Zhong *et al.*
EURASIP Journal on Advances in Signal Processing (2024) 2024:35
https://doi.org/10.1186/s13634-024-01130-6

EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

Joint optimization of UAV communication connectivity and obstacle avoidance in urban environments using a double-map approach

Weizhi Zhong[1*], Xin Wang[1], Xiang Liu[1], Zhipeng Lin[1] and Farman Ali[2]

*Correspondence:
zhongwz@nuaa.edu.cn

[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, Jiangsu, China
[2] Department of Electrical Engineering, Qurtuba University of Science and IT, Dera Ismail Khan 29050, Pakistan

Abstract

Cellular-connected unmanned aerial vehicles (UAVs), which have the potential to extend cellular services from the ground into the airspace, represent a promising technological advancement. However, the presence of communication coverage black holes among base stations and various obstacles within the aerial domain pose significant challenges to ensuring the safe operation of UAVs. This paper introduces a novel trajectory planning scheme, namely the double-map assisted UAV approach, which leverages deep reinforcement learning to address these challenges. The mission execution time, wireless connectivity, and obstacle avoidance are comprehensively modeled and analyzed in this approach, leading to the derivation of a novel joint optimization function. By utilizing an advanced technique known as dueling double deep Q network (D3QN), the objective function is optimized, while employing a mechanism of prioritized experience replay strengthens the training of effective samples. Furthermore, the connectivity and obstacle information collected by the UAV during flight are utilized to generate a map of radio and environmental data for simulating the flying process, thereby significantly reducing operational costs. The numerical results demonstrate that the proposed method effectively circumvents obstacles and areas with weak connections during flight, while also considering mission completion time.

Keywords: Cellular-connected UAV, Trajectory planning, Radio map, DRL, Environment characteristic

1 Introduction

With its low cost, clear line-of-sight (LoS), and deployment flexibility, UAV communication technology has gradually become an integral component of future sixth generation (6G) networks [1]. However, in order to practically realize the application of UAVs in 6G networks, several critical challenges need to be addressed, including high-capacity, low-latency, and reliable links. Nevertheless, most existing civilian UAV links primarily rely on a simplistic point-to-point (P2P) communication pattern and utilize commonly used frequency bands such as ISM 2.4 GHz [2]. Furthermore, it is important to acknowledge certain limitations such as regional constraints, lower transmission rates, inadequate information confidentiality, and increased communication interference. To address

these challenges and meet the escalating data demands of future 6G systems, there is an urgent need for establishing ultra-reliable, high-rate, and secure wireless communication between ground cellular networks and UAVs. In this regard, cellular-connected UAVs have emerged as a promising technology that can fulfill diverse requirements. UAVs can serve as relays [3, 4] or base stations (BSs) [5] to facilitate wireless communications without direct connectivity. In comparison with conventional air-to-ground (A2G) communication, cellular-connected UAVs offer numerous advantages. Firstly, leveraging the global cellular infrastructure enables cost-effective communication links and facilitates extensive UAV operations. Secondly, compared to simple peer-to-peer wireless communication, cellular-connected UAVs provide reduced latency and enhanced data transmission rates, thereby promising substantial performance enhancements. Specifically, cellular-connected UAVs have the potential to expand the conventional two-dimensional (2D) cellular network into a future three-dimensional (3D) architecture, which would greatly benefit both UAV and cellular industries. However, despite the promising application prospects of cellular-connected UAV communication, there are still several challenges that need to be addressed. The existing conventional cellular network is primarily designed for ground users (GUs) [6], resulting in ground base station (GBS) antennas being tilted downwards towards the ground. This configuration limits their ability to provide optimal coverage for air connections. Furthermore, cellular-connected UAVs are susceptible to significant co-channel interference from other unconnected GBS.

To tackle these issues, various strategies have been proposed in the literature. Some studies aim to enhance A2G communication conditions for improved coverage rate and spectral efficiency. In order to maximize the coverage of GUs, researchers in [2] suggested employing a generalized Poisson multinomial distribution to simulate interference information. In [7], the authors proposed a two-stage strategy utilizing Deep Reinforcement Learning (DRL) to optimize the placement of aerial BSs. The GBS antenna inclination was utilized as an optimization objective in [8] to maximize transmission quality and minimize switching time, thereby enhancing the overall performance of the system. In [9], the authors optimized the positioning, user clustering, and frequency band allocation of UAVs to enhance the coverage rate and minimize the required number of UAVs. The authors in [10] proposed a cooperative interference elimination strategy based on the information regarding backhaul links between GBSs in cellular networks to effectively mitigate interference caused by non-associated BSs. The authors propose an alternative scheme in [11] and [12], which utilizes the non-orthogonal multiple access technique to achieve successive interference cancellations at each GBS.

In addition to the aforementioned studies, there has been further research conducted on UAV trajectory control in order to mitigate potential issues arising from weak connectivity between UAVs and GBSs. In [13], the signal-to-interference-plus-noise ratio (SINR) map was constructed, and by employing graph theory, the UAV trajectory was optimized under the constraint of SINR. In [14], the authors employed graph theory to elucidate the correlation between connection interrupt rate and path length, while addressing the connectivity issue by investigating the shortest path with enhanced GBS coverage capabilities in undirected weighted graphs. In [15], the cellular-connected UAVs were subjected to both convex optimization and graph theory techniques, aiming

to minimize the mission execution time while ensuring connectivity with at least one GBS. Additionally, prior studies [16] and [17] have also addressed similar issues. In [18], the authors consider the anti-collision and communication interference constraints between UAVs, and maximizes system throughput by jointly optimizing vehicle communication scheduling, UAVs power distribution, and UAVs trajectory. However, conventional trajectory design approaches tend to oversimplify channel models for diverse environments, rendering them unsuitable for practical applications. For instance, previous studies [15] and [17] simplified the environmental models by making certain assumptions, such as considering free-space path loss and assuming isotropic radiation for antennas. Studies [19–21] have considered statistical channel models incorporating probabilistic LoS and angle-dependent parameters. However, these simplified and constrained models fail to accurately capture real-world channel conditions, rendering them unsuitable for practical environments. Moreover, trajectory optimization poses a challenging non-convex problem with exponentially increasing complexity as the number of optimization variables grows, rendering it difficult to solve. Fortunately, the rapid advancements in machine learning (ML) have led to investigations into trajectory design methods based on DRL aiming to tackle these aforementioned challenges [22–27]. Such approaches acquire navigation strategies by actively interacting with the environment and collecting empirical data.

The advantages of DRL have led to its widespread utilization across various scenarios. For instance, a framework called simultaneous navigation and radio mapping (SNARM) was proposed [22], which employs the Dueling Double Deep Q network (D3QN) to construct a radio map solely based on raw signal measurements. This approach enables accurate prediction of outage probabilities at all significant locations. In [27], the creation of a 3D radio map was described, and the utilization of the multi-step D3QN technique was employed for UAV trajectory design. Although these frameworks are applicable to diverse environments, they do not account for additional factors present in complex settings. For instance, urban areas often pose challenges such as tall buildings, no-fly zones, and flying objects that need to be considered alongside reliable connectivity. Therefore, apart from ensuring dependable communication links, it is crucial to address effective obstacle avoidance. To enable the effective application of cellular-connected UAVs in complex urban environments, it is imperative to ensure both reliable wireless connectivity and obstacle avoidance. However, there is a limited number of studies addressing this crucial aspect at present. In [28], several trajectory planning methods have been proposed solely for obstacle avoidance purposes. In [29], a novel scheme incorporating environment sensing and channel mapping was presented to enhance trajectory planning in unknown 3D airspace with obstacles. Nevertheless, the seamless integration of reliable connectivity and obstacle avoidance in [29] was conducted independently.

To address the aforementioned gap and facilitate joint optimization of reliable wireless connectivity in unknown 3D airspace with obstacles, we propose a novel path optimization method based on environmental awareness within the cellular context. The main contributions and innovations are summarized as follows:

- ·The proposed approach presents a joint optimization strategy for UAV path, integrating obstacle avoidance and communication connectivity. Moreover, we formulate

the optimization problem by introducing a potential function that considers factors such as flying time, communication interruptions, and distance variations between UAVs and obstacles.

- ·We propose a framework for path planning of UAVs called the Double-Map-Assisted UAV (DMAU) framework. This framework utilizes connectivity and obstacle distance information collected by the UAV during its flight to train a map of radio and environmental data. The mapping network generates data that is used to simulate UAV flight training, enabling a combination of simulated and actual flying which accelerates training speed and reduces UAV flight costs.
- ·The proposed framework introduces a learning approach for joint path optimization using an enhanced D3QN. Specifically, by incorporating the prioritized experience replay (PER) mechanism based on the sum tree in the network, diversity sampling replaces traditional uniform sampling to enhance learning efficiency and reduce computational complexity in path optimization.

The remainder of this paper is organized as follows. Section 2 introduces the system model. In Sect. 3, the problem formulation and the proposed algorithm are presented. Section 4 gives the analyzed and simulated results. Finally, conclusions are drawn in Sect. 5.

2 System model

2.1 Scenario model

As depicted in Fig. 1, we consider a scenario model wherein a single UAV functions as an aerial user, establishing communication with a cellular network in a densely populated urban area. The UAV is assigned special missions and is expected to reach the designated destination from its initial location within the shortest possible time while ensuring uninterrupted communication connectivity and avoiding collisions..

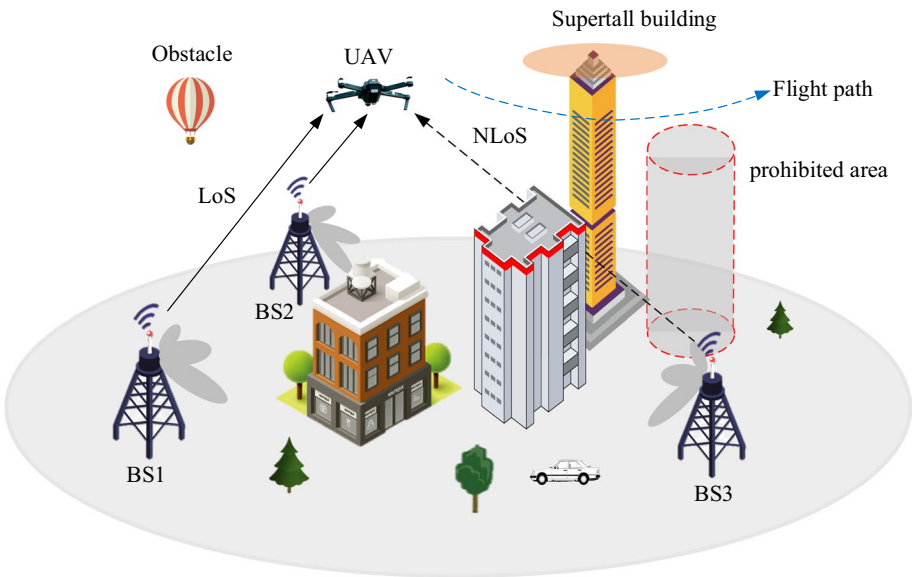


Fig. 1 Path planning for cellular-connected UAV in urban scene

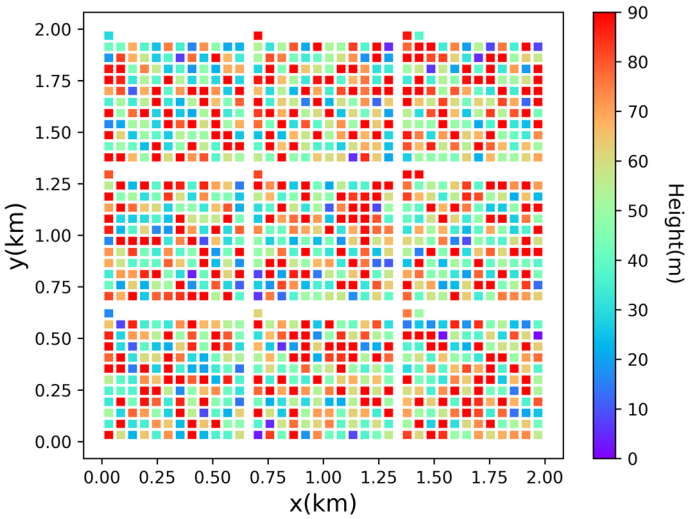


Fig. 2 The spatial distribution and vertical dimensions of the buildings

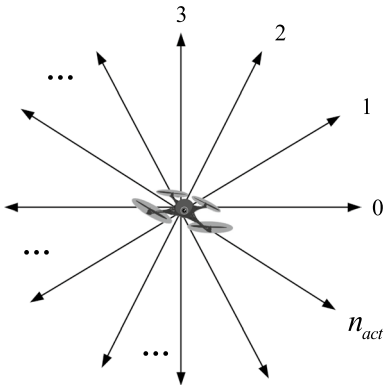


Fig. 3 Action space

To accurately establish the scenario model of cellular-connected UAVs, we consider a flying area of size $D \times D\,km^2$ encompassing high-rise buildings. The heights and positions of these buildings are generated based on the statistical model proposed by the International Telecommunication Union (ITU). Additionally, an overview map depicting the distribution of these buildings is illustrated in Fig. 2.

The constant altitude of the UAV during flight is represented as h. The mission execution time is denoted as T, and the position of the UAV at time t is defined as $g(t) = (x_t, y_t)$, where $t \in [0, T]$,$x_t \in [0, D]$, $y_t \in [0, D]$, and variables x_t and y_t denote the X-coordinate and Y-coordinate of the UAV, respectively.

The definition of motion space significantly impacts the efficacy of UAV path planning. In principle, the motion space of a UAV can be represented in any direction. However, excessive movement of a UAV will considerably augment the training time required for learning model DQN, whereas limited movement of the UAV will result in zigzag motion. As depicted in Fig. 3, we partition the 360° angle into n_{act} equal segments, denoted as $\varphi = 360°/n_{act}$, representing the precision of UAV heading accuracy φ. The spatial range of

UAV movement varies with adjustments made to the heading precision, thereby granting greater flexibility for flight path planning.

2.2 Antenna model

The BS is modeled in this section to represent the antenna radiation. It is assumed that there are 7 GBSs within the designated airspace [30]. These GBSs, equipped with a uniform linear array (ULA) consisting of n elements, are divided into M cells and have a fixed height of h_{bs} meters. Let θ and ϕ denote the UAV's elevation and azimuth angles relative to the base station, respectively. The gain of each pair of angles for the antenna element can be expressed as

$$A_E(\theta,\phi) = G_{E,\max} - \min\left\{-[A_{E,V}(\theta) + A_{E,H}(\phi)], A_m\right\} \tag{1}$$

where $G_{E,\max}$ represents the maximum directional gain of each antenna element in the direction of the main lobe, while A_m denotes the front-back ratio. The vertical and horizontal radiation patterns are denoted by $A_{E,V}(\theta)$ and $A_{E,H}(\phi)$ respectively, which can be defined as

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

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

where θ_{3dB} and ϕ_{3dB} denote the half-power beam widths in the vertical and horizontal dimensions,, while SLA_V represents the limit of side lobe level.

In the case, the array factor can be obtained by

$$AF(\theta,\phi,n) = 10\log_{10}\left[1 + \rho(\left|a \cdot w^T\right|^2 - 1)\right] \tag{4}$$

where n represent the antenna elements, ρ denotes the correlation coefficient, a represents the amplitude vector, and w signifies the beamforming vector. The latter is defined as

$$w = \left[\omega_{1,1}, \omega_{1,2}, \ldots, \omega_{m_V,m_H}\right] \tag{5}$$

and

$$\omega_{p,r} = e^{\pi((p-1)(\cos\theta - \cos\theta_s) + (r-1)(\sin\theta\sin\phi - \sin\theta_s\sin\phi_s))} \tag{6}$$

where m_V and m_H denoted the array elements of the antenna in the vertical and horizontal directions, respectively, $m_V m_H = n$, while the pair of angle (θ_s, ϕ_s) defines as the direction of main lobe.

Combining with (1) and (4), the radiation pattern can be written as

$$A_A(\theta,\phi) = A_E(\theta,\phi) + AF(\theta,\phi,n). \tag{7}$$

The current elevation and azimuth information can be obtained when the coordinates $g(t)$ of the UAV are provided. Consequently, the antenna gain received at position $q(t)$ can be defined as

$$\beta(s(t)) = 10^{\frac{A_A(\theta,\phi)}{10}}. \tag{8}$$

2.3 Signal model

In this section, we establish the received signal model and introduce the concept of expected outage probability. The instantaneous signal power received by UAV from cell m at location $g(t)$ is defined as

$$y_m(t) = P_m\left|h_m(t)\right|^2, m = 1, \ldots, M \tag{9}$$

where P_m represents the transmitting power of cell m, and $h_m(t)$ is the channel gain at time t, which can be written as

$$\left|h_m(t)\right|^2 = \beta(g(t))\bar{h}_m(g(t))\tilde{h}_m(t) \tag{10}$$

where $\beta(s(t))$ given by (8) represents the gain of the GBS antenna at location $g(t)$, $\widetilde{h}_m(t)$ is a random variable with $E\left[\left|\widetilde{h}_m(t)\right|^2\right]$=1, representing the small-scale fading [31]. $\bar{h}_m(g(t))$ is the large-scale channel gain, and can be expressed as

$$\bar{h}_m(g(t)) = \begin{cases} h_m^{LoS}(g(t)), & \text{if LoS link} \\ h_m^{NLoS}(g(t)), & \text{if NLoS link} \end{cases}. \tag{11}$$

Based on the urban Macro (UMa) in 3GPP specification [32], $h_m^{LoS}(g(t))$ and $h_m^{NLoS}(g(t))$ can be defined as

$$h_m^{LoS}(g(t)) = 28 + 22\log_{10}(d_m(g(t))) + 20\log_{10}(f_c) \tag{12}$$

$$h_m^{NLoS}(g(t)) = -17.5 + \left(46 - 7\log_{10}(h)\right)\log_{10}\left(d_m(g(t))\right) + 20\log_{10}\left(\frac{40\pi f_c}{3}\right) \tag{13}$$

where f_c denotes the carrier frequency, h represents the flying altitude of the UAV, which is assumed to be a constant,$d_m(g(t))$ is the distance between UAV and cell m at location $g(t)$, which is given by

$$d_m(g(t)) = \sqrt{(h - h_{bs})^2 + \left\|g(t) - g_m\right\|^2} \tag{14}$$

where $\|\cdot\|$ is the Euclidean norm, and g_m is the location of the GBS corresponding to the cell m.

Based on (9), the signal-to-interference ratio (SIR)[33] between the UAV and the associated GBS at time t can be defined as

$$SIR(t) = \frac{y_{b(t)}(t)}{\sum_{m \neq b(t)} y_m(t)}, b(t) \in \{1, \cdots, M\} \tag{15}$$

where $b(t)$ is the associated cell of UAV, $y_{b(t)}(t)$ denotes the instantaneous signal power received from the associated cell m, $y_m(t)$ mainly depends on the location of UAV, the current associated cell and small-scale fading. In the case, $SIR(t)$ can be rewritten as $SIR(g(t), b(t), \tilde{h}_{b(t)})$. We use the outage probability to evaluate the communication connectivity between UAV and GBS. When the SIR is lower than the set threshold ρ_{th}, the UAV is considered to be in an outage state. In this condition, the outage probability can be defined as

$$P_{out}(s(t), b(t)) \triangleq \Pr\left\{SIR(s(t), b(t), \tilde{h}_{b(t)}) < \rho_{th}\right\}. \tag{16}$$

where $\Pr\{\cdot\}$ represents the probability of event happening.

2.4 Obstacle avoidance model

In addition to ensuring connectivity probability, effective obstacle avoidance plays a crucial role in UAV cellular operations. In an unfamiliar environment, UAVs are unable to anticipate environmental information beforehand. When an obstacle emerges within the observation range of the airborne sensor, the UAV can promptly execute appropriate maneuvers to evade it.

When employing intelligent optimization algorithms like DQN for obstacle avoidance, it is typically imperative to establish rewards for UAVs. The conventional approach to obstacle avoidance reward usually entails assigning a fixed negative value as a consequence of the next action when the UAV approaches an obstacle; conversely, a fixed positive value is assigned otherwise. However, this definition fails to quantify the impact of the action on the UAV. The obstacle avoidance rewards under different circumstances, as depicted in Fig. 4, are quantified based on the UAV's distance from the obstacle. This categorization encompasses four distinct scenarios:

(1) The UAV did not detect any obstacles at time t and $t+1$. In this scenario, the lack of environmental prediction information prevents the UAV from determining whether it will encounter an obstacle in the next moment, despite its actions in the current state. Since this is unrelated to the UAV's actions, a reward of 0 is assigned for encountering an obstacle.

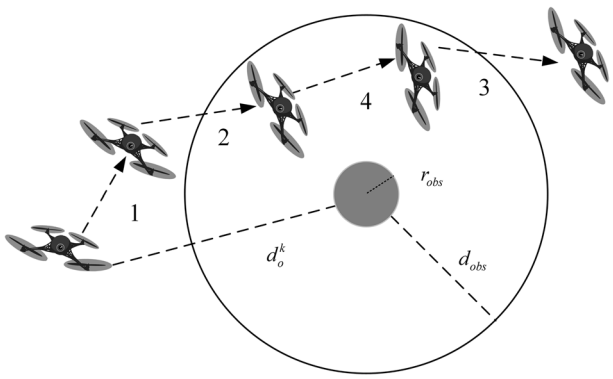


Fig. 4 The relative motion of a UAV with respect to an obstacle

(2) At time t, the UAV does not detect any obstacles; however, at time $t+1$, an obstacle is detected by the UAV. This observation suggests that the current action performed by the UAV is in close proximity to the obstacle, and introducing obstacles yields a negative reward for the UAV.

(3) The UAV detected an obstacle at time t, while no obstacle was found at time $t+1$. This observation suggests that the current action executed by the UAV is to avoid the obstacle, and the presence of the obstacle is considered a positive reinforcement for the UAV.

(4) The UAV detected the obstacle at both time t and $t+1$, which presents a relatively intricate scenario necessitating a quantitative formulation of the reward function for the UAV in relation to the obstacle. The potential function reward associated with an obstacle encountered by the UAV can be defined as

$$r(t)=\begin{cases}0,\ d_o^t > d_{obs}, d_o^{t+1} > d_{obs}\\ -1,\ d_o^t > d_{obs}, d_o^{t+1} \le d_{obs}\\ 1,\ d_o^t \le d_{obs}, d_o^{t+1} > d_{obs}\\ \dfrac{d^{t+1}-d^t}{\left|d^{t+1}-d^t\right|} e^{\frac{1}{\left|d^{t+1}-d^t\right|}}, d_o^t \le d_{obs}, d_o^{t+1} \le d_{obs}\end{cases}, \tag{17}$$

where d_o^t represents the minimum distance between the UAV and any obstacle at time t, d_{os} is the radius of said obstacle and d_{obs} is a constant representing the observation range of the sensor utilized in the airborne sensing system of said UAV. The radius of the obstacle is set to r_{obs}. In case the denominator is zero, when the value of $d^{t+1}-d^t$ is in the range of $[-1,0)$, set $d^{t+1}-d^t=-1$; when the value of $d^{t+1}-d^t$ is in the range of $[0,1]$, set $d^{t+1}-d^t=1$.

3 Preliminary knowledge and DRL based path planning

The present paper proposes a novel approach for UAV trajectory design, taking into account the duration of communication outages, mission completion time, and obstacle avoidance. Figure 5 illustrates the flowchart of the proposed method, which comprises three main parts: modeling optimization objectives, constructing joint optimization objectives, and optimizing trajectories. Firstly, we model three optimization objectives including communication outage duration, mission completion time, and obstacle avoidance probability. Subsequently, a radio map is constructed by utilizing SIR measurement values and an obstacle avoidance strategy is developed based on obstacle information. To achieve joint optimization, we combine the radio map, obstacle avoidance strategy, and mission completion time. Finally, DRL is employed to design trajectories that align with the joint optimization objective.

3.1 Radio map and environmental information map

The Radio map is a tool that facilitates the visualization of communication quality's spatial distribution. In this subsection, we generate a radio map by utilizing the outage probability of the UAV at all locations within the designated area to provide connectivity information during simulated flight. The outage probability is accurately obtained by defining the outage indicator function as

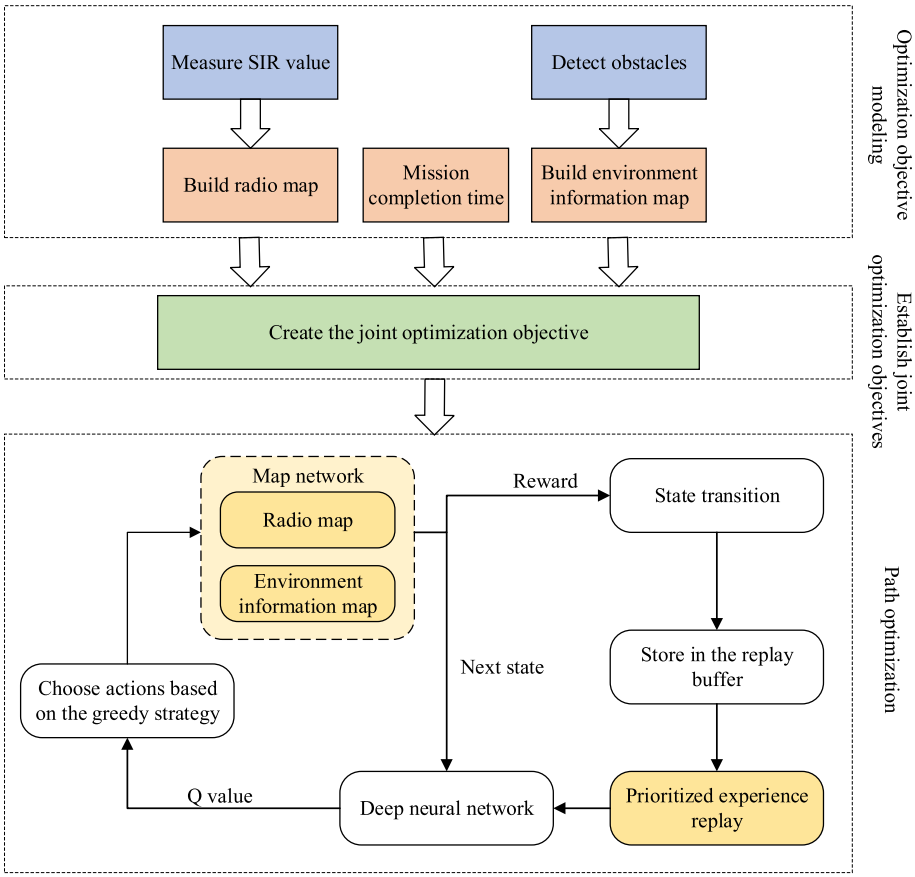


Fig. 5 Flowchart for UAV path planning

$$F(g(t), b(t), \tilde{h}_{b(t)}) = \begin{cases} 1, \ SIR(g(t), b(t), \tilde{h}_{b(t)}) < \rho_{th} \\ 0, \ otherwise \end{cases} \tag{18}$$

The assumption is made that the UAV continuously measures the SIR of each cell M times within a short time period. Let $J = \sum_{j=1}^{K} F(g(t), b(t), \tilde{h}_{b(t)})$, and the outage probability at time t can be obtained by

$$P_{out}(g(t), b(t)) = \frac{J}{K}. \tag{19}$$

Based on the measured outage probability, the best associated GBS at the location $g(t)$ can be determined as

$$b(t)^* = \arg \min_{b(t) \in M} P_{out}(g(t), b(t)) \tag{20}$$

where $\arg\min \cdot$ represents the value of the variable that minimizes the objective function.

The outage probability of the position $g(t)$ can be expressed as

$$P_{out}(g(t)) = \min_{b(t) \in M} P_{out}(g(t), b(t)). \tag{21}$$

In order to construct the map of communication connectivity probability (CCP) in the UAV flying area, the CCP of the location $g(t)$ can be defined as

$$P_{cover}(g(t)) = 1 - P_{out}(g(t)). \tag{22}$$

Based on the aforementioned theory, we can derive the connectivity probability of each position from signal measurements and subsequently construct a radio map based on this probability. Similarly, upon detecting an obstacle, the UAV can record both its current position and distance information to generate an environmental information map. Especially when no obstacles are detected by the UAV, the distance value of the current position can be set to a constant $d_{obs} + a$ that exceeds the sensor detection range, where a is positive.

3.2 Reformulate a collaborative optimization objective

Based on the above discussion, the following three optimization objectives are considered in this paper.

(1) Minimizing the UAV's flight duration from the initial point to the destination.
(2) Minimizing the expected outage time between the UAV and the GBS.
(3) Refraining from colliding with obstacles within the designated airspace.

For the above three objectives, the joint optimization problem can be formulated as

$$\begin{aligned} &\min_{T,g(t)} \; T + \mu \int_0^T P_{out}(g(t))dt - \eta \int_0^T r(t)dt \\ &s.t.\ \mathrm{g}(0) = g_s,\ \mathrm{s}(T) = g_f \\ &0 \le x_t \le D,\ \forall t \in [0, T] \\ &0 \le y_t \le D,\ \forall t \in [0, T] \\ &\mu > 0,\ \eta > 0 \end{aligned} \tag{23}$$

where μ and η are non-negative coefficients that respectively represent the weight coefficients of connectivity and obstacle avoidance, respectively. The greater the value of μ, the higher the emphasis placed on wireless connectivity; similarly, the larger the value of η, the greater attention is given to obstacle avoidance performance. The duration of outage is expected to increase as the mission completion time T improves, while maintaining a constant outage probability $P_{out}(g(t))$. However, as the mission completion times increase, the UAV becomes more adaptable in adjusting its path to avoid areas with weak coverage and reduce expected outage time. Similarly, during obstacle avoidance, the flight path of a UAV tends to become more convoluted, resulting in longer mission completion times. Therefore, there is generally a tradeoff between minimizing mission completion time, expected outage duration, and effective obstacle avoidance. When constructing a joint optimization objective function, it is necessary to assign appropriate weight coefficients to balance their interrelationships.

Given the intricacy of continuous optimization, it is necessary to discretize the flying area and flying actions into a discrete trajectory planning problem on grid points. To achieve this objective, we consider $T = N\Delta t$ and observe that the distance between the UAV and any BS remains approximately constant within Δt, while both the large-scale

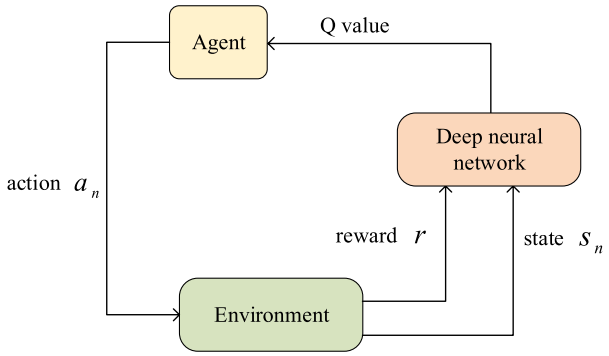


Fig. 6 DRL model

channel gain and the BS antenna gain remain nearly invariant. In the case, (28) can be equivalently written as

$$\begin{aligned}
&\min_{N,g(n)} \; N + \mu \sum_{n=1}^{N} P_{out}(g(n)) - \eta \sum_{n=1}^{N} r(n) \\
&s.t.\ \mathrm{g}(1) = g_s,\ \mathrm{s}(N) = g_f \\
&0 \le x_t \le D,\ \forall n \in [1, N] \\
&0 \le y_t \le D,\ \forall t \in [1, N] \\
&\mu > 0,\ \eta > 0.
\end{aligned} \tag{24}$$

Clearly, the aforementioned problem is non-convex and poses significant challenges in terms of solvability, with its complexity escalating substantially as the number of parameters to be optimized increases. Fortunately, the trajectory planning issue can be formulated as a Markov decision process (MDP), and In addition to ensuring connectivity probability, effective obstacle avoidance plays a crucial role in UAV cellular operations. In an unfamiliar environment, UAVs are unable to anticipate environmental information beforehand. When an obstacle emerges within the observation range of the airborne sensor, the UAV can promptly execute appropriate maneuvers to evade it. algorithms exhibit immense potential in tackling such intricate problems [34]. Consequently, we employ DRL to explore an optimal flight path based on experiential learning through trial and error within a specific environment.

3.3 Basic of DRL

In this subsection, we first present a concise overview of DRL [35], and then introduce our proposed algorithm as detailed in the subsequent section.

The DRL model, depicted in Fig. 6, comprises a combination of RL and DNN. RL, which aims to maximize the cumulative reward through agent-environment interactions, is an effective machine learning technique that adapts well to Markov decision processes (MDP).

In the RL model, there are two pivotal components—the agent and the environment. As the driving force behind the RL algorithm, the agent perpetually engages in a cycle of learning and exploration within its surroundings. Based on the current state s_n provided by the environment, the agent strategically selects an action a_n. The agent

state s_n transitions to s_{n+1} simultaneously, accompanied by the feedback of reward r_{n+1}. By iteratively repeating the aforementioned process, the agent can efficiently attain the optimal strategy and successfully accomplish the learning task within a specific environment. The objective of the agent is to optimize the overall cumulative reward G_n, which can be defined as

$$G_n = \sum_{k=0}^{\infty} \gamma^k r_{n+k+1} \tag{25}$$

where $0 \leq \gamma \leq 1$ is a discount factor, signifying the present-time discounting of future rewards. A higher value of γ emphasizes the significance of long-term returns, while a smaller value of γ highlights the importance of short-term gains.

However, due to the unknown quantity of G_n at time n persisting throughout the episode (where an episode refers to the complete process of the UAV from start to finish, crash, outbound or reaching maximum steps), obtaining an accurate value for G_n becomes unattainable. In this case, we address the problem by employing an expectation-based approach to derive the action-value function Q_π, which is equal to

$$Q_\pi(s_n, a_n) = \mathbb{E}_\pi[G_n | s = s_n, a = a_n] \tag{26}$$

where $\pi(a_n|s_n) = \mathbb{P}[a = a_n|s = s_n]$ is the policy function that represents the probability of selecting and executing action a_n in state s_n. The action-value function Q_π represents the expected return that can be derived by following strategy $\pi(a_n|s_n)$. Suppose there is an optimal strategy π_* with higher return than other strategies and can be expressed as $\pi_* = \arg\max_\pi Q_\pi(s_n, a_n)$, which makes $Q^*(s_n, a_n) = \max_\pi Q_\pi(s_n, a_n)$. $Q^*(s_n, a_n)$ represents the optimal function of action-value, and satisfies

$$Q^*(s, a) = r(s, a) + \gamma \sum_{s'} p(s'|s, a) \max_{a'} Q^*(s', a'). \tag{27}$$

However, the Eq. (27) is nonlinear in nature and generally lacks a closed-form solution. To address this issue, we can employ the concept of temporal difference (TD) learning, which proves to be an effective approach for obtaining an estimation of action-value.

$$Q(s_n, a_n) \leftarrow Q(s_n, a_n) + \alpha \left[r_n + \gamma \max_a Q(s_{n+1}, a) - Q(s_n, a_n) \right] \tag{28}$$

where $r_n + \gamma \max_a Q(s_{n+1}, a) - Q(s_n, a_n)$ is defined as the TD-error and can be represented by ε_n. Specifically, the TD learning algorithm belongs to a category of model-free reinforcement learning methods that estimate value functions by directly sampling state-action-reward-next state sequences, and update the value function estimates using bootstrapping.

The aforementioned RL method is called table-base, which necessitates the storage of each state-action pair and proves unsuitable for scenarios involving an exceedingly large number of states or actions. The present study employs the DQN approach to address this issue. It uses deep neural network (DNN) as a function of approximator and assumes $Q(s, a) \approx \hat{Q}(s, a; \theta)$, where θ is the network parameter, corresponding to

the weights and bias of all links in the DNN. The Q network is updated by minimizing the loss function, which can be modified as

$$\left(r_n + \gamma \max_a \hat{Q}(s_{n+1}, a; \theta) - \hat{Q}(s_n, a_n; \theta)\right)^2. \tag{29}$$

However, applying the standard training algorithm (29) directly may lead to oscillations and divergence. Therefore, a target network with its parameter set to θ^- is introduced in [36].

The parameter θ in the Q network can be updated B times, and then set θ^- be changed for the next B times update. Correspondingly, the loss function in (13) can be rewritten as

$$\left(r_n + \gamma \max_a \hat{Q}\left(s_{n+1}, \arg\max_{a'} Q(s_{n+1}, a'; \theta); \theta^-\right) - \hat{Q}(s_n, a_n; \theta)\right)^2. \tag{30}$$

This contributes to maintaining the target's relative stability, thereby enhancing the convergence characteristics of the training process. Furthermore, we employ a multi-step bootstrapping technique that effectively enhances the training speed by considering the future reward after N_1 steps. The truncated N_1-steps reward is given by

$$r_{n:n+N_1} = \sum_{k=0}^{N_1-1} \gamma^k r_{n+k+1}. \tag{31}$$

It should be noted that when $n + N_1 \geq N$, $r_{n:n+N_1} = r_{n:N}$, i.e., it is accumulated to N -step at most.

Based on the aforementioned analysis, the loss function of (30) can be reformulated as

$$\left(r_{n:n+N_1} + \gamma^N \max_a \hat{Q}\left(s_{n+N_1}, \arg\max_{a'} Q(s_{n+N_1}, a'; \theta); \theta^-\right) - \hat{Q}(s_n, a_n; \theta)\right)^2. \tag{32}$$

3.4 Prioritized experience replay

The experience replay is another important technique in DRL, where transitions (s_n, a_n, r_n, s_{n+1}) are stored in a replay buffer and randomly sampled to update network parameters. The experience replay technology (ERT) facilitates the reuse of sampled information acquired through the interaction between the agent and its environment. The correlation between the samples is broken through random sampling, but this mechanism cannot differentiate the significance of the samples. The limited capacity of the replay buffer further exacerbates the issue of low sampling efficiency. In this case, we propose a PER mechanism to replace the traditional uniform sampling, increasing the frequency of learning useful data and decreasing the frequency of learning useless data. This method enhances learning efficiency, achieves more accurate results, and optimizes UAV paths effectively.

The PER mechanism assigns sampling weights based on the TD-errors of transitions, where the absolute value of TD-error is utilized as the sampling probability denoted by $P_j = |\varepsilon_j| + \sigma$. Additionally, a parameter σ is introduced to prevent the occurrence of zero sampling probabilities. The larger the TD-error, the greater the potential for enhancing prediction accuracy, indicating that learning based on this sample can achieve superior performance. In this case, higher sample priority $P(j)$, which corresponds to a larger TD-error, can be defined as

$$P(j) = \frac{P_j^{\delta}}{\sum_i P_i^{\delta}} \tag{33}$$

where δ determines whether to prioritize sampling, and when $\delta = 0$, the sampling belongs to the uniform random sampling, $\sum_i P_i^{\delta}$ represents the cumulative sum of transition priorities in the replay buffer.

The use of a data structure called sum-tree avoids the need for extensive computation when calculating sampling priorities in each sampling process. The sum-tree is a hierarchical structure resembling a tree, where each leaf node stores the priority value of an individual sample. Each internal node has exactly two child nodes, and its value represents the cumulative sum of its children's values. Consequently, the root node of the sum-tree corresponds to the total sum of all priorities. When the batch sample size is m, priority $(0, \sum_i P_i^{\delta}]$ is evenly divided into intervals. Subsequently, a random value is generated within each interval, and the corresponding transition sample is retrieved from the sum-tree. The sampling process is shown in Algorithm 1.

Algorithm 1 Sampling from a Sum Tree

Input: minibatch m, replay buffer size C, the sum of priorities
The samples in the experience pool should be labeled and stored in descending order based on their size
Divided the priority range into m intervals
for k =1, 2, …, m **do**
    Randomly select a label l from k interval
    While not arrive a leaf node:
        if l <= the value of the left node:
            keep on retrieve left node using l
        else:
            keep on retrieve left node using l - the value of the right node
end for

By prioritizing, the Q network can enhance training efficiency and optimize path results. The introduction of priority alters the sample distribution, necessitating the use of importance sampling weights ω_j to rectify this discrepancy. The sampling weights ω_j can be given by

$$\omega_j = \frac{\left(P(j)\right)^{-\beta}}{\max_i \left((P(i))^{-\beta}\right)} = \left(\frac{P(j)}{\min_i (P(i)}\right)^{-\beta} \tag{34}$$

where β is a hyperparameter, which plays a crucial role in determining the impact of PER on the convergence outcome. Accordingly, the loss function in (32) can be rewritten as

$$\omega_j \left(r_{j:j+N_1} + \gamma^N \max_a \hat{Q}\left(s_{j+N_1}, \arg\max_{a'} Q(s_{j+N_1}, a'; \theta); \theta^-\right) - \hat{Q}(s_j, a_j; \theta)\right)^2. \tag{35}$$

3.5 DMAU algorithm for UAV path planning

The proposed approach integrates a potential function (PF) D3QN, and Prioritized PER algorithms to optimize the connectivity and obstacle avoidance of UAVs. The proposed DMAU algorithm is summarized in Algorithm 2. In this paper, the UAV is considered as an autonomous agent, and the state space S, action space A, and reward function r_n are described as follows.

(1) State space S: The current state of the UAV at time n is denoted as position g_n, while the set of all possible positions within the flying region constitutes the state space.
(2) Action space A: The action space of the UAV, encompasses all feasible directions for flight and is characterized by continuity.
(3) Reward function r_n: Corresponding to the objective function of (24), and the reward function r_n is set to $r_n = 1 + \mu P_{out}(g_{n+1}) - \eta r_{ob}(n+1)$.

The present study introduces several enhancements to address the limitations associated with insufficient prior environmental knowledge, exorbitant training costs relying solely on actual UAV flight, and the suboptimal efficiency of traditional DQN random sampling. The network of obstacle distribution is established in step 8 of Algorithm 2, and the distance information obtained from the sensor is utilized to update the network parameter, enabling the UAV to acquire obstacle avoidance behavior during simulated flight. The sampling operation is performed according to the priority in step 10 of Algorithm 2, while important sampling weights are assigned to the loss function in step 12. Subsequently, steps 15–21 utilize predicted outage probability and obstacle information from neural networks to simulate the UAV's flying process, which significantly accelerates algorithm convergence by incorporating real-world flight data.

Algorithm 2 DMAU for joint optimization of UAV connectivity and obstacle avoidance

Initialization: the replay buffer D with capacity C, initial exploration $\varepsilon = \varepsilon_0$, the DQN network with parameter θ, the target network with parameter θ^-, the radio map E and the map network with parameter θ_{radio}, and the environment information map O and the map network with parameter θ_{obs}

for n_{epi} =1, 2, …, N_{epi} **do**

Initialize the slide window W of size N_1, along with the initial position g_s and the actual flying step $n = 0$.

Select action a_n by $\varepsilon - greedy$ policy

Perform action a_n to get the next state g_{n+1}

Detect the distance of surrounding obstacles $d_o(g_{n+1})$, measure outage probability $P_{out}(g_{n+1})$

Save $(g_{n+1}, P_{out}(g_{n+1}))$ to E and sample minibatch from E to update the parameter θ_{radio}

Save $(g_{n+1}, d_o(g_{n+1}))$ to O and sample minibatch from O to update the parameter θ_{obs}

Set each step reward as $\mathrm{r}_n = 1 + \mu P_{out}(g_{n+1}) - \eta r_{ob}(n+1)$ and store transition (g_n, a_n, r_n, g_{n+1}) in the W

When $n \geq N_1$, calculate N_1-step reward $r_{n-N_1:n}$ based on (30), and store the transition $(g_{n-N_1}, a_{n-N_1}, r_{(n-N_1):n}, g_n)$ in the replay buffer D

Sample minibatch of transition $(g_j, a_j, r_{j:j+N_1}, g_{j+N_1})$ from D based on PER mechanism and obtain the importance-sampling weight ω_j

Set

$$y_j = \begin{cases} r_{j:j+N_1} + r_{des}, & \left\| g_{j+N_1} - g_f \right\| \leq D_{tol} \\ r_{j:j+N_1} - p_{ob}, & \mathrm{g}_{j+N_1} \notin S \\ r_{j:j+N_1} - p_{cra}, & d_o^{j+N_1} \leq D_{obs} \\ r_{j:j+N_1} + \gamma^{N_1} \hat{Q}(g_{j+N_1}, a^*; \theta^-), & otherwise \end{cases}$$

Execute a gradient descent step on $\omega_j (y_j - \hat{Q}(g_j, a_j; \theta))^2$

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

for $\tilde{n}_{epi}$ =1, 2, …, $\tilde{N}_{\text{epi}}$ **do**

Reinitialize the simulated initial position g_s and recalibrate the simulated flying step $n = 0$

Determine the outage probability $P_{out}(g_{n+1})$ by utilizing map E and considering the distance between the UAV position and obstacle $d_o(g_{n+1})$ as depicted in map O

Perform operations similar to steps 4,5 and 9-13 for the simulated experience

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

If $\left\| g_n - g_f \right\| \leq D_{tol}$ or $g_n \notin S$ or $d_o^n \leq D_{obs}$ or $n_{step} = N_{step}$, return to step 15

end for

Repeat step 4-19 until $\left\| g_n - g_f \right\| \leq D_{tol}$ or $g_n \notin S$ or $d_o^n \leq D_{obs}$ or $n_{step} = N_{step}$

After every B episodes, set $\theta^- = \theta$

end for

The initialization of algorithm 2 involves setting the various parameters in step 1. It is important to note that during the initial phase, when the UAV has limited knowledge about the environment, the initialization process should prioritize guiding the UAV to follow the shortest path towards its destination. In each episode of actual flight, the UAV commences from a randomly determined location, executes an action based on strategy $\varepsilon - greedy$, and carries it out. The probability of selecting an action randomly is denoted as ε, while the probability of selecting the action with the highest value is represented by $1 - \varepsilon$, i.e.,

$$a_n = \begin{cases} randi(A) \\ \arg\max\limits_{a \in A} \hat{Q}(q_n, a; \theta) \end{cases}. \tag{36}$$

The UAV employs its sensors in steps 6–7 to detect surrounding obstacles and assess the outage probability at the current location, thereby calculating the reward value. The outage probability and obstacle distance are utilized as input samples to update two network parameters, denoted as θ_{radio} and θ_{obs}, respectively.

The simulated flying process, encompassing steps 14–21, is initialized independently from the actual flight. It is noteworthy that during the simulated flight process, we are unable to acquire information regarding the actual obstacles and outage probability. Therefore, two networks are employed in step 17 to facilitate the generation of a simulated UAV flight experience. The number of episodes $\tilde{N}_{epi} = \min([n_{epi}/100], 10)$ determines the duration of simulated flying in relation to actual flight. As the number of actual flight episodes increases, so does the number of simulated flight episodes, thereby enhancing the reliability of forecasted rewards, expediting the training process, and yielding cost savings.

4 Numerical results

The performance evaluation of the proposed joint optimization algorithm is presented in this section through numerical results. Our proposed algorithm 2, DMAU, extends the traditional D3QN algorithm by creating a joint optimization function of obstacle avoidance and connectivity guaranteeing based on PF, inserting the learning operation of the radio map and environment information map, and adding the PER mechanism based on the sum tree. The DMAU model employs a fully connected feedforward neural network with 5 hidden layers for both the Q network and the target network. The number of neurons is 512, 256, 128, 128, and $n_{act} + 1$, where n_{act} corresponds to the action advantages of n_{act} actions, and the other one corresponds to the estimated value of the state. The radio network and obstacle distribution network are equipped with 5 hidden layers, each consisting of 521, 256, 128, 64, and 32 neurons respectively. The activation function employed in the hidden layer is Rectified Linear Unit (ReLU), while the Adam optimizer is utilized to train the ANN with an objective of minimizing the mean square error (MSE) loss. The designated destination position for the UAV is set at [1400, 1600]. Simulation parameters utilized in model construction are presented in Table 1, while additional parameters relevant to DMAU can be found in Table 2.

The proposed algorithm's validity is verified by comparing it with D3QN and D3QN-PER. Both D3QN and D3QN-PER penalize the UAV for colliding with obstacles, but

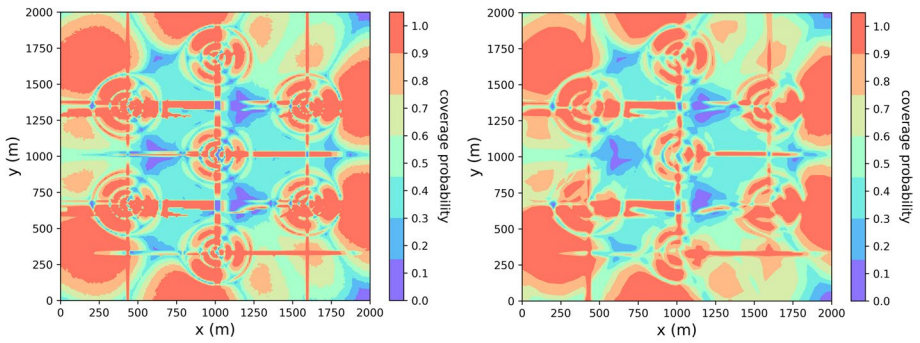
Table 1 Parameters of the system model

Simulation parameter	Value
Flight range D	2 km
The ratio of the built-up area to the total land area α_{bd}	0.3
The mean density of buildings per unit area β_{bd}	300
The average value of the buildings height distribution σ_{bd}	50
The flying height h	100 m
Flying speed V	10 m/s
The height of GBSs h_{bs}	25 m
The maximum directional gain of each individual antenna element $G_{E,\max}$	3 dBi
Front-back ratio A_m	30 dB
The half-power beamwidths θ_{3dB} and ϕ_{3dB}	65°
The limit of the side lobe level SLA_V	30 dB
The transmitted power of cell P_m	0.1 W
The carrier frequency f_c	2 GHz
Outage threshold ρ_{th}	0 dB

Table 2 Parameters for training algorithms

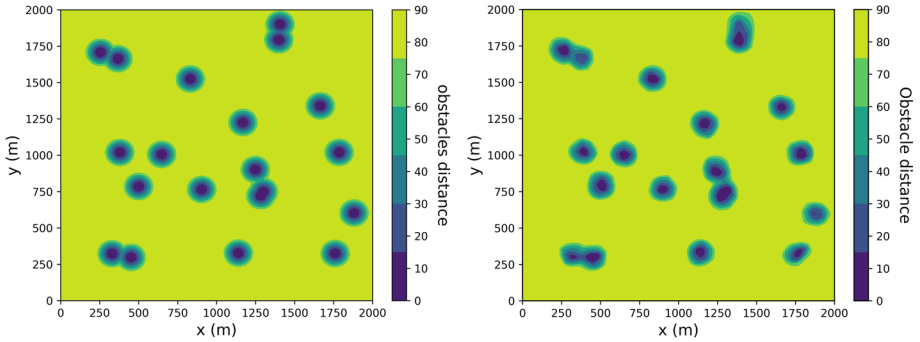
Simulation parameter	Value
Maximum number of episodes N_{epi}	5000
Replay buffer capacity C	100,000
Initial exploration ε_0	0.5
Exploration decay rate α	0.998
Slide window size N_1	32
Outage penalty weight μ	30
Obstacle avoidance weight η	50
Maximum step per episode N_{step}	300
Reaching destination tolerating distance D_{tol}	20 m

unlike D3QN, D3QN-PER incorporates the PER mechanism. The two methods lack a simulation of the flight process, fail to incorporate the radio map network or obstacle information network, and solely rely on real-time measurements during flight for path planning. The actual radio map within the flying area is depicted in Fig. 7a, which is obtained through computer simulation considering building distribution and channels, and can be generated by UAV measurements in practical scenarios. The analysis of Fig. 7a reveals the presence of multiple regions with weak coverage, characterized by a coverage probability below 0.3, in close proximity to the central area. Evidently, for effective cover-aware UAV navigation, it is imperative to steer the UAV away from entering areas with weak coverage in order to ensure uninterrupted communication connectivity. The quality validation of the radio map generated by the DMAU framework proposed in Algorithm 2 is demonstrated in Fig. 7b, which presents the final estimation of the radio map achieved by algorithm 2. The comparison reveals a remarkable similarity between the two radio maps, exhibiting only minor discrepancies. This serves as a compelling demonstration of Algorithm 2's exceptional capability in radio map estimation and coverage-aware path learning.



(a) Actual radio map (b) Learned radio map

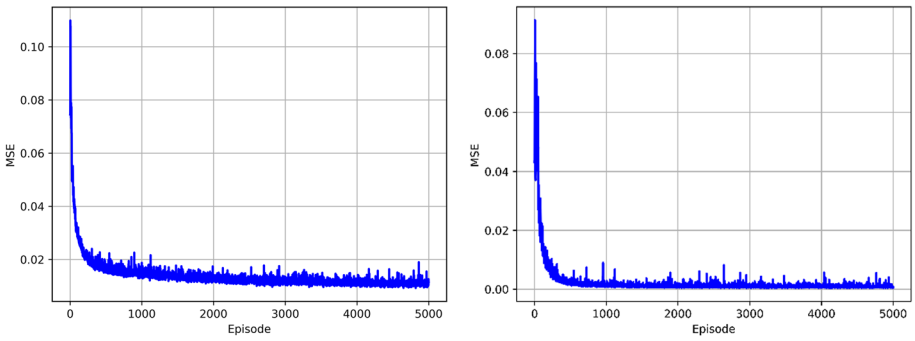
Fig. 7 The comparison of radio maps



(a) Map of actual environmental information (b) Map of acquired environmental information

Fig. 8 The comparison of maps depicting the distribution of obstacles

The obstacle distribution in the flying area, along with the distance to the nearest obstacle for each location, is illustrated in Fig. 8a. The yellow region denotes obstacles that fall beyond the sensor's detection range, resulting in a lack of distance information. Evidently, the navigation system for UAVs should effectively guide them to circumvent these obstacles while ensuring optimal communication connectivity. The environment information map learned through the obstacle distribution network in algorithm 2 is depicted in Fig. 8b. Upon careful observation and comparison, we note a minimal disparity between the two figures, thereby effectively substantiating algorithm 2's robust perception of obstacles. The MSE of the learned radio map and environment information map versus the episode number are illustrated in Fig. 9, respectively. The MSE is calculated by comparing the predicted outage probabilities in the learned radio map with their actual values in the real map for a set of randomly selected locations. In the initial stages, the lack of environmental knowledge resulted in significant inaccuracies. With an increasing number of episodes, there was a noticeable enhancement in signal measurement, leading to a higher quality learned radio map. Similarly, as depicted in Fig. 9b, an increase in the number of episodes enabled more accurate detection of obstacle distances and consequently improved the quality of the learned environment information map.



(a) MSE of radio map (b) MSE of the distribution map for obstacles

Fig. 9 The MSE of radio map and obstacles distribution map

The moving average returns per episode of different algorithms are depicted in Fig. 10, with a moving window length of 200 episodes. It can be observed from the figure that despite experiencing certain fluctuations, all three algorithms exhibit an overall upward trend in average returns.

The paths of multiple UAVs, randomly selected from the last 100 learning episodes, are depicted in Fig. 11. All sub-figures within Fig. 11 share common initial positions indicated by a black cross. The obstacles in the figures are depicted as solid black circles, while the red circles indicate the detectable range of obstacles. In Fig. 11a, b, it is evident that in the absence of obstacle avoidance using PF, the UAV collides with an obstacle at approximately position (300, 1000), leading to a forced termination of flight. The proposed method effectively avoids obstacles and ensures high coverage probability along the routes, as demonstrated in Fig. 11c. For instance, the UAV successfully detects and navigates through a 'radio narrow bridge'located approximately 1000 m along the x-axis without any collision with obstacles. This exemplifies how our method adeptly considers both obstacle avoidance and connectivity requirements. However, due to potential deviations in the UAV's trajectory for obstacle avoidance purposes, it may inadvertently bypass the optimal connectivity path, thereby increasing both the expected outage time and flight duration.

To assess the connectivity in our proposed joint optimization algorithm, we sequentially assign numbers to the paths depicted in Fig. 11 based on their starting positions from left to right and top to bottom. The resulting table (Table 3) presents the aggregated weighted sum of both expected outage time and mission completion time for each route. Additionally, We conducted a comparative analysis between the baseline algorithm (specifically, the D3QN algorithm unaffected by obstacles) and its obstacle-free counterpart. The connectivity of D3QN is relatively good, as evident from Table 3. However, due to the absence of fly process simulation, the training efficiency is compromised and certain paths exhibit poor connectivity, such as 1, 3, and 6. The D3QN-PER algorithm outperforms due to the incorporation of the PER mechanism. Although the DMAU algorithm exhibits proficient obstacle avoidance capabilities, it compromises connectivity to some extent. However, in certain unobstructed paths such as paths 3 and 6, DMAU can even outperform the baseline algorithm by identifying superior flight routes due to the accelerated learning efficiency facilitated by our proposed simulated fly process, thereby enabling the UAV to traverse more

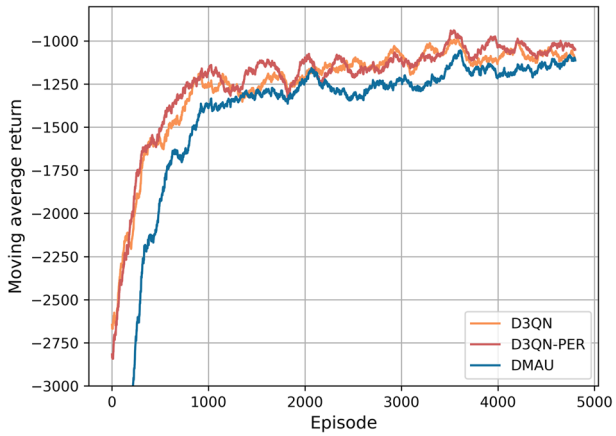
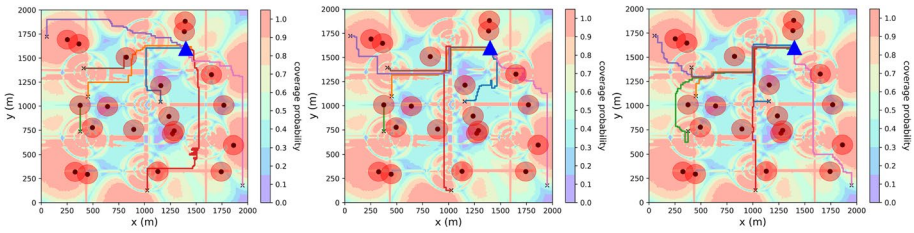


Fig. 10 Moving average return



(a)D3QN (b)D3QN-PER (c) DMAU

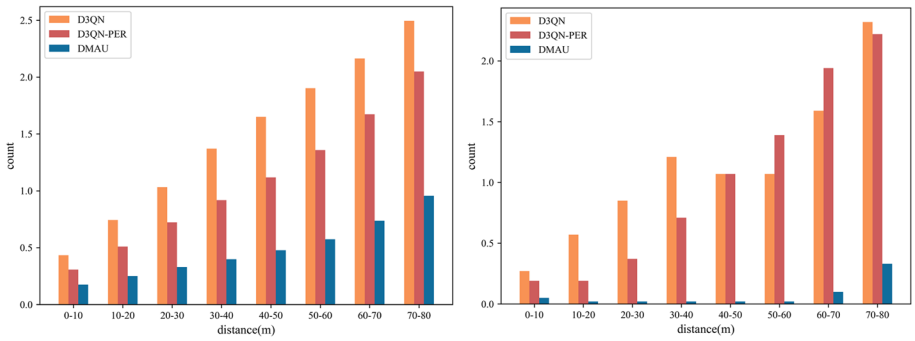
Fig. 11 Resulting path by different algorithms

Table 3 The aggregate of the weighted expected outage time and mission completion time

	1	2	3	4	5	6	7
D3QN	1320.26	–	2313.67	1205.18	–	2188.47	1057.84
D3QN-PER	1148.23	761.00	1236.59	1316.18	–	2015.57	1147.76
DMAU	1163.22	860.86	1024.04	1391.60	1001.60	1871.33	1007.67
No obstacle	1141.07	811.52	1403.38	1138.20	965.70	1955.03	925.23

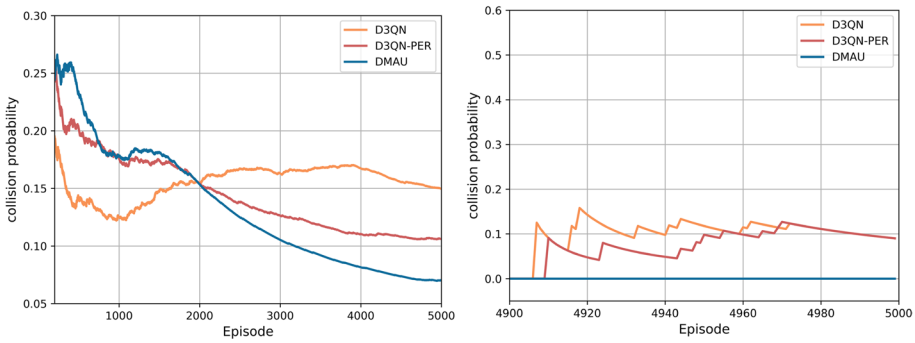
well-connected trajectories. The results demonstrate that the proposed DMAU algorithm effectively ensures enhanced connectivity while successfully circumventing obstacles.

The effectiveness of obstacle avoidance in the proposed joint optimization method is evaluated using a novel evaluation strategy, while ensuring the preservation of certain connectivity. The UAV sensors recorded the distance and frequency of obstacle detection in all episodes, where a lower value indicates fewer instances of the UAV approaching obstacles, thus implying a more effective obstacle avoidance performance. The simulation results based on the aforementioned evaluation methods are depicted in Fig. 12. Specifically, Fig. 12a, b illustrate the average number of obstacles detected by the UAV at varying distances over 5000 episodes and the last 100 episodes, respectively, utilizing the aforementioned evaluation strategy. The obstacle avoidance performance of the proposed DMAU is superior to that of the other two



(a) 5000 episodes (b) Last 100 episodes

Fig. 12 Statistical analysis of the distance between UAVs and obstacles using different algorithms



(a) 5000 episodes (b) Last 100 episodes

Fig. 13 Collision probabilities of different algorithms

algorithms in both cases. In particular, during the last 100 episodes, when the training outcome reaches its optimum, UAVs exhibit infrequent encounters with obstacles. The simulation results demonstrate that the obstacle avoidance strategy proposed in the joint optimization method effectively mitigates obstacles.

The collision probability of different algorithms versus the episode number is illustrated in Fig. 13. Specifically, Fig. 13a presents the variation in collision probability over 5000 episodes. It can be observed that without employing PF for obstacle avoidance, there is no significant enhancement in anti-collision performance as the episode number increases. The collision probability of the algorithm employing PF for obstacle avoidance exhibits a conspicuous decreasing trend, with our proposed method demonstrating a faster rate of decrease compared to the other three methods. Figure 13b illustrates the collision probability over the last 100 episodes. It is evident that our proposed algorithm achieves a 100% success rate in obstacle avoidance, while the two algorithms without PF exhibit higher collision probabilities.

The expected outage time, mission completion time, and weighted sum of the two are calculated in Fig. 14 to assess the connectivity performance of the joint optimization algorithm over the last 100 episodes. The expected outage time follows the order of D3QN, DMAU, and D3QN-PER in decreasing magnitude. This is attributed to the accelerated training efficiency of the PER mechanism in the algorithm, enabling

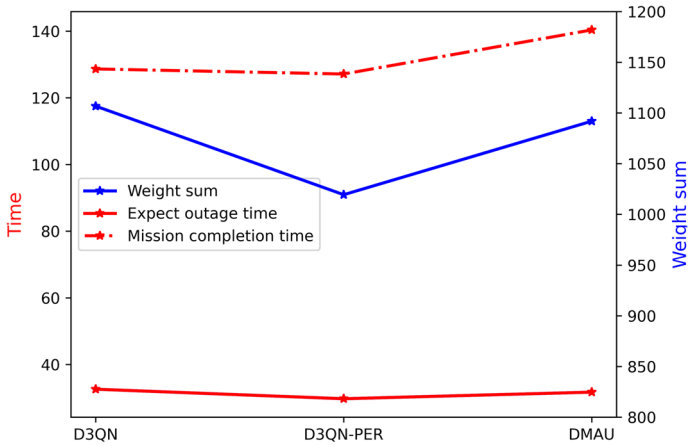


Fig. 14 The weighted summation of the expected outage time and mission completion time by various algorithms

D3QN-PER to discover a path with superior connectivity within a relatively short duration. However, due to the presence of obstacles obstructing the originally optimal connectivity path, UAVs are compelled to choose detours that lead to slightly weaker connectivity paths. Consequently, DMAU exhibits a slightly longer expected outage time compared to D3QN-PER. The implementation of obstacle avoidance inevitably introduces additional flight steps, resulting in a longer mission completion time for the proposed algorithm compared to the other two algorithms without PF obstacle avoidance. In summary, the proposed algorithm ranks second only to D3QN-PER in terms of the weighted sum of expected outage time and mission completion time, thereby demonstrating its effectiveness in ensuring path connectivity.

5 Conclusions

(1) This paper investigates the joint optimization of connectivity, mission completion time, and obstacle avoidance for cellular-connected UAVs through path planning.
(2) We have presented a methodology for constructing a radio map and an environment information map, followed by the creation of a novel optimization function based on PF for joint optimization. Additionally, we propose a DMAU method utilizing D3QN to achieve multi-objective optimization. To enhance learning efficiency, we introduce an advanced PER mechanism. Moreover, we suggest employing radio map and obstacle map networks for simulating UAV flight training, which can expedite the training process, reduce reliance on actual UAV flight data measurements, and yield cost savings.
(3) The numerical results have demonstrated the efficacy of the proposed method in terms of UAV connectivity, mission completion time, and obstacle avoidance, as well as its superior performance compared to alternative approaches. In future research endeavors, our objective is to extend the application of the proposed path planning method to multiple UAVs.

Abbreviations

UAVs	Unmanned aerial vehicles
BS	Base stations
DMAU	Double-map assisted UAV
DRL	Deep reinforcement learning
D3QN	Dueling double deep Q network
PER	Prioritized experience replay
LoS	Line-of-sight
6G	Sixth generation
P2P	Point-to-point
A2G	Air-to-ground
2D	Two-dimensional
3D	Three-dimensional
GUs	Ground users
GBS	Ground base station
SINR	Signal-to-interference-plus-noise ratio
ML	Machine learning
SNARM	Simultaneous navigation and radio mapping
ITU	International Telecommunication Union
CCP	Communication connectivity probability
MDP	Markov decision processes
DNN	Deep neural network
ERT	Experience replay technology
PF	Potential function
Relu	Rectified linear unit
MSE	Mean square error

Acknowledgements
This study was co-supported by the National Natural Science Foundation of China under Grant (No. 62271250), 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).

Author contributions
Zw Z analyzed and proposed the main methods of joint optimization in this paper, XW was the main contributor to writing the manuscript, XL did the work during the simulation, Zp L and FA polished and revised the manuscript, and all authors read and approved the final draft.

Funding
This study was co-supported by the National Natural Science Foundation of China under Grant (No. 62271250), 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).

Availability of data and materials
My manuscript has no associated data.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
No conflict of interest exits in the submission of this manuscript.

Received: 30 December 2023 Accepted: 29 February 2024
Published online: 14 March 2024

References

1. Statista drones: Estimated size of the global commercial drone market in 2021 with a forecast for 2026. Accessed 18 Aug 2011. [Online]. Available: https://www.statista.com/statistics/878018/global-commercial-drone-market-size
2. J. Lyu, R. Zhang, Network-connected UAV: 3-D system modeling and coverage performance analysis. IEEE Internet Things J. **6**(4), 7048–7060 (2019)
3. X. Liu, Y. Yu, F. Li, T.S. Durrani, Throughput maximization for RIS-UAV relaying communications. IEEE Trans. Intell. Transp. Syst. **23**(10), 19569–19574 (2022)
4. X. Liu, Y. Yu, B. Peng, X.B. Zhai, Q. Zhu, V.C.M. Leung, RIS-UAV enabled worst-case downlink secrecy rate maximization for mobile vehicles. IEEE Trans. Veh. Technol. (2022). https://doi.org/10.1109/TVT.2022.3231376

5. X. Liu, Z. Liu, B. Lai, B. Peng, T.S. Durrani, Fair energy-efficient resource optimization for multi-UAV enabled Internet of Things. IEEE Trans. Veh. Technol. **72**(3), 3962–3972 (2023)
6. 3GPP TR 36.873 Study on 3D channel model for LTE. V12.7.0 (2017)
7. J. Qiu, J. Lyu and L. Fu, Placement optimization of aerial base stations with deep reinforcement learning, in *Proceedings of the IEEE International Conference on Communications (ICC)* (2020), pp. 1–6
8. M. M. U. Chowdhury, W. Saad and I. Güvenç, Mobility management for cellular-connected UAVs: a learning-based approach, in *Proceedings of the IEEE International Conference on Communications Workshops* (2020), pp. 1–6
9. C. Zhang, L. Zhang, L. Zhu, T. Zhang, Z. Xiao, X.G. Xia, 3D deployment of multiple UAV -mounted base stations for UAV communications. IEEE Trans. Commun. **69**(4), 2473–2488 (2021)
10. L. Liu, S. Zhang, R. Zhang, Multi-beam UAV communication in cellular uplink: cooperative interference cancellation and sum-rate maximization. IEEE Trans. Wirel. Commun. **18**(10), 4679–4691 (2019)
11. L. Liu, S. Zhang and R. Zhang, Exploiting NOMA for multi-beam UAV communication in cellular uplink, in *Proceedings of the IEEE International Conference on Communications (ICC)* (2019), pp. 1–6
12. W. Mei, R. Zhang, Uplink cooperative NOMA for cellular-connected UAV. IEEE J. Sel. Topics Signal Process. **13**(3), 644–656 (2019)
13. S. Zhang and R. Zhang, Radio map based path planning for cellular-connected UAV, in *Proceedings of the IEEE Global Communications Conference* (2019), pp. 1–6
14. Y.-J. Chen, D.-Y. Huang, Trajectory optimization for cellular-enabled UAV with connectivity outage constraint. IEEE Access **8**, 29205–29218 (2020). https://doi.org/10.1109/ACCESS.2020.2971772
15. S. Zhang, Y. Zeng, R. Zhang, Cellular-enabled UAV communication: a connectivity-constrained trajectory optimization perspective. IEEE Trans. Commun. **67**(3), 2580–2604 (2019)
16. S. Zhang and R. Zhang, Trajectory design for cellular-connected UA V under outage duration constraint, *in Proceedings of the IEEE International Conference on Communications (ICC)* (2019), pp. 1–6
17. E. Bulut and I. Guevenc, Trajectory optimization for cellular-connected UA Vs with disconnectivity constraint, in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)* (2018), pp. 1–6
18. X. Liu, B. Lai, B. Lin, V.C. Leung, Joint communication and trajectory optimization for multi-UAV enabled mobile internet of vehicles. IEEE Trans. Intell. Transp. Syst. **23**(9), 15354–15366 (2022)
19. A. Al-Hourani, S. Kandeepan, S. Lardner, Optimal LAP altitude for maximum coverage. IEEE Wirel. Commun. Lett. **3**(6), 569–572 (2014)
20. M.M. Azari, F. Rosas, K.-C. Chen, S. Pollin, Ultra reliable UAV communication using altitude and cooperation diversity. IEEE Trans. Commun. **66**(1), 330–344 (2018)
21. C. You, R. Zhang, 3D trajectory optimization in Rician fading for UA V-enabled data harvesting. IEEE Trans. Wirel. Commun. **18**(6), 3192–3207 (2019)
22. Y. Zeng, X. Xu, S. Jin, R. Zhang, Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning. IEEE Trans. Wirel. Commun. **20**(7), 4205–4220 (2021)
23. J. Chen, U. Yatnalli and D. Gesbert, Learning radio maps for UAV-aided wireless networks: a segmented regression approach, in *Proceedings of the IEEE International Conference on Communications (ICC)* (2017)
24. U. Challita, W. Saad, C. Bettstetter, Interference management for cellular-connected UA Vs: a deep reinforcement learning approach. IEEE Trans. Wirel. Commun. **18**(4), 2125–2140 (2019)
25. S. Zhang, R. Zhang, Radio map-based 3D path planning for cellular-connected UAV. IEEE Trans. Wirel. Commun. **20**(3), 1975–1989 (2021)
26. X. Wang and M.C. Gursoy, Learning-based UAV trajectory optimization with collision avoidance and connectivity constraints. Available https://arxiv.org/abs/2104.06256
27. H. Xie, D. Yang, L. Xiao et al., Connectivity-aware 3D UAV Path design with deep reinforcement learning. IEEE Trans. Veh. Technol. **70**(12), 13022–13034 (2021)
28. M. Radmanesh, M. Kumar, P.H. Guentert et al., Overview of path planning and obstacle avoidance algorithms for UAVs: a comparative study. Unmanned Syst. (2018). https://doi.org/10.1142/S2301385018400022
29. Y. Huang and Y. Zeng, Simultaneous environment sensing and channel knowledge mapping for cellular-connected UAV, in *2021 IEEE Globecom Workshops (GC Wkshps), Madrid, Spain* (2021), pp. 1–6. https://doi.org/10.1109/GCWkshps52748.2021.9682178.
30. J. Liu, J. Yu, D. Niyato, R. Zhang, X. Gao, J. An, Covert ambient backscatter communications with multi-antenna tag. IEEE Trans. Wirel. Commun. (2023). https://doi.org/10.1109/TWC.2023.3240463
31. B. Hua, H. Ni, Q. Zhu, C.X. Wang, T. Zhou, K. Mao et al., Channel modeling for UAV-to-ground communications with posture variation and fuselage scattering effect. IEEE Trans. Commun. **71**(5), 3103–3116 (2023)
32. K. Mao, Q. Zhu, M. Song, H. Li, B. Ning, G.F. Pedersen et al., Machine learning-based 3D channel modeling for U2V mmwave communications. IEEE Internet Things J. **9**(18), 17592–17607 (2022)
33. M. Shi, K. Yang, D. Niyato, H. Yuan, H. Zhou, Z. Xu, The meta distribution of SINR in UAV-assisted cellular networks. IEEE Trans. Commun. **71**(2), 1193–1206 (2023). https://doi.org/10.1109/TCOMM.2022.3233064
34. J. Pan, N. Ye et al., AI-driven blind signature classification for IoT connectivity: a deep learning approach. IEEE Trans. Wirel. Commun. **21**(8), 6033–6047 (2022)
35. K. Yu, K. Jin and X. Deng, Review of deep reinforcement learning, in *2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China* (2022), pp. 41–48. https://doi.org/10.1109/IMCEC55388.2022.10020015
36. V. Mnih et al., Human-level control through deep reinforcement learning. Nature **518**, 529–533 (2015)

Publisher's Note