Connectivity-Aware 3D UAV Path Design With Deep Reinforcement Learning

Hao Xie, Dingcheng Yang, *Member, IEEE*, Lin Xiao, *Member, IEEE*, and Jiangbin Lyu, *Member, IEEE*

***Abstract*—In this paper, we study the three-dimensional (3D) path planning problem for cellular-connected unmanned aerial vehicle (UAV) taking into account the impact of 3D antenna radiation patterns. The cellular-connected UAV has a mission to travel from an initial location to a destination. In this process, there is a trade-off between the flight time and the expected communication outages duration, which requires planning a suitable route to avoid the weak coverage region. However, as the existing cellular networks are primarily designed for ground coverage, the communication coverage in the sky tends to be intermittent and irregular due to the potential for severe interference and obstructions (such by buildings), which brings great challenges to 3D path planning. To address this challenge, we first construct a 3D coverage map that stores the expected outage probability over each locations. We then propose a multi-step dueling DDQN (multi-step D3QN) based algorithm to design the local optimal UAV path by leveraging the constructed coverage map. In this algorithm, the UAV acts as the agent to learn the appropriate action to complete the flight mission. Numerical results show the effectiveness of the proposed algorithm for connectivity-aware UAV path planning and the superiority of 3D path design over its 2D counterparts.**

***Index Terms*—unmanned aerial vehicle (UAV), path design, radio map, cellular network, deep reinforcement learning.**

I. Introduction

BENEFITING from their high mobility, agility and ease of deployment, unmanned aerial vehicles (UAVs) have found a wide range of applications, including but not limited to cargo delivery, traffic control, rescue and search, and virtual reality [1]. The global commercial UAV market is expected to grow at a compound annual growth rate of over 16 percent starting from 2021, and reach around 58.4 billion U.S. dollars by 2026 [2]. However, most of the existing UAVs usually operate within the visual line-of-sight (LoS) range, since the communication distance is severely limited by the controller or WiFi connection mode. Moreover, there are also problems of low data transmission rates, vulnerability to interference, and even legal risks due to unlicensed communication over the spectrum. The above mentioned issues limit the further development and applications of the UAV technology. Therefore, integrating UAV into the next generation cellular network, namely cellular-connected UAV [1], is a promising research direction. Firstly, thanks to the pervasive cellular infrastructure around the world, it provides cost-effective communication links and almost unlimited range of remote control. Secondly, cellular-connected UAV is expected to achieve significant performance improvements over the simple direct point-to-point wireless communications, thanks to its higher data transmission rates and lower latency enabled by the fifth generation (5G) and beyond mobile networks [3]. Thirdly, the navigation of traditional UAV is typically based on the GPS signal, which is often disrupted due to the influence of poor weather or obstacles, where the cellular signal can complement for the positioning accuracy at this time. Finally, it complies to the air traffic regulations through the configuration of cellular network access permission and background monitoring to eliminate illegal flight activities.

Despite the attractive application prospect, there are still many practical problems with cellular-connected UAV that need to be solved. In the conventional terrestrial cellular network, the ground base stations (GBS) antenna is planned to be downward towards the ground to provide services for more ground users [4], which is not designed for aerial users. Moreover, due to the possible building obstructions in the 3D space, weak coverage for the sky is inevitably caused. Cellular-connected UAV suffers serious co-channel interference, due to its higher probability of LoS channel between other non-associated ground base stations. In [5], the authors analyze the uplink/downlink 3D coverage performance and introduce a generalized Poisson multinomial (GPM) distribution to model the accurate interference information, and show the effect of different downtilt angle of the GBS antenna on the 3D coverage. The authors in [6] investigated the performance of cellular-connected UAV under practical antenna configurations and reveal the impacts of the number of antenna elements on coverage probability and handover rate. The downtilt angle of the GBS antenna is taken as the optimization variable to maximize the UAV's received signal quality while maintaining a good throughput performance of the ground users and reduce the handover times [7]. To deal with the strong aerial ground interference problem, some interference mitigation techniques have been proposed [8]–[11]. In [8], authors introduce a cooperative interference cancellation strategy for the co-channel interference, which exploits the existing backhaul links among the GBSs in the cellular-network. On the other

Manuscript received May 6, 2021; revised August 25, 2021 and October 13, 2021; accepted October 14, 2021. Date of publication October 29, 2021; date of current version December 17, 2021. This work was supported by the Natural Science Foundation of China under Grant 62061027. The review of this article was coordinated by Prof. Sukumar Kamalasadan. *(Corresponding author: Dingcheng Yang.)*

Hao Xie, Dingcheng Yang, and Lin Xiao are with the School of Information Engineering, Nanchang University, Nanchang 330031, China (e-mail: xiehao@email.ncu.edu.cn; yangdingcheng@ncu.edu.cn; xiaolin@ncu.edu.cn).

Jiangbin Lyu is with the School of Informatics, Xiamen University, Xiamen 361005, China, and also with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Xiamen University, Xiamen 361005, China (e-mail: ljb@xmu.edu.cn).

Digital Object Identifier 10.1109/TVT.2021.3121747

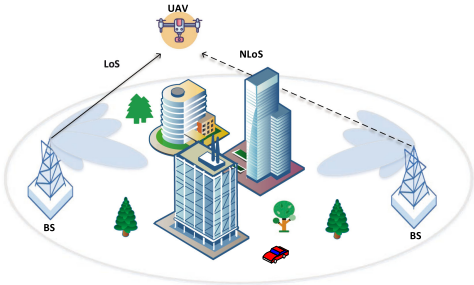


Fig. 1. Connectivity-aware path planning for cellular-connected UAV in urban.

hand, UAV cannot stay aloft too long as a battery-limited device. Communication and flight consume energy all the time [12], so how to save energy or improve energy efficiency as much as possible is a critical issue that needs to be considered. A measure of supplementing energy through ground charging stations was adopted in [13]. The authors in [14] investigate a mobile relay system that maximizes the end-to-end throughput via optimizing transmit power allocation strategy and trajectory of the UAV. The optimization technique is further extended to investigate the energy efficiency of UAV in [15]. In [16], the energy efficiency was maximized by jointly designing the UAV placement, hybrid precoding and power allocation.

An effective path planning should not only ensure satisfactory aerial ground communication conditions, increase the data transmission rate and connection reliability, but also reduce unnecessary movement of UAV and thus improving energy efficiency. It is worth noting that there has been some related work. Authors studied the shortest path planning problem constrained by minimum received signal-to-noise ratio (SNR) target [17]. In [18], [19], authors first established the channel gain map of GBS to provide the large-scale channel gains in 3D space, which is available due to the static and large-size buildings and thus assumed to be time-invariant, and then obtain the signal-to-interference-plus-noise ratio (SINR) map by combining loading factors. Based on the SINR map, the shortest path is designed under the constraint of the specified minimum SINR by leveraging graph theory, and several suboptimal solutions are proposed to reduce the computational complexity. A distributed recursive Gaussian process regression was proposed for constructing received-signal-strength map [20]. Authors used dynamic programming to design an optimal path that subject to a certain outage duration condition [21]. Similar problems have been studied in [22]–[24]. In addition, the recent work [25] considered the 2D path planning issue in a horizontal plane with the radio map. However, In order to explore the mobility benefits of UAV, the 3D path planning issue under the 3D antenna radiation pattern of the GBSs would be more realistic for practical scenarios. To this end, it motivates us to study the optimal 3D path planning of the cellular-connected UAV in the scene of realistic 3D antenna radiation pattern, as illustrated in Fig. 1. More specially, we consider a dedicated UAV that has a specific mission (such as delivery) from a given starting point to a destination, it needs to complete the delivery as quickly as possible while avoiding weak areas of coverage and minimizing expected outage duration during the flight.

The resulted 3D path optimization is difficult to be solved by traditional optimization methods, which typically assume simplified antenna pattern and channel model for analytical tractability. Most of the previous works only assume simplified isotropic radiation for antennas and free-space path loss channels. Although further more sophisticated Rician fading [26] and probabilistic LoS channels [27] have been proposed, these stochastic models only pay attention to the relative positions of communication nodes, and do not consider the environment-awareness attribute. In other words, no matter where the location of each obstacle and how complex the real environment is, communication nodes with the same relative positions have the same channel conditions (in the statistical sense). It's worth noting that this ensures communication performance in an average sense, but dose not reflect real channel realizations. The importance of the environmental-awareness attribute is elaborated in [28]. Worse still, the formulated problem is typically non-convex problem and very difficult to solve, whose complexity increases dramatically with the number of parameters to be optimized.

Fortunately, thanks to the rapid progress in the field of machine learning, deep reinforcement learning (DRL) as a branch of this field provides an alternative solution for such complex optimization problems. In DRL, UAVs act as agents to learn navigation strategies through feedback (empirical data) obtained through continuous interaction with the environment. This data-driven feature enables the learning process without or with little prior knowledge of the radio environment model.

In this paper, we consider a 3D path optimization problem, in which the UAV needs to learn excellent flight strategies in a given environment to shorten the time to reach the destination and improve the communication quality of the whole process. The main contributions of this paper are listed as follows:

- We formulate a 3D path optimization problem by considering the sum of weighted mission completion time and the expected communication outage duration over different 3D antenna radiation pattern of the GBSs, and propose an DRL-based algorithm to learn the navigation strategy.
- Since the formulated path optimization problem is nonconvex and hard to solve, we propose an algorithm based on multi-step D3QN for achieving the local optimal 3D path, which picks three efficient DQN extensions. The first one is double DQN (DDQN), which adopts two network structure to obtain more accurate Q function (action-value function) estimation. The second and third are dueling network architecture and multi-step bootstrapping respectively, which increase stability and accelerate the convergence speed of the algorithm.
- Numerical results demonstrates the effectiveness of the proposed multi-step D3QN based algorithm, and show the influence of different 3D antenna radiation pattern for the path planning. Moreover, we also compare the performance of obtained 3D path with 2D path and show that the 3D trajectory has better flight performance than its 2D counterpart by changing the altitude.

In fact, the application of DRL in UAV trajectory optimization has been developed in various contexts. A decaying deep Q-network (D-DQN) based algorithm in [29] is adopted to minimize energy consumption through the joint optimization of UAV movement, phase shifts of the reconfigurable intelligent surface (RIS), power allocation policy and dynamic decoding order. Authors in [30], [31] exploit Robust-DDPG and MEP-DDPG enable drones to avoid obstacles in a dynamic, complex and unknown environment. In [32], authors used DRL and a map centered on the UAV's position to seek tradeoffs between data collection, time efficiency and security constraints. [33] considered the situation of a joint service of ground users by multi UAVs, the authors assume that each ground user is moving irregularly and use Q-learning to complete the 3D mobile deployment of UAVs, so as to maximize the sum mean opinion score of ground users. Dynamic programming was used in [34] to explore the impact of 3D antenna pattern and backhaul constraint on 3D path of UAV. [35] maximized the real-time downlink capacity under certain coverage constraints by leveraging a constrained Deep Q-Network (cDQN). In [36], the enhanced multi-UAV Q-learning algorithm is proposed to deal with the trajectory design problem of multi-UAV with real-time sensing tasks. A decentralized DRL based framework is used in [37] to control the movement of each UAV in a distributed manner, with the purpose of maximizing the total number of users' coverage and ensuring the fairness of each user being served, as well as reducing the total energy consumption of all UAVs. In [38], the distance between UAVs is considered to avoid collision and reduce the interference of the neighbouring UAV-GBS to the ground users. An improvement was also implemented to improve the performance of system fairness and individual user coverage based on [37]. DRL is used to control the movement of UAV and extend the service time by replenish energy [39]. Different from the above work, we consider the typical cellular-connected UAV system. In this scenario, the proposed algorithm only utilizes the radio map measured offline to learn a navigation strategy, which is an offline design. In the actual flight process, we can uploads the measured data to update the accuracy of the radio map.

The rest of the paper is organized as follows. Section II presents the system model, 3D radio map construction and problem formulation. In Section III, some necessary concepts about DRL are introduced. Section IV describe the proposed algorithm. Simulation settings, experimental results and analysis are introduced in V. Finally, we conclude the paper in Section VI.

II. System Model and Problem Formulation

A. Scenario and UAV Mobility Model

We consider an dense urban area of $D \times D$ km^2, a large number of buildings (obstacles) and several GBSs are distributed in this region. For better simulating the GBS-UAV channel in the area of interest, we specify building locations and heights according to the suggestions of International Telecommunication Union (ITU) recommendation document [40], which involves three statistical parameters α_{bd}, β_{bd} and γ_{bd}. Parameter α_{bd} denotes the proportion of the buildings area to the total land area; parameter β_{bd} denotes the mean number of buildings per unit area (buildings/km^2); parameter γ_{bd} affects the height distribution of buildings according to the Rayleigh distribution with mean value σ_{bd}. The height of buildings was set not to exceed h_{bd} m.

The UAV mobility space is above all buildings, the 3D location of UAV at time t is given by $\mathbf{q}(\mathrm{t}) = (x_t, y_t, h_t) \in \mathbb{R}^{3\times 1}$, $t \in [0, T]$, where $x_t \in [0, D]$ and $y_t \in [0, D]$ denote the 2D X-coordinate and Y-coordinate. $h_t \in [h_{\min}, h_{\max}]$ denote height of the UAV, which should not be less than the lower bound h_{min} and no more than the upper bound h_{max}. $\mathbf{q}_\mathrm{s} = (x_s, y_s, h_s) \in \mathbb{R}^{3\times 1}$ and $\mathbf{q}_\mathrm{f} = (x_f, y_f, h_f) \in \mathbb{R}^{3\times 1}$ represents the initial position and the destination respectively. The UAV flies at a constant speed V meters/second (m/s) during the whole mission process.

B. Antenna Model

We set up the antenna radiation model of the GBS according to the standard of 3GPP [4]. Each GBS is divided into three sectors/cells, where a uniform linear array (ULA) of 8 elements is placed vertically and each of the element radiation pattern provide a high directional gain. Consider a number of M cells, denoted by the set $\mathcal{M} = \{1, 2, \ldots, M\}$. Assume that the GBSs have the same height h_{bs} m.

The element radiation pattern is composed of both horizontal and vertical radiation patterns [41]. The vertical and the horizontal radiation patterns are respectively given by

$$A_{E,V}(\theta) = -\min\left\{12\left(\frac{\theta - 90^\circ}{\theta_{3\,\mathrm{dB}}}\right)^2, SLA_V\right\} \tag{1}$$

$$A_{E,H}(\phi) = -\min\left\{12\left(\frac{\phi}{\phi_{3\,\mathrm{dB}}}\right)^2, A_m\right\} \tag{2}$$

where $\theta_{3\,\mathrm{dB}} = \phi_{3\,\mathrm{dB}} = 65^\circ$ are the half-power beamwidths in vertical and horizontal dimensions, respectively. $SLA_V = 30$ dB is the sidelobe level limit and $A_m = 30$ dB is the front–back ratio. Then we can obtain the 3D antenna element gain $A(\theta, \phi)$ for each pair of angles by combining the vertical and horizontal radiation pattern as

$$A_E(\theta, \phi) = G_{E,max} - \min\left\{-\left[A_{E,V}\left(\theta\right) + A_{E,H}\left(\phi\right)\right], A_m\right\} \tag{3}$$

where $G_{E,max} = 8$ dBi is the maximum directional gain of each antenna element in the main lobe direction.

Then we consider the entire antenna array radiation pattern $A_A(\theta, \phi)$, by the combination of a single element radiation pattern and the array factor [41], which can be obtained as

$$A_A(\theta, \phi) = A_E(\theta, \phi) + AF(\theta, \phi, n) \tag{4}$$

where the array factor $AF(\theta, \phi, n)$ with n antenna elements is defined as

$$AF(\theta, \phi, n) = 10\log_{10}\left[1 + \rho\left(|\mathbf{a}\cdot\mathbf{w}^T|^2 - 1\right)\right] \tag{5}$$

where ρ set as unity, is the correlation coefficient. $\mathbf{a} \in \mathbb{C}^n$ is the amplitude vector with all elements equal to constant $\frac{1}{\sqrt{\mathrm{n}}}$, which assume that each antenna element has the same amplitude. $\mathbf{w} \in \mathbb{C}^n$ is the beamforming vector, which control the steering

direction of the main lobe and is defined as

$$\begin{aligned}\mathbf{w} &= [w_{1,1}, w_{1,2}, \ldots, w_{m_V,m_H}], \quad m_V m_H = n,\\ w_{p,r} &= e^{j2\pi((p-1)\frac{d_V}{\lambda}\Psi_p + (r-1)\frac{d_H}{\lambda}\Psi_r)},\\ &\begin{cases}\Psi_p = \cos\theta - \cos\theta_s\\ \Psi_r = \sin\theta\sin\phi - \sin\theta_s\sin\phi_s\end{cases}\end{aligned} \tag{6}$$

where $d_V = d_H = \frac{\lambda}{2}$ are the vertical and horizontal antenna elements spacing distances of the antenna array respectively, set to half of the carrier frequency wavelength λ. The pair of angle (θ_s, ϕ_s) stand for the main lobe steering direction due to beamforming. For simplicity, the mutual coupling[1] effects is not considered in this paper. According to the above equations, we consider the electrical downtilt technology and steering the main lobe downtilted $\theta_{tilt} = 90^\circ - \theta_s$ ($\theta_{tilt} \leq 0$), and the beamforming weight vector $\mathbf{w}$ can be calculated. The expression in (4) provides the total antenna gain of UAV in different locations.

C. Path Loss Model

In this system model, we simulate the path loss model based on the urban micro (UMI) model in 3GPP specifications [42]. It's worth mentioning that the GBS-UAV path loss is divided into the LoS links and the non Line-of-Sight (NLoS) links. The path loss of the LoS link between UAV and cell m is given in dB as

$$\begin{aligned}h_m^{LoS}(t) = max\{&h_m^{FSPL}, 30.9\\ &+ (22.25 - 0.5\log_{10} h_t)\log_{10} d_m(t) + 20\log_{10} f_c\}\end{aligned} \tag{7}$$

where h_m^{FSPL} is the free-space path loss, h_t is the UAV's altitude at time t, $d_m(t)$ is the 3D distance between UAV and cell m, and f_c is the carrier frequency. The path loss of the NLoS links between UAV and cell m is given in dB as

$$\begin{aligned}h_m^{NLoS}(t) = max\{&h_m^{LoS}(t), 32.4\\ &+ (43.2 - 7.6\log_{10} h_t)\log_{10} d_m(t)\\ &+ 20\log_{10} f_c\}.\end{aligned} \tag{8}$$

Note that in the following process of creating the 3D coverage map, LoS or NLoS channel can be determined by judging whether the line segment is blocked by buildings between the UAV and cell.

Furthermore, in the case of LoS, the small-scale fading is considered as Rician fading with Rician factor K_R dB, while in the case of NLoS, the small-scale fading is Rayleigh fading.

D. Signal Model and 3D Coverage Map Construction

In this subsection, we introduce the signal model and the concept of expected outage probability, and then construct the 3D coverage map by leveraging the expected outage probability of the UAV at each location.

[1]Mutual coupling describes the antenna affected by another adjacent antenna in a complex way, and reducing the antenna radiation efficiency.

We denote $h_m(t)$ as the channel power gain between the UAV and cell m, which are mainly generated from the following parts: the GBS antenna gains, the large-scale channel power gain and the small-scale fading. Therefore, the instantaneous signal power that the UAV received from cell m can be expressed as

$$\begin{aligned}y_m &= P_m|h_m(t)|^2\\ &= P_m\beta(\mathbf{q}(t))\bar{h}_m(\mathbf{q}(t))\tilde{h}_m(t), \quad m \in \mathcal{M}\end{aligned} \tag{9}$$

where P_m denotes the transmit power of cell m, which is assumed to be constant. $\beta(\mathbf{q}(t))$ and $\bar{h}_m(\mathbf{q}(t))$ denote the GBS antenna gain and the large-scale channel power gain respectively, and they are functions of location $\mathbf{q}$(t). $A_A(\theta, \phi) = 10log_{10}(\beta(\mathbf{q}(t)))$, where we can obtain the elevation angle θ and the azimuth angle ϕ of UAV with respect to GBS if the UAV's coordinate $\mathbf{q}(t)$ is given. The large-scale channel power gain $\bar{h}_m(\mathbf{q}(t))$ is determined by the locations of buildings between the UAV and GBS, which is assumed to be time-invariant caused by static and large obstacle blocking, we have

$$\bar{h}_m(\mathbf{q}(t)) = \begin{cases}h_m^{LoS}(\mathbf{q}(t)), & \text{if LoS link,}\\ h_m^{NLoS}(\mathbf{q}(t)), & \text{if NLoS link,}\end{cases} \tag{10}$$

In contrast, the small-scale fading denoted by $\tilde{h}_m(t)$ is a random variable.

We denote $I(t) \in \mathcal{M}$ as the cell that provides service for the UAV at time t. Thus, we have the downlink instantaneous signal-to-interference-plus-noise ratio (SINR)

$$\gamma(t) = \frac{P_{I(t)}|h_{I(t)}(t)|^2}{\sum_{m \neq I(t)} P_m|h_m(t)|^2 + \sigma^2}, \tag{11}$$

Note that, $\gamma(t)$ is also a random variable for any given location $\mathbf{q}(t)$ and associated cell $I(t)$, due to the small-scale fading $\tilde{h}_m(t)$.

We use the outage probability to evaluate the reliability of GBS-UAV link, i.e.,

$$P_{out}(\mathbf{q}(t), I(t)) = \Pr\{\gamma(t) < \gamma_{th}\}. \tag{12}$$

where $\Pr\{\cdot\}$ is the probability of the occurrence of the event. The GBS-UAV connection is considered interrupted when $\gamma(t)$ is less than the outage threshold γ_{th}.

Next, for a given γ_{th}, the connection strategy for any $\mathbf{q}$(t) will be introduced, i.e., select the connected cell as $I(t)$. We rewrite the instantaneous SINR $\gamma(t)$ as $\gamma(\mathbf{q}(t), I(t), \tilde{h}_{I(t)})$, where $\tilde{h}_{I(t)}$ is the small-scale fading between the UAV and cell $I(t)$. Then the outage indication function $c(t)$ is defined as

$$c(\mathbf{q}(t), I(t), \tilde{h}_{I(t)}) = \begin{cases}1, & \gamma(\mathbf{q}(t), I(t), \tilde{h}_{I(t)}) < \gamma_{th}\\ 0, & \text{otherwise.}\end{cases} \tag{13}$$

Therefore, the outage probability function in (12) can be expressed as the expectation of $\tilde{h}_{I(t)}$

$$\begin{aligned}P_{out}(\mathbf{q}(t), I(t)) &= \Pr\left\{\gamma(\mathbf{q}(t), I(t), \tilde{h}_{I(t)}) < \gamma_{th}\right\}\\ &= \mathbb{E}_{\tilde{h}_{I(t)}}\left[c(\mathbf{q}(t), I(t), \tilde{h}_{I(t)})\right]\end{aligned} \tag{14}$$

To obtain the outage probability of each time t, we continuously measure the SINR J times for each cell in a very short period of time, which could be completed by querying continuous RSRP and RSRQ reports in practical application. The j-th measurement of the small-scale fading is denoted as $\tilde{h}_{I(t)}(t,j)$, the corresponding SINR and the outage indication function are denoted as $\gamma(\mathbf{q}(t), I(t), \tilde{h}_{I(t)}(t,j))$ and $c(\mathbf{q}(t), I(t), \tilde{h}_{I(t)}(t,j))$. From this we can derive the empirical outage probability as

$$\hat{P}_{out}(\mathbf{q}(t), I(t)) = \frac{1}{J}\sum_{j=1}^{J} c(\mathbf{q}(t), I(t), \tilde{h}_{I(t)}(t,j)). \tag{15}$$

According to the law of large numbers, the real outage probability can be replaced by the empirical outage probability when J is large enough, which is expressed as

$$\lim_{J\to\infty} \hat{P}_{out}(\mathbf{q}(t), I(t)) = P_{out}(\mathbf{q}(t), I(t)). \tag{16}$$

Then it's easy to obtain the optimal connection strategy selected by the UAV at the $\mathbf{q}(t)$, namely the connected cell $I(t)$, is the one with the minimum empirical outage probability among M cells, which can be expressed as

$$I(t) = \arg\min_{m\in\mathcal{M}} \hat{P}_{out}(\mathbf{q}(t), m), \tag{17}$$

Then the $\hat{P}_{out}(\mathbf{q}(t), m)$ can be reduced to

$$\hat{P}_{out}(\mathbf{q}(t)) = \min_{m\in M} \hat{P}_{out}(\mathbf{q}(t), m). \tag{18}$$

For simplicity and the purpose of exposition, we focus on the basic cell association criterion in this paper, which is suitable for applications where connectivity and communication rate are more important or low mobility scenarios where there is enough time for handover, and leave extended investigation on handover issues for our future work.

Based on the above analysis, the expected outage probability of the UAV at any location can be obtained, whereby we can construct the coverage map over the 3D space, which can be measured offline by dedicated UAVs. The constructed coverage probability map will be shown in section V, where coverage probability = 1 - outage probability.

E. Problem Formulation

Based on the obtained outage probability map, this paper considers the following two targets by controlling the 3D trajectory $\{\mathbf{q}(t)\}$ of the UAV:

1) minimizing the flying time T from $\mathbf{q}_\mathrm{s}$ to $\mathbf{q}_\mathrm{f}$.

2) minimizing the expected outage duration.

Therefore, the optimization problem can be formulated as:

$$\max_{T,\{\mathbf{q}(t)\}} \quad -T - \int_0^T \hat{P}_{out}(\mathbf{q}(t))dt \tag{19a}$$

$$s.t. \quad \mathbf{q}(0) = \mathbf{q}_\mathrm{s}, \tag{19b}$$

$$\mathbf{q}(T) = \mathbf{q}_\mathrm{f}, \tag{19c}$$

$$\|\dot{\mathbf{q}}(t)\| = V, \quad \forall t \in [0,T] \tag{19d}$$

$$0 \le x_t \le D, \quad \forall t \in [0,T] \tag{19e}$$

$$0 \le y_t \le D, \quad \forall t \in [0,T] \tag{19f}$$

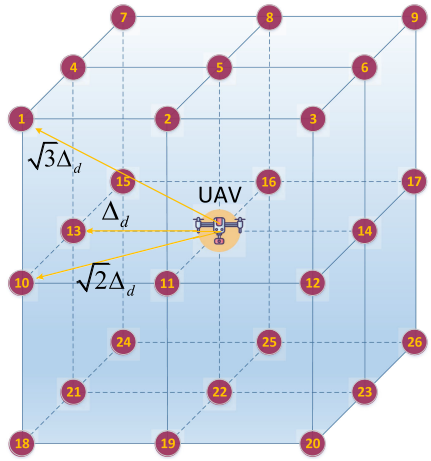


Fig. 2. Illustration of available 3D grid point and index.

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

where (19b) and (19c) represent the start and end location constraints. (19d) denotes the speed of the UAV. (19e) and (19f) limits the 2D boundary of the UAV. (19g) is the altitude range constraint. For the first goal, the optimal strategy is to fly in a straight line from a fixed initial location and destination. However, this approach may pass through the areas with weak coverage and resulting in significant outage duration. For the second goal, it's required to bypass areas with high outage probability as much as possible, which would increase the mission completion duration.

Considering the complexity of the nonconvex problem, it's quite challenging to solve. Fortunately, the 3D path planning problem is a Markov decision process (MDP) and can be conveniently solved by leveraging DRL. However, (19) is a continuous optimization problem, and the continuous state space and action space tend to increase the complexity sharply and thus lead to divergence. Therefore, we convert (19) to the following approximate discrete path planning problem over the 3D grid point:

$$\max_{N,\{\mathbf{q}_n\}_{n=0}^{N}} \quad -\sum_{n=0}^{N-1}\left(\|\mathbf{q}_{n+1}-\mathbf{q}_n\| + \hat{P}_{out}(\mathbf{q}_{n+1})\right) \tag{20a}$$

$$s.t. \quad \mathbf{q}_0 = \mathbf{q}_\mathrm{s}, \tag{20b}$$

$$\mathbf{q}_N = \mathbf{q}_\mathrm{f}, \tag{20c}$$

$$\|\mathbf{q}_{n+1}-\mathbf{q}_n\| \le \sqrt{3}\Delta_d, \quad \forall n \tag{20d}$$

$$0 \le x_n \le D, \quad \forall n \tag{20e}$$

$$0 \le y_n \le D, \quad \forall n \tag{20f}$$

$$h_{min} \le h_n \le h_{max}, \quad \forall n \tag{20g}$$

where N represents the number of grid point that the UAV passes through. we assume that the 3D movement space of UAV is divided into a series of adjacent grid points as shown in Fig. 2, the maximum distance of each adjacent grid point is not more than $\sqrt{3}\Delta_d$. Assume that $\mathbf{q}_s$ and $\mathbf{q}_f$ are both located on the grid points. When the grid granularity Δ_d is small enough,

we aim to minimize the sum of distances between each pair of adjacent grid points and the expected outage probability. In the following sections, we will introduce the proposed 3D path planning algorithm based on the multi-step D3QN to solve this path optimization problem.

III. Preliminries

In the section, we briefly introduce the relevant knowledge of DRL. For a more comprehensive description, please refer to [43].

RL has been well applied in solving MDP problems. In the RL framework, the subject of learning and decision making is called agent, and all the things that can interact with it are called environment. At each discrete moment, the agent will interact with the environment to produce the following sequence:

$$s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_2, r_3, \dots \tag{21}$$

where s, a, r represent the state of agent, the action and the reward from environmental feedback, respectively. For the MDP, s_{n+1}, r_{n+1} are only related to s_n, a_n. The task of the RL agent is to maximize the cumulative sum of all rewards G_n by selecting a series of actions from time slot n, which can be defined as:

$$G_n = r_{n+1} + \gamma r_{n+2} + \gamma^2 r_{n+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{n+k+1} \tag{22}$$

where $\gamma \in [0, 1]$ is a discount factor, the larger γ, the more long-term returns are considered; conversely, the less long-term return are considered.

The policy function $\pi(s)$ controls the agent's action by giving the probability of choosing different actions for each state. Obviously, the best action should be selected for each state to maximize the accumulated reward, and thus the optimal policy $\pi_*(s)$ is found. Another important concept is the action-value function $Q_\pi(s, a) = \mathbb{E}_\pi[G_n | s_n = s, a_n = a]$, which is the expected return that can be obtained by following the policy $\pi(s)$ after taking action a for state s. The optimal action-value function $Q_*(s, a) = \max_\pi Q_\pi(s, a)$ satisfies the Bellman optimality equation

$$Q_*(s, a) = r(s, a) + \gamma \sum_{s'} p(s'|s, a) \max_{a'} \mathrm{Q}_*(\mathrm{s},' \mathrm{a}'). \tag{23}$$

Theoretically, the optimal $Q_*(s, a)$ can be obtained by traversing all transitons $(s_n, a_n, r_{n+1}, s_{n+1})$ and constant iteration, and then the optimal strategy $\pi_*(s)$ can be found. In order to overcome the situation that Q learning is not applicable in continuous and high-dimensional state space or action space, the classical DQN utilize deep neural network (DNN) as the function approximator instead of Q-table and updates network parameters by minimizing the loss function:

$$(r_{n+1} + \gamma \max_a Q(s_{n+1}, a|\theta) - Q(s_n, a_n|\theta))^2 \tag{24}$$

After the proposal of DQN, a series of extensions to improve its performance have also been proposed. By utilize two neural networks with the same structure, nature DQN [44] can suppress the disadvantage that DQN is difficult to convergence due to the strong correlation caused by only one neural network. Double DQN (DDQN) [45] solves the overestimation problem inherent in the Q-learning. In addition, DQN with the dueling structure neural network [46] and the multi-step bootstrapping method [43] can improve the convergence efficiency, which will be described in detail in Section IV.

IV. Proposed Algorithm for 3D Path Planning

In the DRL model, the UAV as an agent, selects the optimal action based on the current state, obtains the reward from the 3D coverage map and then transfers to the next state. The state space, action space and reward function are described in detail as follows:

1) *State:* The state is expressed as $s_n = (x_n, y_n, h_n) \in \mathcal{S}$, which represents the location information of the n-th grid point where the UAV is located.
2) *Action:* The action space $\mathcal{A}$ consists of 26 directions, as shown in the Fig. 2, the UAV can selects action a_n to fly to any adjacent grid point. If the UAV flies out of the specified boundary, it will be transferred back to the state before flying out.
3) *Reward:* The reward r_{n+1} is obtained when the action a_n is performed in the state s_n. We set the reward function as:

$$r_{n+1} = \begin{cases} r_a, & \text{if arrived at the destination area;} \\ r_b, & \text{if out of the specified boundary;} \\ \mu_1 \|\mathbf{q}_{n+1} - \mathbf{q}_n\| + \mu_2 \hat{P}_{out}(\mathbf{q}_{n+1}) & \\ +\mu_3 (d_{pre} - d_{cur}), & \text{otherwise;} \end{cases} \tag{25}$$

where $d_{pre} = \|\mathbf{q}_n - \mathbf{q}_f\|$ and $d_{cur} = \|\mathbf{q}_{n+1} - \mathbf{q}_f\|$ represent the previous and current relative distance between the UAV and the destination, respectively, and these two forces UAV towards the destination area. μ_1, μ_2 and μ_3 are the weighting factor.

A. Three Effective Components

The proposed algorithm uses three extensions of DQN that each have addressed a limitation and helps to improve the overall performance.

Double DQN. The DQN is known to have the problem of overestimation and the reason can be found in the above (24), where the max operating selects the action with the maximum evaluated value to update. There is a bias between the evaluated action value and the real which thus degrades the learning performance. To solve this problem, [45] proposes the DDQN algorithm on the basis of DQN, whereby the loss function is changed to

$$(r_{n+1} + \gamma Q(s_{n+1}, \arg\max_{a'} Q(s_{n+1}, a'|\theta)|\theta') - Q(s_n, a_n|\theta))^2, \tag{26}$$

which obtains the state-of-the-art performance by changing the structure of the loss function.

Dueling networks. The dueling networks is a structural design of neural networks, which combines two stream to evaluate the action-value $Q(s, a)$. The first is value stream, which is only

Algorithm 1: N_1-Step D3QN for Connectivity-Aware UAV 3D Path Planning.

Initialization: the replay buffer with size B, the initial exploration probability ϵ, the exploration decay rate α_ϵ and the maximum number of steps per episode N_{max}.
Initialization: the evaluation network with parameters θ and the target network with parameters θ'.
Algorithm:
1: **for** episode $= 1, 2, \ldots, K$ **do**
2: $n \leftarrow 0$
3: Randomly initialize the state s_0;
4: **while** $n < N$ and $n < N_{max}$ **do**
5: Select action $a_n \in \mathcal{A}$ according to ϵ-greedy policy;
6: Execute a_n and transfer to next state s_{n+1};
7: Measure the outage probability $P_{out}(s_{n+1})$ from 3D coverage map;
8: Obtain reward r_{n+1} according to Eqn. (25);
9: $\tau \leftarrow n - N_1 + 1$.
10: **if** $\tau \geq 0$ **then**
11: Calculate the N_1-Step reward $R_{\tau:min(\tau+N_1,N)}$ according to Eqn. (28);
12: Store $(s_\tau, a_\tau, R_{\tau:min(\tau+N_1,N)}, s_{min(\tau+N_1,N)})$ into the replay buffer.
13: **end if**
14: Randomly sample a minibatch H of transition $(s_j, a_j, R_{j:j+N_1}, s_{j+N_1})$ from the replay buffer;
15: Update the parameters θ of the evaluation Q network by gradient descent step on Eqn. (30);
16: Update target Q network parameters every C step $\theta' \leftarrow \theta$;
17: $n \leftarrow n + 1$;
18: **end while**
19: $\epsilon \leftarrow \alpha_\epsilon \epsilon$;
20: **end for**

related to the state. The second is advantage stream, which is related to both the state and the action. The benefit is that it updates $Q(s, a)$ of all actions instead of only one action in a certain state, so that more values can be updated with less time and thus the updating process is more stable [46]. This corresponds to the following form:

$$Q(s, a|\theta, \alpha, \beta) = V(s|\theta, \beta) + A(s, a|\theta, \alpha) - \frac{1}{|A|} \sum_{a'} A(s, a'|\theta, \alpha) \tag{27}$$

where θ denotes the parameters of the convolutional layers, β and α are the parameters of the value stream and the advantage stream, respectively. $|A|$ is the size of action space.

Multi-step bootstrapping. The N_1-step bootstrapping takes into account the future return after N_1 steps. Compared with the temporal-difference (TD) learning with single-step update such as Q-learning, it can effectively accelerate the training process. Despite that it usually requires more storage space and a higher amount of calculation, it's worth for its advantages. The truncated N_1-step return is given by

$$R_{n:n+N_1} = \sum_{k=0}^{N_1-1} \gamma^k r_{n+k+1} \tag{28}$$

Specifically, it's accumulated to the termination step N at most, i.e., if $n + N_1 \geq N$, then $R_{n:n+N_1} = R_{n:n+N}$. A suitable size N_1 is very important to improve performance throughout the learning process [43].

B. Overall Multi-Step D3QN Algorithm

In the multi-step D3QN model, the UAV obtain a state from the state space $\mathcal{S}$ and selects an action from the action space $\mathcal{A}$ according to the following ϵ-greedy policy:

$$a = \begin{cases} \text{randomly selected from } \mathcal{A}, & \text{with probability } \epsilon; \\ \underset{a \in \mathcal{A}}{\text{argmax}}\, Q(s, a|\theta), & \text{with probability } 1 - \epsilon; \end{cases} \tag{29}$$

which is used to balance the exploration of new experience and the exploitation of the existing ones. A replay buffer of size B is used to store the N_1-step transitions (or experience) $(s_n, a_n, R_{n:n+N_1}, s_{n+N_1})$, and a minibatch of size H was randomly sampled from it to update the evaluation Q network by minimizing the loss function:

$$(R_{n:n+N_1} + \gamma^{N_1} Q'(s_{n+N_1}, \arg\max_{a'} Q(s_{n+N_1}, a'|\theta)|\theta') - Q(s_n, a_n|\theta))^2 \tag{30}$$

where N_1-step truncated return (28) is used in (26). θ and θ' are the parameters of the evaluation Q network and the target Q network respectively, both of which adopt the dueling structure in IV-A.

The pseudo-code of the proposed algorithm is presented in Algorithm 1. During the preparation stage of the training process, we initialize the replay buffer with capacity B. The parameters θ and θ' of the evaluation Q network and the target Q network are randomly initialized.

There are a total of K episodes, where a random discrete locations is selected as the initial state of UAV at the beginning of each episode. The UAV selects the action according to the ϵ-greedy policy. Since there is not enough experience to form a good flight strategy in the early stage of training, a higher probability of exploration is needed, and then as the strategy matures, we can reduce exploration. The initial exploration rate is ϵ, which decays at a rate of α and is not less than 0.01. After the action is executed, it's transferred to the next state, the outage probability is measured and the single-step reward is obtained according to Eqn. (25). The N_1-step truncated return is calculated and the N_1-steps transition $(s_n, a_n, R_{n,n+N_1}, s_{n+N_1})$ is stored into the replay buffer. (Lines 1-13).

Next, sample a minibatch of size H randomly from the replay buffer, calculate the loss function according to Eqn. (30) and update the evaluation Q network by performing gradient descent. The evaluation network parameter is periodically copied to the

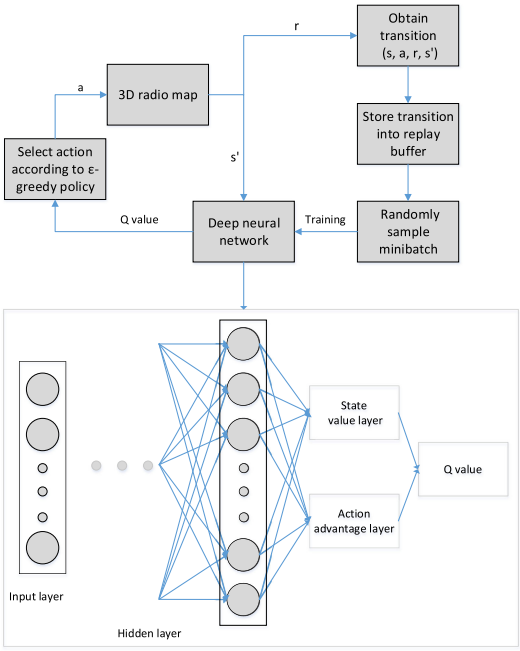


Fig. 3. Framework of the proposed algorithm for UAV path planning.

TABLE I
MAIN SYSTEM PARAMETERS

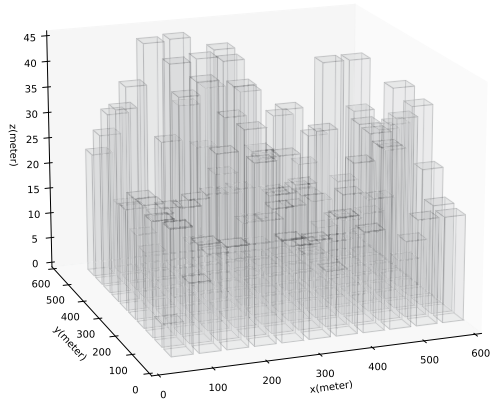
Notation	Definition	Simulation value
D_{tol}	Reaching destination tolerating distance	$10\sqrt{2}$ m
V	The velocity of the UAV	10 m/s
h_{max}	Maximum flight altitude	100 m
h_{min}	Minimun flight altitude	60 m
f_c	Carrier frequency	2 GHz
h_{bs}	The height of GBSs	10 m
J	Number of signal measurements	1000
Δ_d	The small grid point granularity	10 m
N_1	Multi-step bootstrapping size	4
ϵ	Initial exploration probability	0.7
α_ϵ	Exploration decay rate	0.9995
B	Replay buffer size	200000
H	Batch size	128
N_{max}	Maximum step per episode	600
C	Update interval steps for target network	1000
γ	Discount factor	0.99

target network every C steps. Finally, the next episode is started if the destination is reached or the maximum number of steps is exceeded (Lines 14-20). The framework of the proposed algorithm is shown in Fig. 3.

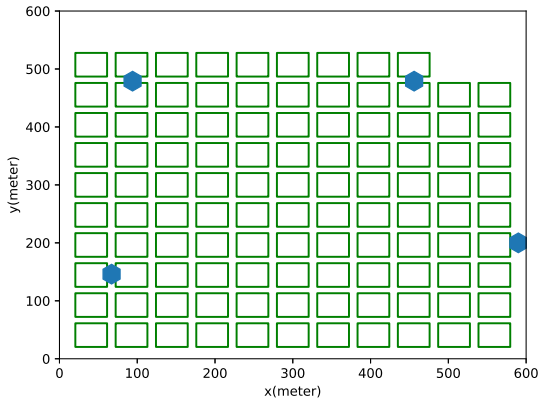
C. Analysis of the Proposed Algorithm and Suboptimal Solutions

1) *Convergence analysis:* To analyze the convergence of the proposed multi-step D3QN algorithm, the first step is to prove the convergence of the double Q-learning algorithm. According to theorem 1 in [47], it has been proved that the double Q-learning will converge to the optimal value with probability 1 as long as certain conditions are fulfilled, such as $0 \leq a_n \leq 1$, $\sum_n a_n = \infty$ and $\sum_n {a_n}^2 < \infty$. Additionally, by following the Universal Approximation Theorem [48], [49], the neural network can approximate any non-linear continuous function if it has enough parameters, and thus will succeeds in identifying the generated Q-values. In terms of the dueling networks and multi-step bootstrapping, it only relates to the training stability and speed. Overall, the convergence of the multi-step D3QN algorithm can be guaranteed. It is worth noting that the proposed algorithm is a suboptimal solution since the optimality of a reinforcement learning model can not be guaranteed.

2) *Complexity analysis:* Similar to [29], [35], we discuss the computational complexity of the proposed algorithm which consist of two main aspects, namely, the complexity related to fully connected layer model and the learning process. First, the neural network contains three hidden layer with fully connected layer model. As demonstrated in [50], the complexity of fully connected layer model is given by $\mathcal{O}(\sum_{l=1}^{3} F_l F_{l-1})$, where F_l is the number of neural units of the fully connected layer l. The neural networks are trained by leveraging backpropagation technique, and its parameters are updated continuously adopting the gradient descent until convergence. Second, the complexity of the learning process can be expressed as $\mathcal{O}(|\mathcal{S}| \cdot |\mathcal{A}|)$, where $|\mathcal{S}|$ is the total number of states determined by the grid



(a)



(b)

Fig. 4. The 3D and 2D views of building distribution in dense urban area. (a) 3D view. (b) 2D view and GBS locations marked by blue hexagon.

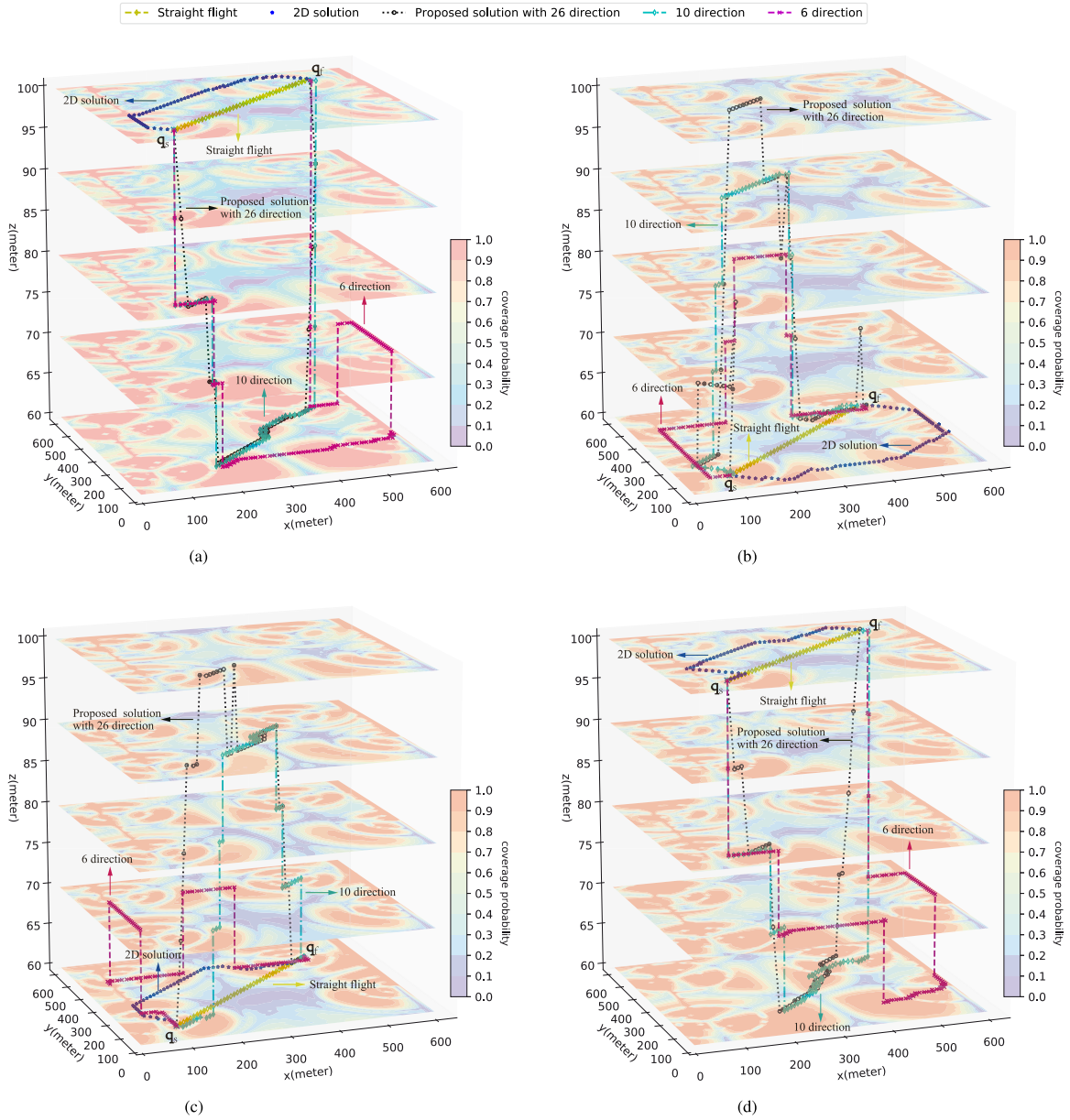


Fig. 5. Coverage map under different antenna downtilt angles and illustration of the proposed path solution. (a) Under $\theta_{\text{tilt}} = -6°$. (b) Under $\theta_{\text{tilt}} = -10°$. (c) Under $\theta_{\text{tilt}} = -14°$. (d) Under $\theta_{\text{tilt}} = -20°$.

point spacing granularity Δ_d, and $|\mathcal{A}|$ is the total number of actions determined by flight directions.

In addition, we propose two suboptimal solutions by adjusting the number of flight directions of the UAV which can reduce the complexity of learning process to some extent:

1. Suboptimal solution 1 with 10 directions: In this scheme, we assume that the number of action space $|\mathcal{A}| = 10$, which is expressed as {**front, back, left, right, left front, right front, left back, right back, ascent, descend**}, with specific reference to the numbers {5, 10–17, 22} in the Fig. 2.

2. Suboptimal solution 2 with 6 directions: In this scheme, we assume that the number of action space $|\mathcal{A}| = 6$, which is expressed as {**front, back, left, right, ascent, descend**}, with specific reference to the numbers {5, 11, 13, 14, 16, 22} in the Fig. 2.

V. Simulation and Performance Evaluation

In this section, numerical result will be provided to verify the effectiveness and superiority of the proposed 3D UAV path planning algorithm. As shown in the Fig. 4, we consider a dense urban area with a size of 600 m $\times$ 600 m, the building height limit $h_{bd} = 45$ m, and we set the three parameter of building distribution with $\alpha_{bd} = 0.3$, $\beta_{bd} = 300$, $\sigma_{bd} = 20$. We assume that there are 4 GBSs in the considered area, so the number of cells is $M = 12$. The transmit power of each cell is equal to

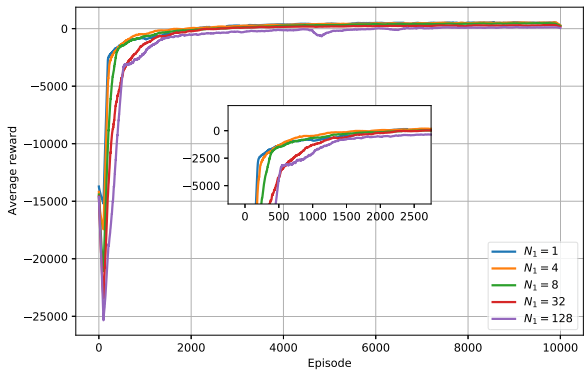


Fig. 6. Average reward trends with training episodes.

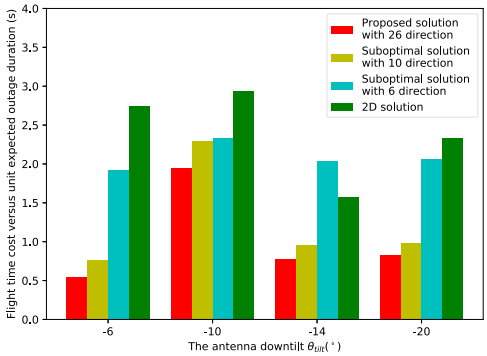


Fig. 7. Flight time cost versus unit expected outage duration(s).

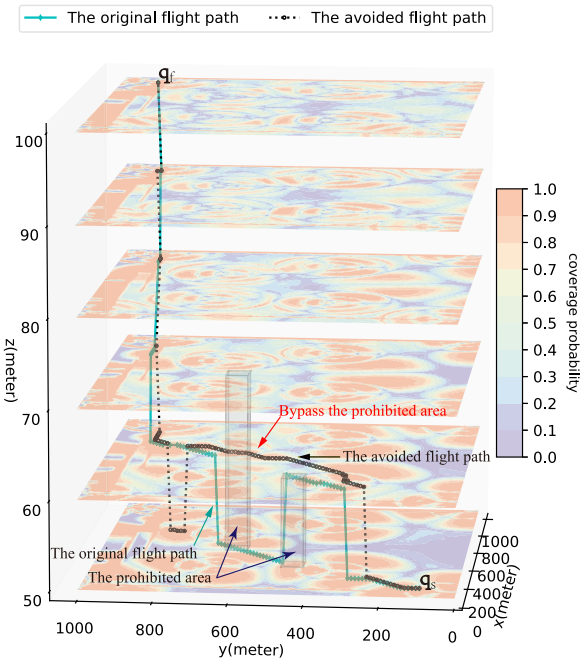


Fig. 8. 3D view of the prohibited area exists.

$P_m = 20$ dBm and the carrier frequency is 2 GHz. The Rician fading factor $K_R = 15$ dB in the case of LoS channel. The outage SINR threshold γ_{th} is considered as 3 dB and the small grid point spacing granularity is $\Delta_d = 10$ m. The minimum allowable altitude of UAV is $h_{min} = 60$ m, and the maximum is

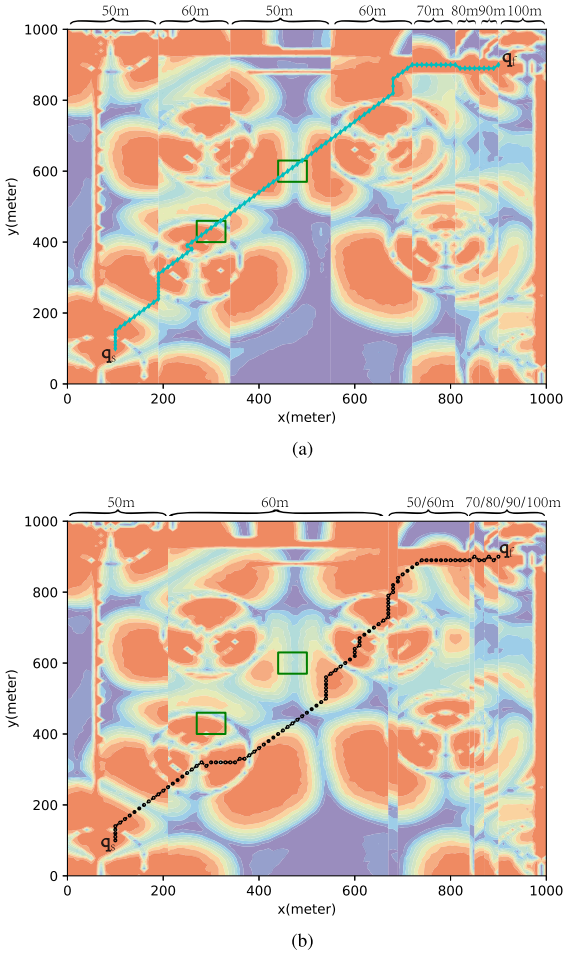


Fig. 9. Projections of the original and avoided flight path on their plane. (a) The original flight path and its spliced radio map. (b) The avoided flight path and its spliced radio map.

$h_{max} = 100$ m. The parameters in reward function (25) were set to: $r_a = 200$, $r_b = -100$, $\mu_1 = -0.1$, $\mu_2 = -50$ and $\mu_3 = 2$.

Simulation was implemented in Python 3.7 and Tensorflow 2.1. We establish the evaluation Q network and target Q network consisting of 3 hidden layers by $300 \times 300 \times 100$ fully connected feedforward ANN, ReLU is the activation function and AdamOptimizer is used to train the ANN. Other simulation parameters are shown in the Table I.

To demonstrate the validity of the proposed 3D path, we use the 2D path and straight flight path as benchmarks.

1) *2D Path:* The 2D path gives up the ability to change the altitude and only considering the flight at the same altitude, so the number of action space $|\mathcal{A}| = 4$, which is expressed as {**front, back, left, right, left front, right front, left back, right back**}, with specific reference to the number (10-17) in the Fig. 2.
2) *Straight Flight Path:* The straight flight path aim to minimize the flight time and without considering the effect of the expected outage time.

Fig. 5 shows the coverage map at different θ_{tilt} and the paths of the proposed solution. By comparing the coverage probability

maps of (a)-(d), it can be found that the change of θ_{tilt} affects the 3D radiation over the airspace. For the same plane, it will lead to the change of the total coverage. For the example plane with height of 60 m, there is a better connectivity probability overall when the $\theta_{tilt} = -6^\circ$ or -20°. However, the center area appears to have high outage probability with $\theta_{tilt} = -10^\circ$, -14°. Considering this situation, 2D path planning requires circumventing extremely long distances to avoid the high expected outage time. Fortunately, this disadvantage can be addressed by slightly adjusting the UAV's flight altitude, so that the UAV can fly on a plane with good coverage. Therefore, 3D path saves a lot of flight time and achieves better performance compared with its 2D counterpart. To better reflect the above mentioned situation, we set the plane location of the $\mathbf{q}_s$ and $\mathbf{q}_f$ as (100,100) and (500,500) respectively, and the height are divided into two types, 60 m and 100 m. The reaching destination plane tolerance distance D_{tol} is $10\sqrt{2}$ m.

Fig. 5(a) shows all the paths under $\theta_{tilt} = -6^\circ$. Straight flight is chosen to directly go through the low coverage area in the middle of the plane with a height of 100 m, which leads to the shortest flight time but also a very high expected outage time. The 2D solution takes more time to bypass the middle area, and even finds narrow road with high coverage probability due to direct signals at the X-axis around 80 m. In contrast, the proposed 3D solution and suboptimal solution 1 both choose a better strategy, that is, first descend to the 60 m plane with better coverage, where the high outage rate area has disappeared, and then return to the destination at 100 m, which produces a trajectory similar to the straight flight at the 2D level. In addition, it's observed that the suboptimal solution 2 chooses to fly along the boundary (X-axis around 600 m) due to the limitation of its direction.

Fig. 5(b) shows all the paths under $\theta_{tilt} = -10^\circ$. 2D solution chooses to bypass the low coverage area and detour to the destination along the boundary of the X-axis around 600 m. We can observe that the proposed 3D solution and suboptimal solution 1 still have a similar strategy, but the proposed 3D solution chooses to fly a certain distance on plane with height of 100 m. Suboptimal solution 2 also finds a narrow road with high coverage of the X-axis around 80 m, and then raises it's height to 80 m to skip the coverage fault.

Fig. 5(c) shows all the paths under the $\theta_{tilt} = -14^\circ$. 2D solution selects a detour strategy similar to that mentioned above in Fig. 5(a). The proposed 3D solution directly rises the height to 100 m and then drops to 90 m and overlap with the suboptimal solution 1. Suboptimal solution 2 choose to rise the height to 70 m to skip the coverage fault.

Fig. 5(d) shows all the paths under the $\theta_{tilt} = -20^\circ$. 2D solution also avoids the area with the most serious outage. Both the proposed 3D solution and the suboptimal solution 1 choose to descend to 60 m height at the beginning, and the only difference in the final ascent. Suboptimal solution 2 also circumvents along the boundary of X-axis around 600 m.

We show the convergence trend of the training process of the proposed algorithm. For the convenience of comparison, we average the reward every 200 episodes and the training process has a total of 10,000 episodes. As shown in Fig. 6, it can be observed that the average reward shows a decreasing trend at the beginning of training, the reason is that the UAV dose not have a good flight strategy to reach the destination or obtain the out of specified boundary punishment r_b. After that, the average reward gradually increased with episodes and finally tended to be stable. In addition, we can also see the influence of the size N_1 of multi-step bootstrapping on the convergence speed, the 4-step reward faster than the single-step after about 500 episodes.

Next, we show the corresponding detailed numerical results in Table II. As expected, straight flight takes the shortest time, but has suffer from the longest expected outage duration. 2D solution greatly reduces the expected outage duration, but the flight time is also significantly increased due to the poor coverage of its plane. In contrast, the proposed 3D solution with 26 direction has better performance. It can be observed that the flight time and the expected outage duration are lower than 2D solution under $\theta_{tilt} = -6^\circ$, -14° and -20°. Under $\theta_{tilt} = -10^\circ$, the expected outage duration is increased by 1.447 s but the flight time is decreased by 15.7 s.

For comparison purposes, we define an proportion index η_l based on straight flight as follows:

$$\eta_l = \frac{F_a - l_a}{|F_b - l_b|} \tag{31}$$

where l_a and l_b are the flight time and expected outage duration of straight flight path respectively. F_a and F_b are the flight time and expected outage duration of the path to be compared respectively. η_l means the flight time cost of decreasing the expected outage duration by one second. Fig. 7 compares η_l for each path solution, the proposed 3D solution with 26 direction has the lowest flight time cost and suboptimal solution 1 with 10 direction take second place. 2D solution is lower than suboptimal solution 2 with 6 direction only under $\theta_{tilt} = -14^\circ$.

In fact, the 3D path design may be affected by constraints other than communication requirements due to the complex real-world environment factors, such as the presence of particularly high buildings or non-flying areas restricted by laws and regulations. With these factors in mind, we present the relevant path planning in Fig. 8 and Fig. 9, which can be obtained by adding a dominantly large penalty term to the reward function (25) in case that the UAV enters the prohibited areas, thus discouraging such behaviors. As shown in the Fig. 8, the UAV's flying altitude and plane range have been further expanded, and the number of base stations has been increased to eight. Two of the buildings are prohibited areas and the UAV has to bypass them to avoid collision, we show the path planning with and without prohibited areas. Fig. 9(a) first shows the original flight path projected to the 2D plane, and then presents the spliced radio map, which can be obtained by splicing the radio map of different heights where the original flight path is located according to the x coordinates. Moreover, we use braces at the top to indicate the height of parts. Similarly, Fig. 9(b) shows the avoided flight path and its spliced radio map. By comparing Fig. 9(a) and Fig. 9(b), it is shown that the avoided flight path is planned to bypass the prohibited areas when the original flight path is not feasible.

TABLE II
NUMERICAL COMPARISON OF THE PROPOSED PATH SOLUTION

	$\theta_{tilt} = -6^\circ$		$\theta_{tilt} = -10^\circ$		$\theta_{tilt} = -14^\circ$		$\theta_{tilt} = -20^\circ$	
	Flight time	Expected outage duration	Flight time	Expected outage duration	Flight time	Expected outage duration	Flight time	Expected outage duration
Proposed solution with 26 direction	66.353	13.652	77.696	14.493	64.671	14.205	72.450	9.723
2D solution	80.639	25.007	93.396	13.046	74.083	14.346	77.882	20.857
Straight flight	55.154	34.290	55.154	26.085	55.154	26.350	55.154	30.598
Suboptimal solution 1 with 10 direction	72.154	11.937	78.468	15.918	66.811	14.205	76.154	9.209
Suboptimal solution 2 with 6 direction	107	7.232	93	9.886	90	9.283	104	6.863

VI. CONCLUSION

This paper studies the connectivity-aware 3D path planning problem for cellular-connected UAV to trade off between flight time and expected outage duration. We first introduce the 3D radio map construction considering the combined influence of the 3D antenna radiation patterns, the large-scale channel power gain and the small-scale power channel gain. Based on the radio map, we further propose a multi-step D3QN based algorithm to solve the MDP problem for achieving the local optimal path solution. Numerical results demonstrate that the 3D path of our algorithm significantly improves the overall performance compared to the trajectory with fixed heights. we also show the effects of 3D antenna radiation patterns on airspace coverage and the resulting changes in the UAV trajectory. In the future work, we plan to include the downtilt angle of the GBS into the overall problem study as an optimization variable to bring further performance improvement. Moreover, we have considered constant UAV speed in the current paper for simplicity and it is possible to extend our current framework by allowing the UAV to adapt its flying speed as well to further improve the performance. For example, a higher speed can be adopted to reduce the flight and outage duration through the area with high outage probability.

REFERENCES

[1] Y. Zeng, J. Lyu, and R. Zhang, "Cellular-connected UAV: Potential, challenges, and promising technologies," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 120–127, Feb. 2019.

[2] "Statista drones: Estimated size of the global commercial drone market in 2021 with a forecast for 2026." Accessed: Aug. 18, 2011. [Online]. Available: https://www.statista.com/statistics/878018/global-commercial-drone-market-size

[3] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. IEEE*, vol. 107, no. 12, pp. 2327–2375, Dec. 2019, doi: 10.1109/JPROC.2019.2952892.

[4] 3GPP TR 36.873: "Study on 3D channel model for LTE," V12.7.0, Dec. 2017.

[5] J. Lyu and R. Zhang, "Network-connected UAV: 3-D system modeling and coverage performance analysis," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7048–7060, Aug. 2019.

[6] R. Amer, W. Saad, B. Galkin, and N. Marchetti, "Performance analysis of mobile cellular-connected drones under practical antenna configurations," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–7.

[7] M. M. U. Chowdhury, W. Saad, and I. Güvenç, "Mobility management for cellular-connected UAVs: A learning-based approach," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2020, pp. 1–6.

[8] L. Liu, S. Zhang, and R. Zhang, "Multi-beam UAV communication in cellular uplink: Cooperative interference cancellation and sum-rate maximization," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4679–4691, Oct. 2019.

[9] W. Mei and R. Zhang, "Uplink cooperative NOMA for cellular-connected UAV," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 644–656, Jun. 2019.

[10] A. Rahmati, X. He, I. Guvenc, and H. Dai, "Dynamic mobility-aware interference avoidance for aerial base stations in cognitive radio networks," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2019, pp. 595–603.

[11] H. C. Nguyen, R. Amorim, J. Wigard, I. Z. KováCs, T. B. Sørensen, and P. E. Mogensen, "How to ensure reliable connectivity for aerial vehicles over cellular networks," *IEEE Access*, vol. 6, pp. 12 304–12 317, Feb. 2018.

[12] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[13] A. Trotta, M. D. Felice, F. Montori, K. R. Chowdhury, and L. Bononi, "Joint coverage, connectivity, and charging strategies for distributed UAV networks," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 883–900, Aug. 2018.

[14] Y. Zeng, R. Zhang, and T. J. Lim, "Throughput maximization for UAV-enabled mobile relaying systems," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 4983–4996, Dec. 2016.

[15] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3747–3760, Jun. 2017.

[16] X. Pang, J. Tang, N. Zhao, X. Zhang, and Y. Qian, "Energy-efficient design for mmWave-enabled NOMA-UAV networks," *Sci. China- Inf. Sci.*, vol. 64, no. 4, Apr. 2021, Art. no. 140303.

[17] 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.

[18] S. Zhang and R. Zhang, "Radio map based path planning for cellular-connected UAV," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.

[19] 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.

[20] F. Yin and F. Gunnarsson, "Distributed recursive Gaussian processes for RSS map applied to target tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 492–503, Apr. 2017.

[21] E. Bulut and I. Guevenc, "Trajectory optimization for cellular-connected UAVs with disconnectivity constraint," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2018, pp. 1–6.

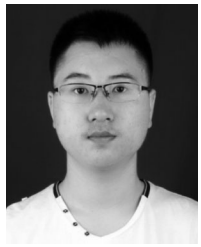
[22] H. Yang, J. Zhang, S. H. Song, and K. B. Lataief, "Connectivity-aware UAV path planning with aerial coverage maps," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2019, pp. 1–6.

[23] B. Khamidehi and E. S. Sousa, "Federated learning for cellular-connected UAVs: Radio mapping and path planning," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.

[24] B. Khamidehi and E. S. Sousa, "A double Q-learning approach for navigation of aerial vehicles with connectivity constraint," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–6.

[25] 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.

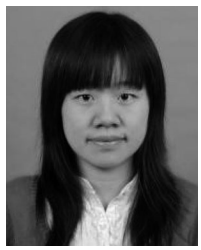
[26] M. M. Azari, F. Rosas, K. Chen, and S. Pollin, "Ultra reliable UAV communication using altitude and cooperation diversity," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 330–344, Jan. 2018.
[27] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
[28] Y. Zeng and X. Xu, "Toward environment-aware 6G communications via channel knowledge map," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 84–91, Jun. 2021.
[29] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042–2055, Jul. 2021.
[30] K. Wan, X. Gao, Z. Hu, and G. Wu, "Robust motion control for UAV in dynamic uncertain environments using deep reinforcement learning," *Remote Sens.*, vol. 12, no. 4, pp. 1–21, Feb. 2020.
[31] Z. Hu, K. Wan, X. Gao, Y. Zhai, and Q. Wang, "Deep reinforcement learning approach with multiple experience pools for UAV's autonomous motion planning in complex unknown environments," *Sensors*, vol. 20, no. 7, Apr. 2020, Art. no. 1890.
[32] H. Bayerlein, M. Theile, M. Caccamo, and D. Gesbert, "UAV path planning for wireless data harvesting: A deep reinforcement learning approach," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.
[33] X. Liu, Y. Liu, and Y. Chen, "Reinforcement learning in Multiple-UAV networks: Deployment and movement design," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8036–8049, Aug. 2019.
[34] M. M. U. Chowdhury, S. J. Maeng, E. Bulut, and Güvenç, "3-D trajectory optimization in UAV-Assisted cellular networks considering antenna radiation pattern and backhaul constraint," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 3735–3750, Oct. 2020.
[35] W. Zhang, Q. Wang, X. Liu, Y. Liu, and Y. Chen, "Three-dimension trajectory design for multi-UAV wireless network with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 600–612, Jan. 2021.
[36] J. Hu, H. Zhang, and L. Song, "Reinforcement learning for decentralized trajectory design in cellular UAV networks with sense-and-send protocol," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6177–6189, Aug. 2019.
[37] C. H. Liu, X. Ma, X. Gao, and J. Tang, "Distributed energy-efficient Multi-UAV navigation for long-term communication coverage by deep reinforcement learning," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1274–1285, Jun. 2020.
[38] H. V. Abeywickrama, Y. He, E. Dutkiewicz, B. A. Jayawickrama, and M. Mueck, "A reinforcement learning approach for fair user coverage using UAV mounted base stations under energy constraints," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 67–81, Feb. 2020.
[39] H. Qi, Z. Hu, H. Huang, X. Wen, and Z. Lu, "Energy efficient 3-D UAV control for persistent communication service and fairness: A deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 53 172–53 184, Mar. 2020.
[40] ITU-R, Rec. P.1410-5, "Propagation data and prediction methods required for the design of terrestrial broadband radio access systems operating in a frequency range from 3 to 60 GHz," *Radiowave Propag.*, Feb. 2012.
[41] M. Rebato, L. Resteghini, C. Mazzucco, and M. Zorzi, "Study of realistic antenna patterns in 5G mmWave cellular scenarios," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–6.
[42] 3GPP TR 36.777, Technical specification group radio access network: study on enhanced LTE support for aerial vehicles, V15.0.0, Dec. 2017.
[43] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[44] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.
[45] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 2094–2100.
[46] Z. Wang *et al.*, "Dueling network architectures for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, vol. 48, 2016, pp. 1995–2003.
[47] H. van Hasselt, "Double Q-learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 2613–2621.
[48] A. Sannai, Y. Takai, and M. Cordonnier, "Universal approximations of permutation invariant/equivariant functions by deep neural networks," 2019, *arXiv:1903.01939*. [Online]. Available: https://arxiv.org/abs/1903.01939
[49] V. Timofte, A. Timofte, and L. Khan, "Stone-Weierstrass and extension theorems in the nonlocally convex case," *J. Math. Anal. Appl.*, vol. 462, no. 2, pp. 1536–1554, 2018.
[50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.



Hao Xie received the B.S. degree in communication engineering from the Nanchang University College of Science and Technology, Nanchang, China, in 2018. He is currently working toward the master's degree with the Information Engineering School, Nanchang University. His research interests include wireless communications, and machine learning.

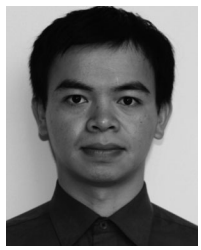
Dingcheng Yang (Member, IEEE) received the B.S. degree in electronic engineering and the Ph.D. degree in space physics from Wuhan University, Wuhan, China, in 2006 and 2012, respectively.

He is currently a Professor with the Information Engineering School, Nanchang University, Nanchang, China. He has authored or coauthored more than 50 papers including journal papers on IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, etc. and conference papers such as IEEE GLOBECOM. His research interests include cooperation communications, IoT/cyber-physical systems, UAV communications, and wireless resource management.



Lin Xiao (Member, IEEE) received the Ph.D. degree in electronic engineering from the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K., in 2010.

After that, she worked with the China Academy of Telecommunication Research, MIIT for one year. She is currently a Professor with the Information Engineering School, Nanchang University, Nanchang, China. Her research interests include wireless communication and networks, in particular, UAV network planning and optimization, radio resource management, relay, and cooperation communication.



Jiangbin Lyu (Member, IEEE) received the B. Eng. degree (Hons.) in control science and engineering, and completed the Chu Kochen Honors Program with Zhejiang University, Hangzhou, China, in 2011, and the Ph.D. degree from NUS Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore (NUS), Singapore, in 2015, under the NGS scholarship.

From 2015 to 2017, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, NUS. He is currently an Assistant Professor with the School of Informatics, Xiamen University, China. His research interests include UAV communications, intelligent reflecting surface, cross-layer network optimization, etc. He was the recipient of the IEEE ComSoc Heinrich Hertz Award for Best Communications Letters in 2020, and also the Best Paper Award at Singapore-Japan Int. Workshop on Smart Wireless Communications in 2014. He was the Invited Track Co-Chair at the 2021 IEEE/CIC ICCC conference.