QN Model

In this section, we will briefly introduce the relevant knowledge of DRL and provide an overview of the specific components of the multi-step D3QN model employed in this paper.

In the reinforcement learning model, the agent and the environment play crucial roles. The agent selects actions a_n based on the current state s_n provided by the environment, while its own state changes to s_{n+1} according to state transition function, with rewards r_{n+1} being fed back to the environment. By iteratively following this process, the agent can efficiently converge towards an optimal strategy within a specific environment. Execution of this optimal policy leads to maximum cumulative reward G_n for agent movements, which can be defined as

$$G_n = \sum_{k=0}^{\infty} \gamma^k r_{n+k+1} \tag{28}$$

where $\gamma \in [0, 1]$ represents the discount factor, denoting the future reward discounted at the prevailing rate. A higher value of γ signifies greater emphasis on long-term gains, while a lower value indicates prioritization of short-term benefits.

Since the cumulative reward G_n is unknown prior to the completion of the agent's trajectory, we estimate the expected reward instead of its actual value to derive the action value function Q_π

$$Q_\pi(s,a) = \mathrm{E}[G_n|s_n = s, a_n = a] \tag{29}$$

where $\pi(a_n, s_n) = \mathrm{P}[a = a_n|s = s_n]$ represents the state transition function, denoting the probability of an action a_n being performed by the agent while in a particular state s_n. The action value function Q_π signifies the expected return obtained by adhering to a given policy $\pi(a_n, s_n)$. If there exists a strategy function capable of selecting the optimal action for the agent at each state during its trajectory, it is referred to as the optimal strategy $\pi_*(s)$. Under this guidance, the optimal action value function $Q_*(s,a)$ can be expressed as follows

$$Q_*(s,a) = \max_\pi Q_\pi(s,a) = r(s,a) + \gamma \sum_{s'} p(s'|s,a) \max_{a'} Q_*(s',a'). \tag{30}$$

In principle, by exhaustively traversing all possible sequences $(s_n, a_n, r_{n+1}, s_{n+1})$ and iteratively optimizing, we can obtain the optimal value for $Q_*(s,a)$ and subsequently determine the optimal strategy $\pi_*(s)$. However, to address the limitation of Q-learning in dealing with continuous high-dimensional state or action spaces, we employ the classical DQN network model instead of the Q table as a function approximator, and update the network parameters by minimizing the loss function

$$loss = (r_{n+1} + \gamma \max_a \hat{Q}(s_{n+1}, a|\theta) - \hat{Q}(s_n, a_n|\theta))^2 \tag{31}$$

where θ denotes neural network parameter vector. However, the direct utilization of (31) in the standard training algorithm may give rise to the issue of overestimating Q value, thereby leading to learning instability and inefficiency. To address this challenge, we introduce Double DQN into our research, aiming to mitigate overestimation. This approach separates the selection of the target Q value from the estimation process by leveraging the policy network to determine the optimal action and utilizing the target network to estimate the corresponding Q value. In accordance with the Double DQN model, we can reformulate the loss function as

$$loss = (r_{n+1} + \gamma \hat{Q}(s_{n+1}, \arg\max_{a'} \hat{Q}(s_{n+1}, a'|\theta)|\theta') - \hat{Q}(s_n, a_n|\theta))^2 \tag{32}$$

where θ' denotes the parameter vector of target network. Additionally, to enhance the effectiveness of learning state value information and address bias-variance trade-off in training, this study introduces the dueling network and n-step bootstrapping techniques to improve the Double DQN model, and the improved model was represented as multi-step D3QN model. The dueling network models both the state value function V_s and the action advantage function $A(s,a)$, respectively, enabling the network to learn the relative value of each state as well as the advantages of different actions. By decomposing the network's output into status value and action advantages, we obtain a comprehensive Q value by combining these two components, which can be expressed as

$$\hat{Q}(s,a|\theta,\alpha,\beta) = V(s|\theta,\beta) + A(s,a|\theta,\alpha) - \frac{1}{|A|}\sum_{a'} A(s,a'|\theta,\alpha) \tag{33}$$

where α and β are the parameters of the advantage stream and the value stream, respectively, and $|A|$ represents the size of the action space.

Multi-step bootstrap is an improved learning style in reinforcement learning that aims to improve the efficiency of learning by considering the rewards of multiple future time steps N_1, which can be expressed as

$$R_{n:n+N_1} = \sum_{k=0}^{N_1-1} \gamma^k r_{n+k+1}. \tag{34}$$

It is worth noting that the return accumulates to a maximum of N steps, when $n + N_1 \geq N$, $R_{n:n+N_1} = R_{n:N}$.

The loss function of the D3QN model, incorporating multi-step bootstrap technology, can be summarized as follows

$$loss = (R_{n:n+N_1} + \gamma^{N_1} \hat{Q}(s_{n+N_1}, \arg\max_{a'} \hat{Q}(s_{n+N_1}, a'|\theta)|\theta') - \hat{Q}(s_n, a_n|\theta))^2 \quad (35)$$

3.2. Priority Experience Replay

Experiential playback is a crucial technique in DRL. Its fundamental concept involves storing the experiences acquired through agent-environment interactions and sampling them randomly for learning, thereby reducing sample correlation. However, randomly selecting samples may result in the loss of crucial experiences, thereby impacting the learning efficacy and, consequently, the effectiveness of UAV path planning. To address this issue, we propose employing PER instead of traditional RER by assigning priorities to each experience. During the process of sample extraction, samples with higher priority are more likely to be selected, thus enhancing the efficiency of sample training.

The PER mechanism assigns sampling weights based on the absolute value of the temporal difference error (TD-error). In this mechanism, the priority of each experience is set to $p_i = |\delta_i| + \sigma$, where $|\delta_i|$ represents TD-error, and the parameter σ is a constant greater than 0, which is used to ensure all $p_i > 0$. Notably, higher TD-errors correspond to greater experience priorities. Consequently, the sampling probability for each experience can be defined as follows

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha}. \quad (36)$$

the hyperparameter $\alpha \geq 0$ controls the intensity of priority playback, while $\alpha = 0$ represents traditional random experience playback where each experience is sampled with equal probability. Additionally, $\sum_k p_k^\alpha$ denotes the sum of all experience priorities in the buffer. The sampling probability $P(i)$ can be utilized for calculating the loss function.

To mitigate the computational complexity arising from priority sampling as the number of experiences increases, we employ a sum-tree data structure to store priorities and conduct sampling operations. Given a sample size of k, priority $\left(0, \sum_k p_k^\alpha\right]$ is divided into an average of intervals. A random value is generated in each interval and the corresponding transition sample is extracted from the sum-tree. However, changing the priority of the sample will introduces errors into the data distribution. To compensate for this error, importance sampling weights are introduced and can be expressed as follows

$$w_j = \left(\frac{P(j)}{\min_i P(i)}\right)^{-\beta} \quad (37)$$

where β is a hyperparameter that determines how much PER affects the convergence result, and the loss function in (35) can be rewritten as

$$loss = w_j(R_{n:n+N_1} + \gamma^N \hat{Q}(s_{n+N_1}, \arg\max_{a'} \hat{Q}(s_{n+N_1}, a'|\theta)|\theta') - \hat{Q}(s_n, a_n|\theta))^2 \quad (38)$$

3.3. SNARM Framework

Due to the lack of prior environmental knowledge, relying solely on the actual flight of UAVs not only incurs high training costs and a slow learning process, but also poses a significant risk of accidents. To address this issue, we propose the SNARM framework in this paper, which utilizes UAV measurement signals during flight to generate a simulated 3D radio map and create a virtual flight trajectory. By doing so, the UAV can predict the expected outcome for each path without physically traversing, thereby reducing the cost of measured flight and mitigating potential risks. Furthermore, we employ the Dyna

framework to integrate simulation experience with real-world experience in updating UAV flight strategies within deep learning algorithms, thus enhancing the accuracy of neural network. The Dyna framework is shown in Figure 3 below.

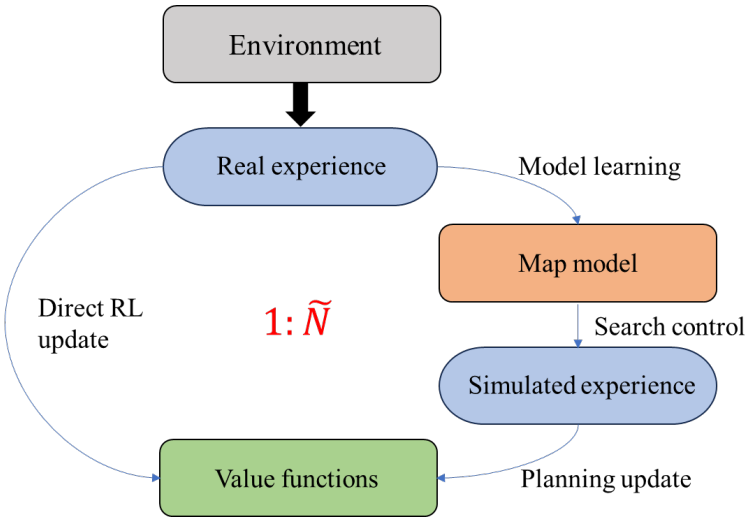


Figure 3. Dyna algorithm framework.

Notably, the simulated trajectory is utilized more frequently than the actual path, and for each real episode taken by the UAV, $\tilde{N} = \min([n/200], 5)$ episodes are employed in the simulated trajectory. Initially, the limited efficacy of the map model in learning resulted in a relatively low reference value for simulation experience, leading to a reduced contribution of simulation experience towards neural network updates. As the accuracy of the local map model improves, there will be an increased proportion of simulation experience involved in network updates. Since acquiring simulation experience does not necessitate actual UAV measurements, it is possible to appropriately increase the proportion of simulation experience without concerns about additional UAV operating costs and algorithm runtime consumption.

3.4. Path Planning Based on Improved DRL

In the enhanced DRL model, the UAV functions as an autonomous agent that strategically selects the optimal course of action based on its current state, subsequently receiving rewards from the environment and transitioning to subsequent states. The comprehensive depiction of the state space, action space, and reward function is expounded upon in meticulous detail as follows:

- State: The state serves as the input of the neural network, representing the UAV's 3D positions. The state space S encompasses all potential UAV positions within the terrain of interest $S = \left\{ q : q_s \leq q_n \leq q_f \right\}$. For each episode, the initial location of the UAV is randomly generated, while the final location is predetermined.
- Action: The action space A corresponds to the UAV flying direction. Considering the limited vertical range of the UAV's flying area, the action space of the UAV consists of 10 directions, including 8 horizontal directions spaced 45 degrees apart, as well as upward and downward directions, as shown in Figure 4. The selection of UAV motion direction relies on the model's estimation of the Q value for each direction in specific position.
- Reward: The reward R is defined as $R(q) = -1 - \mu \hat{P}_{out}(q)$, and the UAV incurs a penalty of 1 for each step taken before reaching the endpoint q_f. Additionally, if it enters an area with weak coverage, it will be penalized by a weighted value of μ. This encourages the UAV to consider both flight time and interruption time to determine the optimal path towards the endpoint.

Algorithm 1 N1-Step D3QN with PER for Connectivity-Aware UAV Path Planning

Initialize: number of episodes N_{epi}, maximum number of steps per episode N_{step}, number of multi-step learning steps N_1, experience buffer D of size C, initial exploration rate ε_0, exploration decay rate α and experience extraction number B

Initialize: Q network with parameter θ, target network with parameter θ^-, network update rate δ, and radio map network E with parameter θ_{radio}

for $n_{\text{epi}} = 1, \cdots, N_{\text{epi}}$ **do**

  Initialize the sliding window W of size N, the actual starting position q_s, the simulated starting position $\tilde{q}_s$, and the flight step $n = 0$, $\tilde{n} = 0$

  Select the action with ε-greedy policy

  Perform action a_n to obtain the next state q_{n+1}, measure the probability of communication interruption $\hat{P}_{out}(q_{n+1})$, and save it in map network E

  Update the map network E with $(q_{n+1}, \hat{P}_{out}(q_{n+1}))$

  Set single-step reward $R_n = -1 - \mu\hat{P}_{out}(q_{n+1})$ and store sequence (q_n, a_n, R_n, q_{n+1}) in slide window W

  When $n \geq N_1$, calculate $R_{(n-N_1):n}$ and store $(q_{n-N_1}, a_{n-N_1}, R_{(n-N_1):n}, q_n)$ in experience buffer D

  Extract B sequence $(q_j, a_j, R_{j:j+N_1}, q_{j+N_1})$ and its priority w_j from D according to PER mechanism

  Set $y_j = \begin{cases} R_{j:j+N_1} + R_{des}, q_{j+N_1} = q_f \\ R_{j:j+N_1} + R_{out}, q_{j+N_1} \notin S \\ R_{j:j+N_1} + \gamma^{N_1}\hat{Q}(q_{j+N_1}, a^*|\theta^-), \text{otherwise} \end{cases}$

  Perform a gradient descent step on $w_j(y_j - \hat{Q}(q_j, a_j|\theta))^2$ with respect to network parameters θ

  for $\tilde{n}_{epi} = 1, \cdots, \tilde{N}_{epi}$ **do**

    Perform steps (3–4,6–10) for the simulated experience, where the interrupt probability of q_{n+1} is predicted by map network E.

    $\tilde{n} = \tilde{n} + 1$

    Until $\left\|\tilde{q}_n - q_f\right\| \leq D_{tol}$, $\tilde{h}_n = h_{\text{target}}$; $\tilde{q}_n \notin S$ or $\tilde{n} = N_{step}$, reinitialize $\tilde{q}_s$, $\tilde{n} = 0$

  end for

  $n = n + 1$,$\varepsilon = \varepsilon\alpha$

  Repeat steps 3–16 until $\left\|q_n - q_f\right\| \leq D_{tol}$, $h_n = h_{\text{target}}$; $q_n \notin S$ or $n = N_{step}$,

  After every δ episodes, set the target network parameters $\theta^- = \theta$

end for

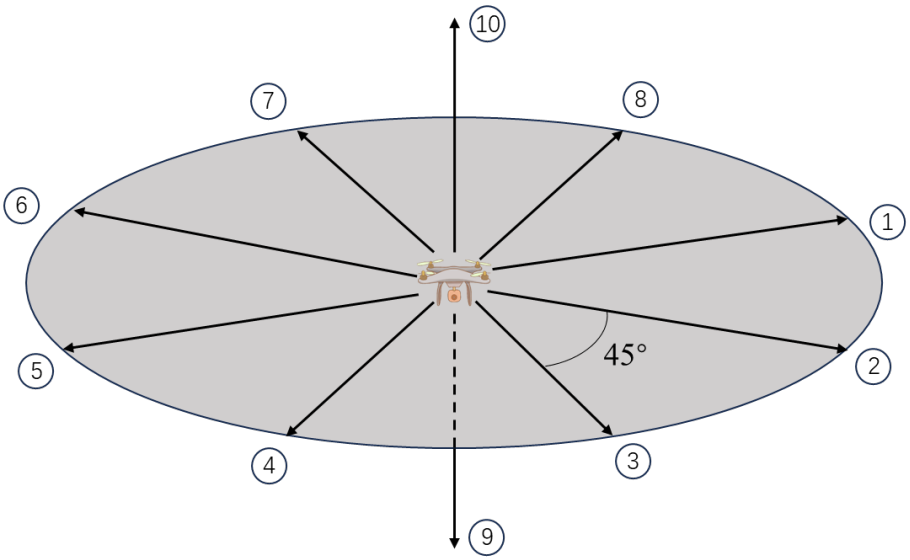


Figure 4. Diagram of UAV flight direction.

In the multi-step D3QN model, the UAV obtains the state from the state space and selects the action from the action space according to $\varepsilon - greedy$ strategy, i.e.,

$$a = \begin{cases} \text{randomly selected from } A, & p = \varepsilon \\ \arg\max_{a \in A} \hat{Q}(s, a|\theta), & p = 1 - \varepsilon \end{cases} \tag{39}$$

where $\varepsilon \geq 0$ represents the random exploration rate , θ denotes a multi-step D3QN network parameter, and the value of ε gradually decreases as the number of iterations increases. During the initial learning stage, the UAV conducts random exploration with a high probability to gather sufficient environmental information. As the UAV accumulates more experience, it becomes more inclined to select directions corresponding to maximum Q values. To enhance algorithm convergence, parameter θ is initialized based on the distance between the UAV and endpoint. After initialization, $\hat{Q}(q, a|\theta) = -\left\|q' - q_f\right\|$ can be obtained, where q' represents the next state of UAV after action a is performed in state q, and q_f signifies the endpoint coordinate. This encourages optimal path selection when radio environment understanding is limited during early stages. Additionally, θ_{radio} serves as a parameter for radio map network E and undergoes random initialization. The 3D path planning framework of UAV is shown in Figure 5.

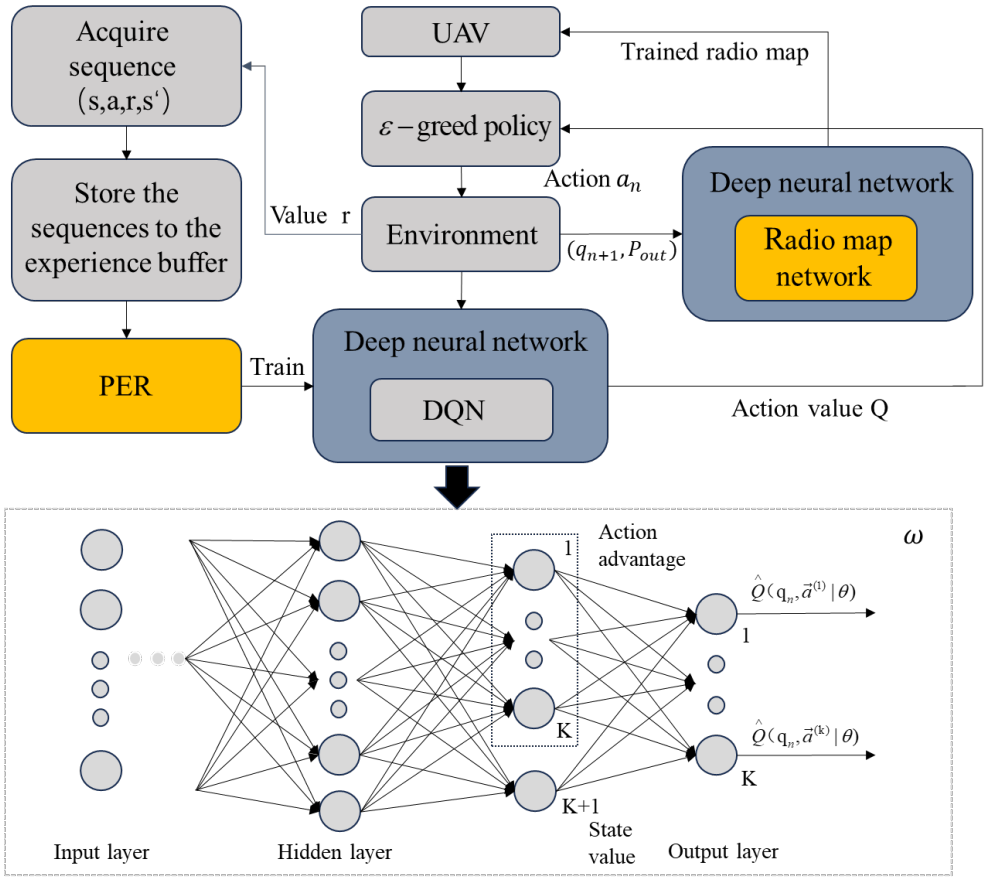


Figure 5. The framework of 3D path planning algorithm. PER mechanism and radio map network are utilized to assist the DQN network in learning Q-values. The specific algorithm structure of dueling DQN is presented in the dotted box.

4. Simulation Verification and Analysis

In order to validate the efficacy of the proposed approach, this section conducts simulations on radio mapping and path planning based on the enhanced DRL algorithm. Furthermore, we conduct a comparative analysis between 2D and 3D trajectories in the path

planning simulation to substantiate the indispensability of incorporating 3D path planning for UAVs under connection constraints. In simulations, each GBS has a transmitting power $P_m = 20$ dBm, with an interrupt SIR threshold set at $\gamma_{th} = 0$ dB. Other simulation parameters are presented in Table 1.

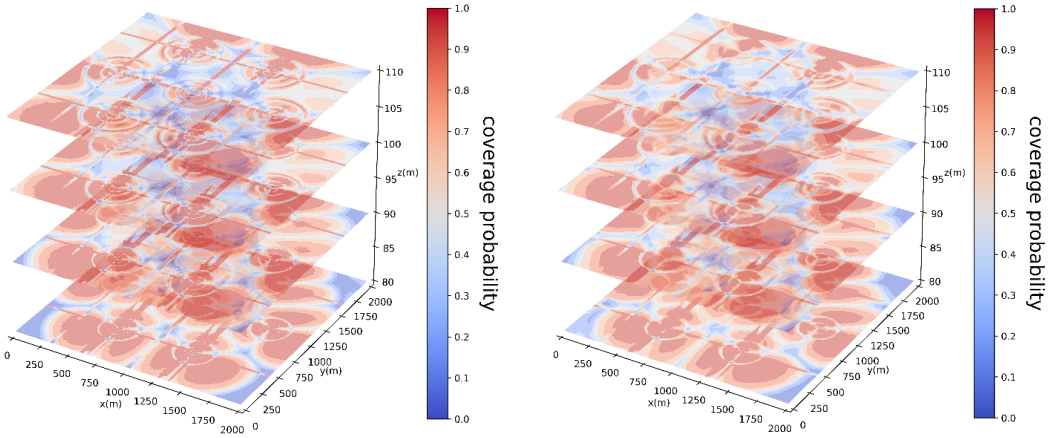
Table 1. Main simulation parameters.

Simulation Parameter	Description	Value
N_{epi}	Training episodes	10,000
N_{step}	Maximum steps per episode	400
N_1	Multi-step bootstrapping size	30
C	The capacity of experience pool	16,000
ε_0	Initial exploration rate	0.5
α	Decay rate of exploration	0.999
γ	Discount factor	0.9999
B	The number of experience extraction	32
δ	Update interval steps for target network	5
J	Signal measurement frequency	1000
R_{out}	Outbound reward	−10,000
R_{des}	Finish reward	500
V	The velocity of the UAV	10 m/s
D_{tol}	Sensing distance	20 m
Δt	Time interval	1 s
h_{max}	Maximum flight altitude	110 m
h_{min}	Minimun flight altitude	80 m
h_{bs}	The height of GBSs	25 m

4.1. Radio Mapping Based Environmental Learning

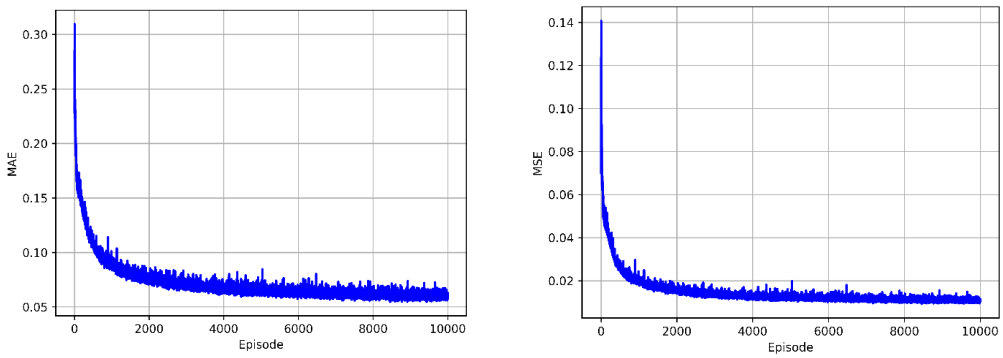
For radio mapping, we employ artificial neural networks (ANN) for map learning, which are trained using Adam optimizers to minimize mean square error (MSE) losses. The radio map network comprises five hidden layers with 512, 256, 128, 64, and 32 neurons respectively. The input consists of the UAV's 3D coordinates q_n , while the output represents the predicted probability of interruption $\hat{P}_{out}(q(n))$ at that location. The objective of network learning is to accurately align the 3D radio map with the real environment, thereby providing precise interconnection probabilities for each spatial point during simulated flight and enhancing the accuracy of the multi-step D3QN algorithm.

The actual global coverage of the 3D region under consideration is depicted in Figure 6a, which is obtained through numerical simulations using a computer based on the aforementioned model of the 3D environment and channel. Therefore, direct utilization of this simulated data in the algorithm is not feasible. As shown in Figure 6a, due to the combined influence of GBS antenna inclination and building occlusion, the coverage map exhibits irregularities in high altitude areas, while low altitude areas demonstrate a more regular pattern. Figure 6b illustrates the spatial 3D coverage probability map acquired from radio mapping. By comparing Figure 6a with Figure 6b, it is evident that the acquired CPG exhibits a remarkable alignment with its corresponding actual counterpart, thereby substantiating the effective application of SNARM in path learning. The algorithm effectiveness was further verified through simulations, which evaluated the MSE and mean absolute error (MAE) of the learned radio map in relation to the episode count, as depicted in Figure 7. MSE and MAE are derived by comparing predicted outage probabilities obtained from radio network measurements against actual outage probabilities at randomly selected locations. Episodes ranging from 0 to 500 correspond to an initial learning stage where large MSE and MAE values indicate poor quality of initially learned radio maps. However, as episode count increases, accumulating more signal measurement data leads to gradual decline in both MSE and MAE values, indicating improved approximation between learned radio maps and real maps.



(**a**) actual coverage map (**b**) learned coverage map

Figure 6. The diagram of 3D space coverage probability.



(**a**) MAE (**b**) MSE

Figure 7. The MSE and MAE of radio mapping versus episode number.

4.2. UAV Path Planning

The UAV operates within a 3D environment established in Section 2.1. It is assumed that the UAV's starting point is randomly generated, while its endpoint is located at coordinates [1400,1600,100] and labelled by the big blue triangle in simulated graph. When the UAV reaches the specified altitude and the distance between it and the endpoint satisfies condition $d \leq D_{tol}$, it is considered to have reached the endpoint.

The multi-step D3QN network consists of five hidden layers, with 512, 256, 128, 128, and 11 neurons respectively. The last hidden layer consists of one neuron representing the estimated state value, while the remaining ten neurons represent action advantages. These action advantages capture the discrepancy between each state's action value and its corresponding state value. By aggregating these differences in the output layer, we obtain ten estimates for action values. The objective of multi-step D3QN network learning is to accurately estimate the Q value for each action, enabling the UAV to determine an optimal flight strategy that minimizes the cumulative flight time and interruption-weighted time.

Authors in the paper [27] thoroughly investigate the 2D trajectory planning of UAVs subject to connectivity constraints. Nonetheless, overlooking the 3D attributes of the environment and neglecting the vertical movement of UAVs may lead to missed opportunities for optimal connectivity points and improved communication pathways. In this section, we conducted simulations of both 3D and 2D trajectories for UAVs, as outlined below

Comparing the 2D and 3D motion trajectories of the UAVs in Figure 8, it is evident that UAV prioritizes descending during 3D motion to seek better communication conditions. When the UAV is at the lowest altitude of 80 m, the convergence of its trajectory is lower compared to the 2D trajectory due to weakened spatial connectivity constraints. Nevertheless, upon comparing the weighted time of 2D and 3D trajectories from the same starting

point in Figure 9, it is observed that the UAV's flight process with 3D motion has lower weighted time. Even at starting point 6, its weighted time is only half of that of the 2D trajectory. Therefore, it can be inferred that within the confines of connectivity limitations, the superiority of UAVs' 3D motion compared to 2D motion becomes apparent.

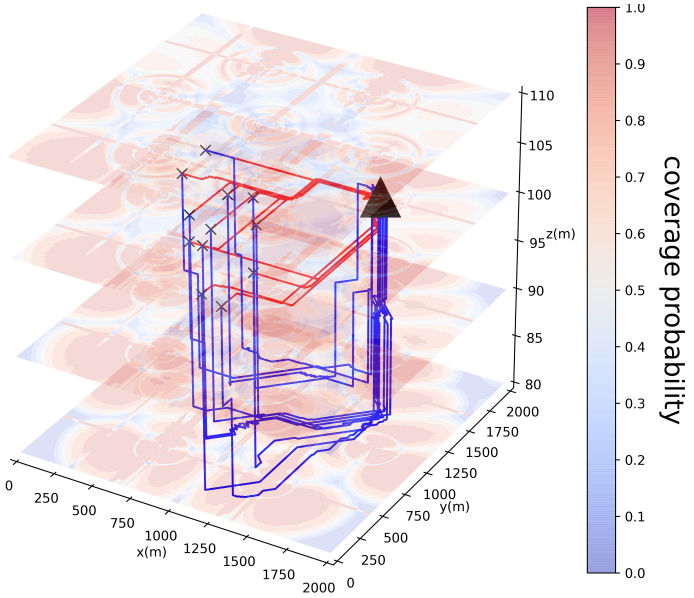


Figure 8. Comparison diagram of 2D and 3D trajectory. The red route represents the 2D trajectory, while the blue route represents the 3D trajectory. The starting height of the 12-episode route is set at 100 m and the weigh coefficient is set as $\mu = 40$.

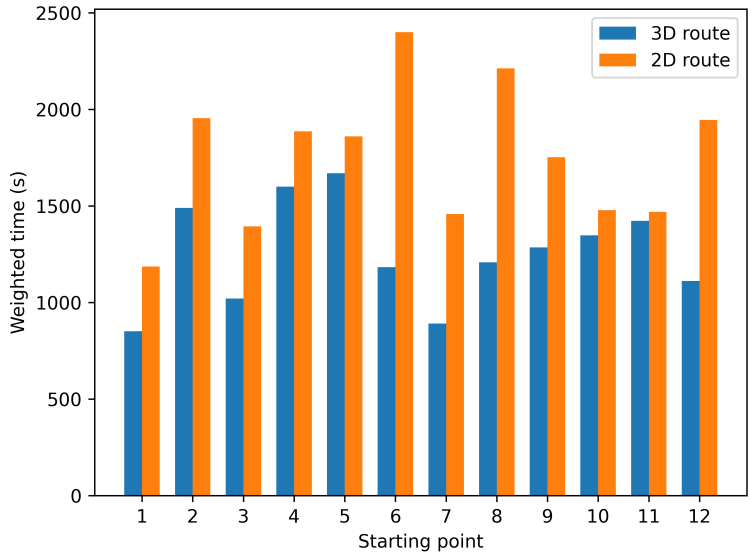
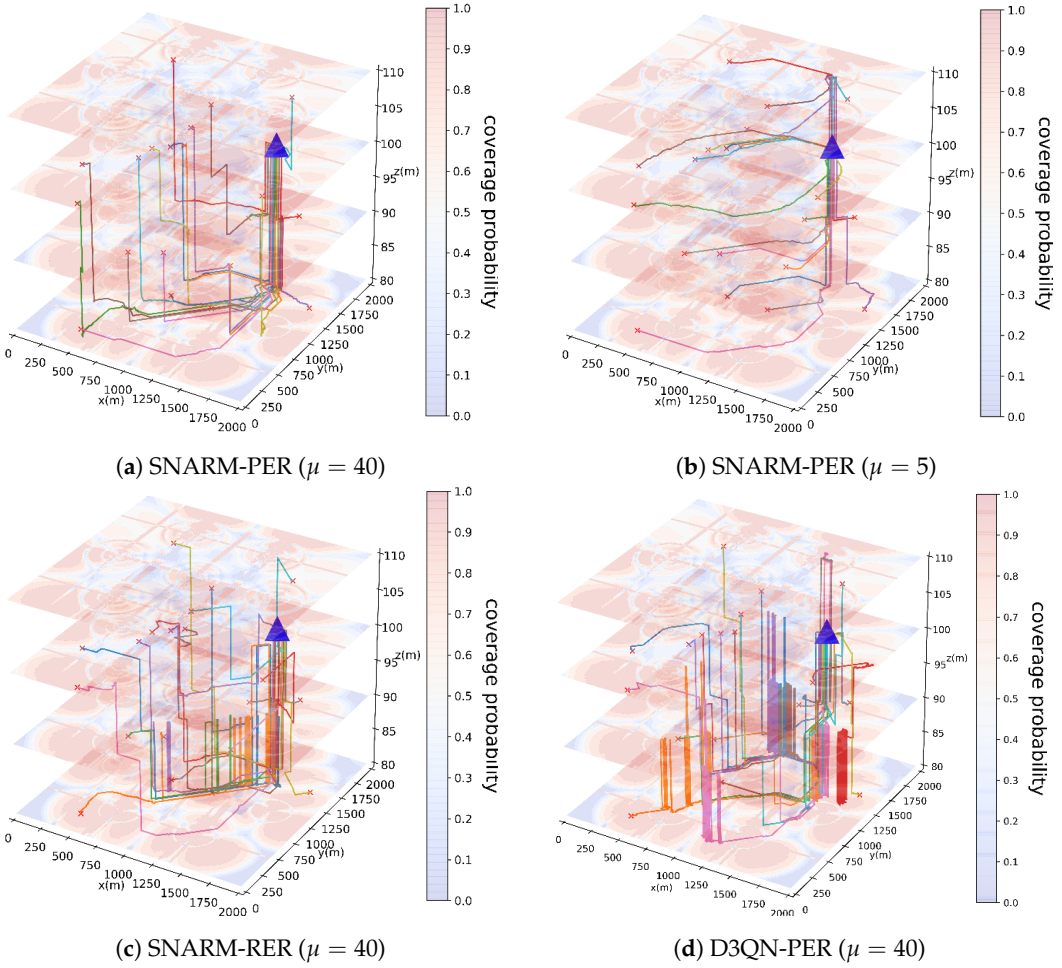


Figure 9. Temporal comparison diagram between 2D and 3D trajectories.

To further emphasize the merits of the proposed SNARM-PER technique in path planning, a comparative analysis is conducted with other approaches such as SNARM-RER [27] and D3QN-PER [34].

Figure 10 illustrates the final 20 episodes of UAV flight paths under different simulation conditions. Among them, the SNARM-PER algorithm integrates a multi-step D3QN algorithm with a radio map network and a PER mechanism, while the SNARM-RER and D3QN-PER algorithms serve as comparative algorithms, incorporating a multi-step D3QN algorithm with a radio map network and a RER mechanism, and utilizing a PER mechanism without a radio map network, respectively. In Figure 10a,b depict flight trajectory

maps using the target algorithm with different weight values. It can be observed that when the weight value is sufficiently large, UAVs tend to prioritize avoiding areas with weak communication coverage by descending to seek better communication conditions. Conversely, when the weight value is small, UAVs tend to follow more direct paths towards the destination with less consideration given to communication connectivity. This demonstrates the influence of weight coefficients μ in the objective function (20) on UAV flight paths. Specifically, a higher weight coefficient directs the UAV's focus more towards maintaining connectivity with the base station, consequently diminishing its emphasis on seeking the shortest route.



(**a**) SNARM-PER ($\mu = 40$) (**b**) SNARM-PER ($\mu = 5$)

(**c**) SNARM-RER ($\mu = 40$) (**d**) D3QN-PER ($\mu = 40$)

Figure 10. The diagram of UAV flight trajectory. These trajectories are all derived from the final training set of 20 episodes.

Following the principle of controlling variables, we compare (a) with (c) and (d) in the Figure 10. Under the same weighting coefficients, UAVs exhibit significant differences in their trajectories. it is evident that UAVs using the proposed SNARM-PER algorithm exhibit more convergent flight paths, allowing for precise avoidance of communication weak coverage areas, and completion of flight missions over shorter distances. However, UAVs using the comparative algorithms, due to insufficient learning of their Q-networks, show only partial convergence in their flight paths, along with oscillations in trajectory altitude.

Figure 11 illustrates how the average return of the UAV flight path changes with the number of episodes across various algorithms. The average return of the path is calculated as the mean value of the returns from the previous 200 episodes, thereby introducing data smoothing and enhancing trend visibility through averaging within a moving window. The average mobile return serves as a pivotal index for assessing the overall efficacy of UAV movement processes. Incorporating the settings of return values as outlined in

Section 3.4, a higher average movement return indicates lower cumulative flight and interruption times. A clear observation from Figure 11 is that, in the learning phase after 1000 episodes, UAVs leveraging the SNARM-PER algorithm, as proposed in this paper, exhibit notably superior average movement returns and enhanced motion performance compared to the contrasting algorithms. Figure 12 illustrates the total mission time of different algorithms during the last 20 episodes, representing a weighted sum of actual flight time and estimated interruption time. The weighted time of the last 20 episodes serves as an indicator of learning outcomes. As depicted in Figure 12, compared to SNARM-RER and direct-line approaches, the SNARM-PER algorithm excels in minimizing the weighted sum of UAV flight time and interruption time. Consequently, the UAV achieves a better balance between flight energy conservation and the avoidance of areas with weak communication coverage. All these findings serve to numerically validate the superiority of the proposed algorithm over other comparative methods. The UAV employing the SNARM-PER algorithm demonstrates enhanced capability in path planning under connectivity constraints while minimizing the weighted sum of flight time and interruption time.

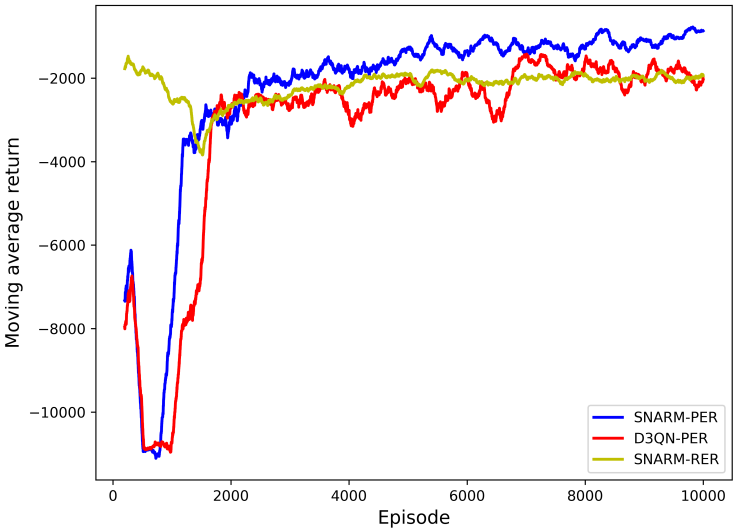


Figure 11. Moving average return.

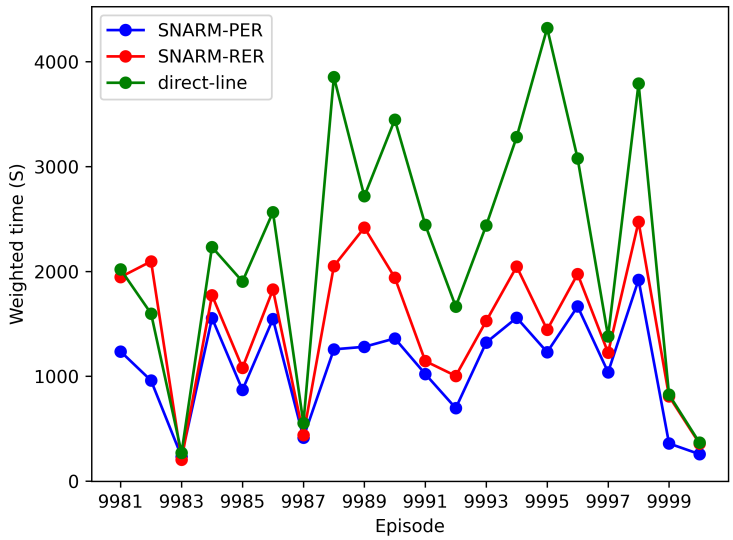


Figure 12. Weighted time of the last 20 episodes.

5. Conclusions

The quality of communication between UAV-BS in cellular network assisted UAV communication can be enhanced by strategically planning the 3D path of UAVs. Based on

this, we propose an improved DRL approach integrated with a radio prediction network for efficient 3D path planning of UAVs. Building upon the multi-step D3QN model, this method replaces conventional RER with PER and leverages the Dyna framework to combine real flight data with simulated flights under the radio prediction network, resulting in significant cost savings and improved algorithm performance. However, it should be noted that this method is only applicable to scenarios where the movement space of UAVs is discrete and there are no obstacles or no-fly zones at high altitudes. Therefore, future research should focus on comprehensive investigation into continuous motion space for UAV path planning and 3D obstacle avoidance.

Author Contributions: Conceptualization, X.L. and W.Z.; methodology, X.L.; software, X.L.; validation, X.L., W.Z. and X.W.; formal analysis, X.L. and Z.F.; investigation, X.L. and H.J.; resources, X.L. and Y.H.; data curation, X.L. and Z.L.; writing—original draft preparation, X.L.; writing—review and editing, X.L. and W.Z.; visualization, X.L. and H.D.; supervision, X.L. and W.Z.; project administration, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was co-supported by the Key Technologies R&D Program of Jiangsu (Prospective and Key Technologies for Industry) under Grants (No.BE2022067, BE2022067-1, BE2022067-2 and BE2022067-3) and the Key R&D Plan of Jiangsu Province under Grant BE2021013-4.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We would like to express our sincere thanks to all the editors, reviewers and staff who participated in the review of this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, J.; Zhu, Q.; Lin, Z.; Wu, Q.; Huang, Y.; Cai, X.; Zhong, W.; Zhao, Y. Sparse Bayesian learning-based 3D radio environment map construction—Sampling optimization, scenario-dependent dictionary construction and sparse recovery. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *10*, 80–93. [CrossRef]
2. Huang, Y.; Cui, H.; Hou, Y.; Hao, C.; Wang, W.; Zhu, Q.; Li, J.; Wu, Q.; Wang, J. Space-Based Electromagnetic Spectrum Sensing and Situation Awareness. *Space: Sci. Technol.* **2024**, *4*, 109. [CrossRef]
3. Zeng, Y.; Wu, Q.; Zhang, R. Accessing from the sky: A tutorial on UAV communications for 5G and beyond. *Proc. IEEE* **2019**, *107*, 2327–2375. [CrossRef]
4. Mondal, B.; Thomas, T.A.; Visotsky, E.; Vook, F.W.; Ghosh, A.; Nam, Y.H.; Li, Y.; Zhang, J.; Zhang, M.; Luo, Q.; et al. 3D channel model in 3GPP. *IEEE Commun. Mag.* **2015**, *53*, 16–23. [CrossRef]
5. Lin, Y.; Na, Z.; Feng, Z.; Lin, B.; Lin, Y. Dual-game based UAV swarm obstacle avoidance algorithm in multi-narrow type obstacle scenarios. *EURASIP J. Adv. Signal Process.* **2023**, *2023*, 118. [CrossRef]
6. Mao, K.; Zhu, Q.; Qiu, Y.; Liu, X.; Song, M.; Fan, W.; Kokkeler, A.B.J.; Miao, Y. A UAV-aided real-time channel sounder for highly dynamic non-stationary A2G scenarios. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 6504515. [CrossRef]
7. Lyu, Y.; Wang, W.; Sun, Y.; Rashdan, I. Measurement-based fading characteristics analysis and modeling of UAV to vehicles channel. *Veh. Commun.* **2024**, *45*, 100707. [CrossRef]
8. Lyu, Y.; Wang, W.; Sun, Y.; Yue, H.; Chai, J. Low Altitude UAV Air-to-Ground Multi-Link Channel Modeling and Analysis at 2.4 and 5.9 GHz. *IEEE Antennas Wirel. Propag. Lett.* **2023**, *22*, 2135–2139. [CrossRef]
9. Lyu, J.; Zhang, R. Network-connected UAV: 3-D system modeling and coverage performance analysis. *IEEE Internet Things J.* **2019**, *6*, 7048–7060. [CrossRef]
10. Qiu, J.; Lyu, J.; Fu, L. Placement optimization of aerial base stations with deep reinforcement learning. In Proceedings of the ICC 2020-2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
11. Amer, R.; Saad, W.; Galkin, B.; Marchetti, N. Performance analysis of mobile cellular-connected drones under practical antenna configurations. In Proceedings of the ICC 2020-2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–7.
12. Chowdhury, M.M.U.; Saad, W.; Güvenç, I. Mobility management for cellular-connected UAVs: A learning-based approach. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
13. Liu, L.; Zhang, S.; Zhang, R. Multi-beam UAV communication in cellular uplink: Cooperative interference cancellation and sum-rate maximization. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 4679–4691. [CrossRef]
14. Mei, W.; Zhang, R. Uplink cooperative NOMA for cellular-connected UAV. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 644–656. [CrossRef]

15. Rahmati, A.; He, X.; Guvenc, I.; Dai, H. Dynamic mobility-aware interference avoidance for aerial base stations in cognitive radio networks. In Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications, Paris, France, 29 April–2 May 2019; pp. 595–603.
16. Nguyen, H.C.; Amorim, R.; Wigard, J.; Kovács, I.Z.; Sørensen, T.B.; Mogensen, P.E. How to ensure reliable connectivity for aerial vehicles over cellular networks. *IEEE Access* **2018**, *6*, 12304–12317. [CrossRef]
17. Na, Z.; Liu, Y.; Shi, J.; Liu, C.; Gao, Z. UAV-supported clustered NOMA for 6G-enabled Internet of Things: Trajectory planning and resource allocation. *IEEE Internet Things J.* **2020**, *8*, 15041–15048. [CrossRef]
18. Na, Z.; Ji, C.; Lin, B.; Zhang, N. Joint optimization of trajectory and resource allocation in secure UAV relaying communications for Internet of Things. *IEEE Internet Things J.* **2022**, *9*, 16284–16296. [CrossRef]
19. Zhang, S.; Zeng, Y.; Zhang, R. Cellular-enabled UAV communication: A connectivity-constrained trajectory optimization perspective. *IEEE Trans. Commun.* **2018**, *67*, 2580–2604. [CrossRef]
20. Zhang, S.; Zhang, R. Radio Map Based Path Planning for Cellular-Connected UAV. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
21. Zhang, S.; Zhang, R. Radio map-based 3D path planning for cellular-connected UAV. *IEEE Trans. Wirel. Commun.* **2020**, *20*, 1975–1989. [CrossRef]
22. Yin, F.; Gunnarsson, F. Distributed recursive Gaussian processes for RSS map applied to target tracking. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 492–503. [CrossRef]
23. Yang, H.; Zhang, J.; Song, S.H.; Lataief, K.B. Connectivity-aware UAV path planning with aerial coverage maps. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–6.
24. Khamidehi, B.; Sousa, E.S. Federated learning for cellular-connected UAVs: Radio mapping and path planning. In Proceedings of the GLOBECOM 2020-2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6.
25. Khamidehi, B.; Sousa, E.S. A double Q-learning approach for navigation of aerial vehicles with connectivity constraint. In Proceedings of the ICC 2020-2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
26. Bulut, E.; Guevenc, I. Trajectory optimization for cellular-connected UAVs with disconnectivity constraint. In Proceedings of the 2018 IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
27. Zeng, Y.; Xu, X.; Jin, S.; Zhang, R. Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 4205–4220. [CrossRef]
28. Luo, X.; Zhang, T.; Xu, W.; Fang, C.; Lu, T.; Zhou, J. Multi-Tier 3D Trajectory Planning for Cellular-Connected UAVs in Complex Urban Environments. *Symmetry* **2023**, *15*, 1628. [CrossRef]
29. Mughal, U.A.; Ahmad, I.; Pawase, C.J.; Chang, K. UAVs path planning by particle swarm optimization based on visual-SLAM algorithm. In *Intelligent Unmanned Air Vehicles Communications for Public Safety Networks*; Springer Nature: Singapore, 2022; pp. 169–197.
30. Pan, H.; Liu, Y.; Sun, G.; Fan, J.; Liang, S.; Yuen, C. Joint power and 3D trajectory optimization for UAV-enabled wireless powered communication networks with obstacles. *IEEE Trans. Commun.* **2023**, *71*, 2364–2380. [CrossRef]
31. Wang, S.; Qi, N.; Jiang, H.; Xiao, M.; Liu, H.; Jia, L.; Zhao, D. Trajectory Planning for UAV-Assisted Data Collection in IoT Network: A Double Deep Q Network Approach. *Electronics* **2024**, *13*, 1592. [CrossRef]
32. Gong, Q.; Wu, F.; Yang, D.; Xiao, L.; Liu, Z. 3D Radio Map Reconstruction and Trajectory Optimization for Cellular-Connected UAVs. *J. Commun. Inf. Netw.* **2023**, *8*, 357–368. [CrossRef]
33. Al-Hourani, A.; Kandeepan, S.; Lardner, S. Optimal LAP altitude for maximum coverage. *IEEE Wirel. Commun. Lett.* **2014**, *3*, 569–572. [CrossRef]
34. Zhong, W.; Wang, X.; Liu, X.; Lin, Z.; Ali, F. Joint optimization of UAV communication connectivity and obstacle avoidance in urban environments using a double-map approach. *EURASIP J. Adv. Signal Process.* **2024**, *2024*, 1–26. [CrossRef]
35. Zeng, Y.; Zhang, R. Energy-efficient UAV communication with trajectory optimization. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 3747–3760. [CrossRef]
36. Zeng, Y.; Xu, J.; Zhang, R. Energy minimization for wireless communication with rotary-wing UAV. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 2329–2345. [CrossRef]