2023 IEEE/CIC International Conference on Communications in China (ICCC) | 979-8-3503-4538-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCC57788.2023.10233369

Interpretable and Secure Trajectory Optimization for UAV-Assisted Communication

Yunhao Quan*, Nan Cheng*, Xiucheng Wang*, Jinglong Shen*, Longfei Ma*, and Zhisheng Yin†

*School of Telecommunications Engineering, Xidian University, Xi'an, China

†School of Cyber Engineering, Xidian University, Xi'an, China

Email: {qyh, xcwang_1, jlshen, lfma}@stu.xidian.edu.cn,, dr.nan.cheng@ieee.org, zsyin@xidian.edu.cn

Abstract—**Unmanned aerial vehicles (UAVs) have gained popularity due to their flexible mobility, on-demand deployment, and ability to establish line-of-sight wireless communication. However, existing UAV-assisted communication schemes often overlook the critical issue of collision avoidance during UAV flight. This paper proposes an interpretable UAV-assisted communication scheme that addresses this challenge through decomposition into two sub-problems. The first sub-problem involves constrained UAV coordinates and power allocation, solved using the Dueling Double DQN (D3QN) method. The second sub-problem deals with constrained UAV collision avoidance and trajectory optimization, addressed through the Monte Carlo tree search (MCTS) method. This approach ensures reliable and efficient UAV operation. To enhance the transparency and reliability of system decisions, a scalable explainable artificial intelligence (XAI) framework is proposed. The interpretability of the scheme generates explainable and trustworthy results, facilitating comprehension, validation, and control of UAV-assisted communication solutions. Extensive experiments demonstrate the superiority of the proposed algorithm in terms of performance and generalization compared to existing techniques. The proposed model improves the reliability, efficiency, and safety of UAV-assisted communication systems, offering a promising solution for future applications.**

Index Terms—**UAV network, explainable artificial intelligence, trajectory optimization, collision avoidance**

I. Introduction

In recent times, there has been an upsurge in the popularity of unmanned aerial vehicles (UAVs), due to their remarkable maneuverability, the ability to deploy them as per requirement, and their proficiency in establishing wireless communication links with a high likelihood of line-of-sight (LOS) connectivity [1]. Nonetheless, achieving optimal performance in UAV networks poses a formidable challenge, primarily due to the intricate interplay between the channel gain, which relies on the UAV-user distance, and the transmission power employed by the UAV itself [2], [3]. Furthermore, when contemplating the optimization process, it becomes crucial to acknowledge the criticality of steering clear from UAV collisions, a paramount aspect ensuring the overall security of the network [1].

Henceforth, a substantial body of research has been ardently dedicated to the optimization of UAV trajectory. For example, the work of Zhan et al. [4] delves into an optimization methodology aiming to maximize the energy efficiency of sensor networks by meticulously fine-tuning the UAV trajectory. Similarly, the authors in [5] adopt an optimization stance encompassing trajectory, transmission power, and connection optimization between UAVs and nodes, all with the aim of minimizing the cumulative transmission power within the system. Baek et al. [6] have traversed the domain of UAV trajectory and route design, employing a hovering flight model as the foundation for modeling UAV behavior. Moreover, a cluster of researchers in [7], [8] have put forth proposals concerning UAV trajectory optimization schemes, meticulously crafted to minimize energy consumption or, alternatively, extend the overall duration of UAV flight.

In addition to the aforementioned optimization paradigms, the imperative of collision avoidance looms large, given its intrinsic significance in ensuring the integrity and dependability of UAV networks. Consequently, the notion of jointly optimizing UAV trajectory, resource allocation strategy, and collision avoidance strategy has gained substantial traction as a viable solution. Towards this pursuit, Yang et al. [9] bring forth an optimization paradigm underpinned by the deep deterministic policy gradient (DDPG) algorithm, which endeavors to maximize energy efficiency through a cohesive optimization of UAV trajectory, resource allocation strategy, and interference strategy. Zhang et al. [10] harness the deep Q-network (DQN) method to conjointly craft UAV transmission scheduling, power allocation, and trajectory optimization, ultimately seeking to maximize the system transmission rate. As for Liu et al. [11], they adopt a multi-agent deep deterministic policy gradient (MADDPG) method, replete with a convolutional neural network (CNN) to extract pertinent features. This particular methodology excels in the joint optimization of UAV operational trajectory and collision avoidance. By leveraging deep reinforcement learning techniques, these endeavors effectively address the vexing concern of collision avoidance during UAV service operations. However, a lingering issue surrounding the interpretability of these approaches raises legitimate concerns pertaining to UAV safety, potentially culminating in avoidable legal disputes [12]. Consequently, prioritizing interpretability and the reliability of decision-making processes assumes critical importance when devising algorithms tailored to UAV operations.

In a bid to augment the overall performance of UAV-assisted communication networks, in this paper, a joint optimization approach encompassing trajectory and power allocation is set forth, all while taking into account the imperatives of collision avoidance. Furthermore, a cutting-edge architecture grounded in the principles of explainable artificial intelligence (XAI) takes center stage, affording an efficient means of grappling with this multifaceted challenge. The main contribution of this paper are as follows.

1) A scalable framework based on explainable artificial intelligence (XAI) is proposed, which serves as the bedrock for effectively optimizing the flying trajectory and power allocation, all the while ensuring robust collision avoidance. The trustworthiness and interpretability of the proposed XAI methodology bolster its credibility and efficacy in addressing this complex problem.
2) The joint optimization problem is judiciously decomposed into two sequential sub-problems. In the first phase, a power allocation and service coordinate conundrum, ensconced within collision avoidance constraints, finds an optimal resolution through the adept utilization of a Double Dueling DQN (D3QN)-based approach. Subsequently, the trajectory optimization quandary, also encumbered by collision avoidance restrictions, is masterfully resolved using a Monte Carlo tree search (MCTS)-based method. Notably, the entire process adheres to the tenets of explainability, further enhancing the trustworthiness of the outcomes.
3) The simulation results show the proposed method, not only in terms of overall performance but also in terms of generalization capabilities. Furthermore, the tree search method unveils the decision paths followed during the search process, thereby substantially augmenting the interpretability of our algorithm and instilling confidence in its decision-making mechanisms.

II. System Model and Problem Formulation

As depicted in Fig. 1, we consider a UAV-assisted communication network comprised of multiple users and UAVs. The set of users served by the UAVs is denoted by $k \in \mathbb{K}=\{1,2,3,\ldots,K\}$. During flight, the altitude of the UAV is denoted by $h(t)$, while its distance from the k-th user at time t is expressed as $d_k(t)$, computed as

$$d_k\left(t\right)=\sqrt{h_u^2\left(t\right)+\left[x_u\left(t\right)-x_k\left(t\right)\right]^2+\left[y_u\left(t\right)-y_k\left(t\right)\right]^2}. \tag{1}$$

The average path loss between the UAV and the k-th user can be expressed as

$$L_k(t)=P_{\mathrm{LoS}}\cdot L_{\mathrm{LoS}}+P_{\mathrm{NLoS}}\cdot L_{\mathrm{NLoS},} \tag{2}$$

where P_{LoS} and P_{NLoS} are the probabilities of line-of-sight and non-line-of-sight conditions, respectively.

Considering small-scale fading, the channel gain between the UAV and user k at time t, denoted by $g_k(t)$, can be calculated as

$$g_k(t)=H_k(t)\cdot 10^{-L_k(t)/10}, \tag{3}$$

where $H_k(t)$ is the channel fading coefficient. The variable $v_k(t)$ serves as a performance metric, with a value of 1 indicating that the UAV is serving the k-th user and 0 otherwise. The power allocated to user k is given by $p_k(t)$, and the data rate between the UAV and user k is represented as

$$R_{(k)}(t)=B\log 2\left(1+\gamma_{(k)}(t)\right). \tag{4}$$

Here, $\gamma_k(t)$ is the signal-to-noise ratio (SNR) of the channel between the k-th user and the UAV, which is calculated as

$$\gamma_k(t)=\frac{v_k(t)g_k(t)P_k(t)}{\sum_{i=1,i\neq k}^{K}v_k(t)g_k(t)P_k(t)+\sigma_k(t)^2}. \tag{5}$$

with $\sigma_k(t)$ representing the white Gaussian noise. The communication bandwidth of the UAV is given by B. The overall rate of the system is calculated as

$$R(t)=\sum_{k=1}^{K}\mathcal{R}_{(k)}(t). \tag{6}$$

The throughput of the system over time T is represented as

$$R=\sum_{t=0}^{T}\mathcal{R}(t). \tag{7}$$

where $\mathcal{R}(t)$ is the instantaneous rate at time t. The location coordinates of the UAV during service time, $\mathbf{H}=\{h(t),x(t),y(t),0\leq t\leq T\}$, are used to denote its movement trajectory. The power allocated by the UAV to each user is represented by $\mathbf{P}=\{p_k(t),0\leq t\leq T,k\in\mathbb{K}\}$, and the connectivity between users and the UAV is quantified using $\mathbf{V}=\{v_k(t),0\leq t\leq T\}$. The number of steps taken by the UAV at time t is represented by S_t with S_{max} being the maximum number of steps that the UAV can fly. The specific control action executed by the UAV during flight, such as its movement trajectory or any adjustments made to maintain a stable position in the air, are denoted by $\mathbf{U}=\{u_k(t),0\leq t\leq T\}$. $\mathcal{C}\left(H,U\right)$ represents the collision statistics function. With the objective of maximizing system throughput and minimizing collision probability, subject to constraints on maximum power, spatial limitations, and Quality of Service (QoS) requirements, the problem of reliable service provision by UAVs can be formulated as follows

$$\max_{\mathbf{H},\mathbf{V},\mathbf{P},\mathbf{U}}\mathcal{G}=\sum_{t=0}^{T}(R(t)-\mathcal{C}(H,U)), \tag{8}$$

$$\text{s.t.}\quad h_{\min}\leq h(t)\leq h_{\max},\forall t\in[0,T], \tag{8a}$$

$$x_{\min}\leq x(t)\leq x_{\max},\forall t\in[0,T], \tag{8b}$$

$$y_{\min}\leq y(t)\leq y_{\max},\forall t\in[0,T], \tag{8c}$$

$$\sum_{k\in\mathbb{K}}v_k(t)P_k\leq P_{max},\forall t\in[0,T],\forall k\in\mathbb{K}, \tag{8d}$$

$$S_t\leq S_{max},\forall t\in[0,T], \tag{8e}$$

$$R_k(t)\geq R_{\mathrm{Qos}},\forall t\in[0,T],\forall k\in\mathbb{K}. \tag{8f}$$

It should be noted that the optimization problem described above is a mixed exponential non-convex problem, which is known to be an NP hard problem. Furthermore, in the scenario under consideration, both large-scale fading and small-scale fading are dependent on the instantaneous position of the UAV and users, making it difficult to solve the optimization problem using traditional optimization methods. Therefore, sub-problem decomposition and reinforcement learning have

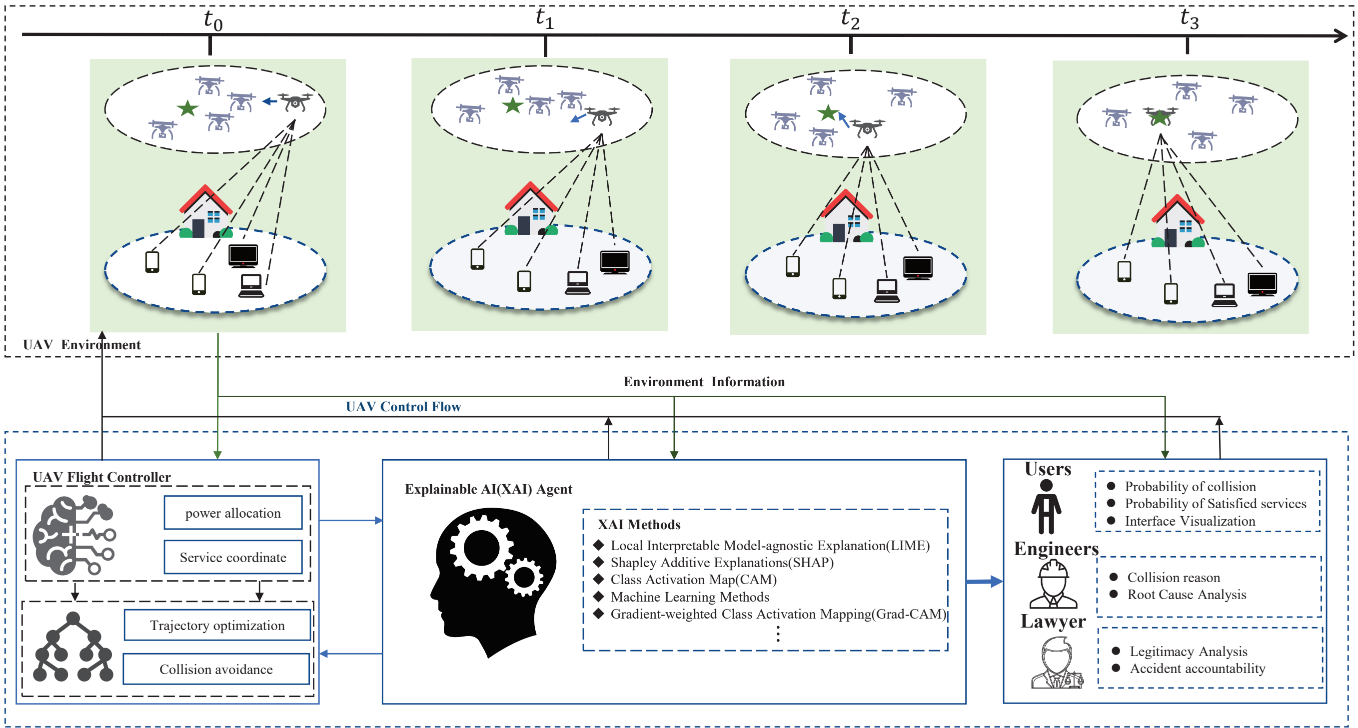

Fig. 1. Scalable and interpretable artificial intelligence framework based on UAV-assisted communication.

proven to be effective methods for dealing with complex control problems in high-dimensional continuous spaces. In the next section, we adopt the idea of sub-problem decomposition to decompose the problem and solve it using reinforcement learning and MCTS methods, while also designing a corresponding XAI framework, as shown in Fig. 1.

III. UAV-Assisted Communication Methods

The original problem was decomposed into two sub-problems to reduce its complexity. The first sub-problem involves power allocation and coordinate solving, while the second sub-problem involves trajectory optimization and collision avoidance.

A. coordinate and power allocation

UAV service coordinate solving and power allocation problems can be expressed as

$$\max_{\mathbf{H},\mathbf{V},\mathbf{P}} \mathcal{R} = \sum_{t=0}^{T} \mathcal{R}(t), \tag{9}$$

$$\text{s.t.} \quad h_{\min} \le h(t) \le h_{\max}, \forall t \in [0, T], \tag{9a}$$

$$x_{\min} \le x(t) \le x_{\max}, \forall t \in [0, T], \tag{9b}$$

$$y_{\min} \le y(t) \le y_{\max}, \forall t \in [0, T], \tag{9c}$$

$$\sum_{k \in \mathbb{K}} v_k(t) P_k \le P_{max}, \forall t \in [0, T], \forall k \in \mathbb{K}, \tag{9d}$$

$$R_k(t) \ge R_{\text{Qos}}, \forall t \in [0, T], \forall k \in \mathbb{K}. \tag{9e}$$

The D3QN reinforcement learning algorithm, which utilizes two neural networks to fit the state and action values, as well as an additional layer to estimate the advantage values of each action, is used to solve the problem. The Q-value for each action at each time step is calculated based on the average advantage value of other actions and the action's Q-value. Algorithm 1 provides a more detailed description of the algorithm.

- **Action Space:** The action space is a vector of size $K \times 6$ consisting of the UAV's moving direction and the power allocated to each user. The UAV has seven available movement options, which include moving left, right, forward, backward, ascending, descending, or remaining stationary. Additionally, the sum of all power allocation values must be within the power constraint limit.
- **State:** The state space includes the UAV's three-dimensional position and the channel gain between the UAV and the users.
- **Reward:** To maximize the overall throughput, we design the reward function as follows, where λ represents the penalty factor.

$$R = \frac{\mathcal{R}(t)}{2^{\lambda}}. \tag{10}$$

In the D3QN model, the evaluation network first receives the abstract state information from the connected UAV and users to determine the optimal action. The reward value is calculated next, and the corresponding action executed in the environment. Upon the completion of a UAV-terminal user pair's service, we compute the data rate for that specific period.

Algorithm 1 D3QN algorithm for UAV service coordinates solution

for each episode **do**
  Initialize initial positions of UAV and users
  Initialize the network parameter θ
  Update ϵ in action policy
  for each step $t_0 \leq t \leq t_0 + T_r$ **do**
    Calculate $g_k(t)$
    Generate state abstraction array s
    Choose A according to action policy and $Q(s, a, \theta)$
    Take action a,observe r and $s^{'}$
    Store $D = (s, s^{'}, r, a)$
    Sample random mini-batch of transitions (s_j, a_j, r_j, s_{j+1}) from D
    Set $y_j = r_j + \gamma \max_{a'} \hat{Q}(s_{j+1}, a'; \theta^{-})$
    Update the action-value function using gradient descent: $\Delta\theta = \alpha(y_j - Q(s_j, a_j; \theta))\nabla_\theta Q(s_j, a_j; \theta)$
  end for
end for

B. trajectory optimization and collision avoidance

The trajectory optimization and collision avoidance problems can be formulated as

$$\min_{\mathbf{H},\mathbf{U}} \mathcal{C} = \sum_{t=0}^{T} \mathcal{C}(H, U), \tag{11}$$

$$\text{s.t. } h_{\min} \leq h(t) \leq h_{\max}, \forall t \in [0, T], \tag{11a}$$

$$x_{\min} \leq x(t) \leq x_{\max}, \forall t \in [0, T], \tag{11b}$$

$$y_{\min} \leq y(t) \leq y_{\max}, \forall t \in [0, T], \tag{11c}$$

$$S_t \leq S_{max}, \forall t \in [0, T], \tag{11d}$$

We addressed the issue by treating it as a Markov Decision Process (MDP) problem. The position coordinates and velocity of the kth intruder are represented as ($p_x^{(k)}$,$p_y^{(k)}$) and ($v_x^{(k)}$,$v_y^{(k)}$), respectively. Similarly, the position coordinates and velocity of the ownership are represented as (o_x, o_y) and (v_x, v_y). The ownership's heading and tilt angles are represented by A_ψ and A_ϕ. We solved the problem using the MCTS method.

Action space: At the beginning of each time step, the target aircraft adjusts its tilt angle and acceleration at a certain rate. $\mathcal{A}_\phi$ and $\mathcal{A}_a$ represent the directional and acceleration action spaces, respectively. $\mathcal{A}_\phi$ consists of three actions: left turn, straight, and right turn, while $\mathcal{A}_a$ includes three actions: speed up, slow down, and maintain constant speed.

Termination state: In consideration of safety, we define d_{min} as the minimum collision distance between two UAVs. When the distance between two UAVs becomes less than d_{min}, it results in a collision. There are three kinds of termination states for the entire process

1) Collision, which occurs if the distance between the intruder and the ownership is less than d_{min}.
2) Time out: if the ownership leaves the defined map or is unable to reach its destination within the specified steps.
3) Goal state: if the ownership reaches its set destination.

In MCTS, the nodes of the search tree correspond to states in the state space. The leaf nodes of the tree represent all possible subsequent states that can occur from performing different actions on the current state. As each time step involves 9 action spaces, a node can have up to 9 leaf nodes. The MCTS algorithm selects actions by carrying out a forward search of the search tree. Each edge (s, a) in the tree stores an action value $Q(s, a)$ and its number of visits $N(s, a)$. The tree is traversed by simulating from the root node (the initial state). The MCTS algorithm can be divided into four steps:

1) The ownership will select the leaf node with the highest value according to Equation (12), which maximizes the sum of the average action value and the uncertainty reward.

$$UCT = \bar{X}_j + 2C\sqrt{\frac{2\ln n}{n_j}}. \tag{12}$$

The variable $\bar{X}_j$ approximately represents the state-action value of the child node, $UCT = 2C\sqrt{\frac{2\ln n}{n_j}}$ known as the exploration term, n_j represents the number of times child node j has been visited, and n represents the number of times the parent node has been visited. C is a constant that balances exploration and exploitation. If multiple child nodes have the same maximum value, the leaf node will be randomly selected. If a child node has never been visited, it will be prioritized, ensuring that each leaf node is visited at least once.
2) At the point when the ownship enters a new node (state) that it has not visited before, a new child node is created in the search tree as a subnode of the previous state, or parent node. The visit count of the new node is initially set to 1, and the cumulative reward value is initialized to 0.
3) Ordinarily, a large number of iterations are required by the conventional approach to reach a termination state by following a random policy and determine the corresponding ultimate reward score. This leads to high time complexity. we leverage the value function estimation method to overcome this limitation. This approach sets the iteration's search depth and employs the value function to compute the final reward. From a subjective perspective, a state where the drone is approaching the destination without any collision is considered to be better. Based on this, we utilize the estimation function shown in Equation (13) for non-termination states.

$$\tilde{V}(s) = 1 - \frac{d(o, g)}{\max d(o, g)}, \quad \text{if } s \text{ is non-terminal state} \tag{13}$$

The distance between the ownship and its goal is constant and equal to the diagonal length of the map. If there are no collisions with other drones or boundaries, the ownship receives a reward whose magnitude is determined by the distance between itself and the goal.

Specifically, the closer the ownship gets to the goal, the higher the reward it receives.

4) The process of updating the final reward and visit count for all traversed edges is called backpropagation. Once the termination state is reached through the value function estimation function described earlier, each traversed edge's final reward and visit count are updated. As the ownship transverses each edge, the edge is updated with a reward increment while counting the number of visits. The reward value for each edge is determined by dividing its accumulated reward by its visit count.

A single simulation consists of executing the four steps described earlier once. To improve decision accuracy, we perform a large number of simulations.

C. XAI Framework

In UAV-assisted communication scenarios, strict reliability requirements must be met in addition to the key connection requirements for high-speed and stable data transmission. This section proposes a scalable XAI framework that takes into account the characteristics of UAV trajectory optimization and collision avoidance, as depicted in Fig. 1. During flight, UAVs gather information about their surroundings and take appropriate control measures. Real-time environment data is transmitted to both the UAV flight controller and the XAI agent. The proposed framework integrates scalable XAI methods that increase confidence in decision-making for artificial intelligence systems. The framework is scalable, incorporating location and velocity information of UAVs and surrounding aircraft as part of the environment data. XAI methods within the framework use the MCTS method.

Fig. 1 illustrates the types of questions that the UAV may present, and which XAI can address. Scalable XAI methods can improve wireless network service quality and enhance fault detection efficiency for service providers. Engineers can easily detect decision-making errors. XAI can also provide flight decision-making details for individual users, increasing their trust. Lastly, for legal regulators, XAI can explain model decisions in a quantifiable manner, establishing trust.

IV. Simulation Result

A. UAV coordinate and power allocation

To simulate UAV service coordinates and power allocation, we randomly distribute users within the service area and deploy the UAV near the initial height boundary of 100 meters. The UAV's flight range is 500 meters, with a width of 500 meters, and we employ a neural network with three layers and 40 hidden nodes. The activation function used is a rectified linear unit. The Adam optimizer is used to train the neural network.The greedy action strategy ϵ is set to linearly decrease from 0.9 to 0.1.

• **DQN**: The traditional Q-value coupled DQN inputs state information and outputs the action value of each action in this state.

• **Random**: The random method is a traditional approach for solving problems, which involves randomly choosing points within a specified area and computing the corresponding values at those points.

• **D3QN**: D3QN introduces double and dueling improvements, where the input of D3QN is state information and the output is the action value and advantage value of each action in this state.

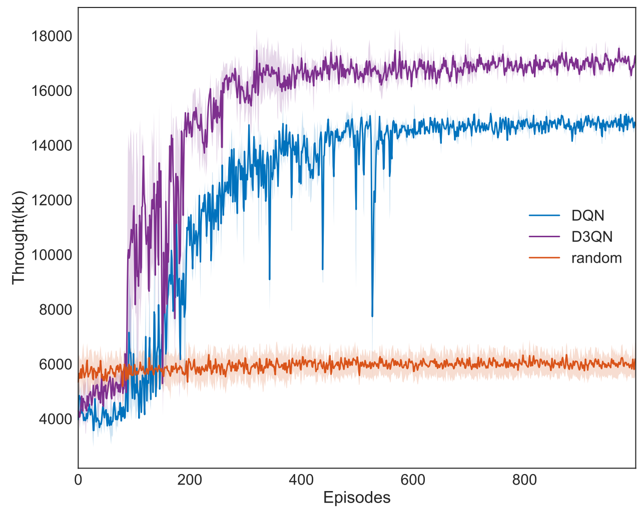

Fig. 2. Convergence performance of different algorithm with $K = 10$.

Fig. 2 illustrates the convergence of the proposed D3QN algorithm. It can be observed that the D3QN algorithm requires approximately 300 episodes to converge, which is significantly less than the number of episodes required for the DQN algorithm to converge. Furthermore, Fig. 2 shows that the D3QN algorithm is able to converge to a performance of around 17000, which is significantly greater than the convergence value of approximately 14000 achieved by the DQN algorithm. Overall, the results presented in Fig. 2 demonstrate the superior convergence performance of the D3QN algorithm compared to the DQN algorithm and Random algorithm.

B. trajectory optimization and collision avoidance

In this section, we use the UAV service coordinates obtained in the previous section as the goal of the task. Intruders are randomly distributed within an area with a length and width of 2000 meters. reward is set according to Equation(14), d_{min} = 50.

$$R(s) = \begin{cases} 1, \text{ if } s \text{ is goal state} \\ 0.1, \text{ if } s \text{ is time-out state} \\ 0, \text{ if } s \text{ is collision state} \end{cases} \tag{14}$$

• **DQN**: The traditional Q-value coupled DQN is trained in an environment with a fixed number of UAVs (using the states of all surrounding intruders as inputs)

• **Safe-DQN**: A safety-aware DQN model consists of two DQNs: one ensures the UAV reaches its goal safely, while the other guarantees that the UAV does not collide with other intruders [13].

•**Tree-fast**: A fast Monte Carlo Tree search method with low steps per iteration [14].

•**Tree-depth**: Our proposed Monte Carlo Tree search method with a large number of steps per iteration and a search depth of 3 or 4.

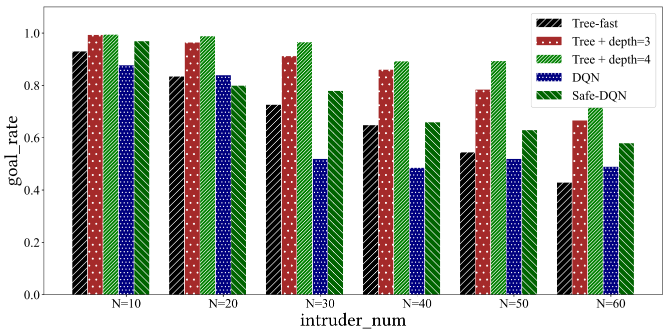

Fig. 3. Performance goal rates of different algorithms with varying numbers of intruders.

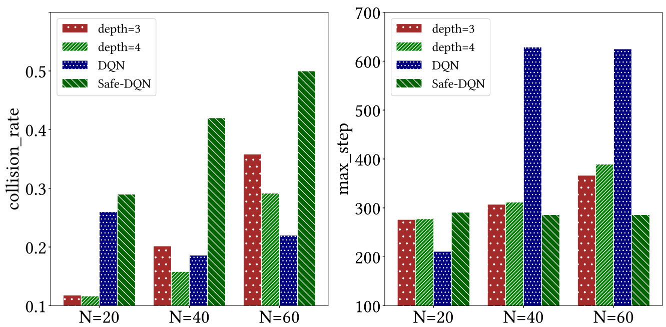

Fig. 4. Collision rate and number of steps for different algorithms with varying numbers of intruders.

We evaluated the performance of DQN, Safe-DQN, Tree-fast, and Tree-depth in trajectory optimization and collision avoidance under different intruder numbers, as depicted in Fig. 3. As the number of intruders increases, the MCTS algorithm can consistently sustain optimal performance compared to other methods. Additionally, as depicted in Fig. 4, the MCTS algorithm maintains its overall performance while exhibiting a lower collision probability and shorter execution steps as compared to other algorithms. Notably, when the number of intruders increases, the algorithm's ability to generalize its performance is superior to other algorithms.

V. Conclusion

In conclusion, this research paper has presented a remarkable trajectory optimization solution for UAV-assisted communication. By addressing the challenge of ensuring reliable UAV services, the problem has been dissected into two crucial sub-problems. The first sub-problem tackles constrained UAV coordinate and power allocation, effectively determining optimal coordinates considering spatial constraints, service quality thresholds, and power constraints. This optimization process has enhanced reliability and efficiency in UAV operations. The second sub-problem focuses on constrained UAV collision avoidance and trajectory optimization, employing a meticulously crafted framework that prioritizes reliability and security. This approach mitigates collision risks and ensures steadfast performance. Additionally, a scalable and comprehensive XAI framework has revolutionized decision-making, fostering transparency, trust, and reliability in UAV collision avoidance and trajectory optimization processes. By using the proposed method in UAV networks, the performance can be significantly increased while guaranteeing trustworthy collision avoidance. Future research aims to explore interpretability in complex domains such as the Internet of Vehicles and mobile communication.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (2020YFB1807700), the National Natural Science Foundation of China (NSFC) under Grant No. 62071356.

References

[1] N. Cheng, S. Wu, X. Wang, Z. Yin, C. Li, W. Chen, and F. Chen, "Ai for uav-assisted iot applications: A comprehensive review," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[2] N. Cheng, F. Lyu, W. Quan, C. Zhou, H. He, W. Shi, and X. Shen, "Space/aerial-assisted computing offloading for iot applications: A learning-based approach," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 5, pp. 1117–1129, 2019.

[3] X. Wang, L. Fu, N. Cheng, R. Sun, T. Luan, W. Quan, and K. Aldubaikhy, "Joint flying relay location and routing optimization for 6g uavndash;iot networks: A graph neural network-based approach," *Remote Sensing*, vol. 14, no. 17, 2022.

[4] C. Zhan, Y. Zeng, and R. Zhang, "Energy-efficient data collection in uav enabled wireless sensor network," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 328–331, 2018.

[5] S. Fu, Y. Tang, Y. Wu, N. Zhang, H. Gu, C. Chen, and M. Liu, "Energy-efficient uav-enabled data collection via wireless charging: A reinforcement learning approach," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 10 209–10 219, 2021.

[6] J. Baek, S. I. Han, and Y. Han, "Energy-efficient uav routing for wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1741–1750, 2020.

[7] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing uav," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, 2019.

[8] Z. Wang, L. Duan, and R. Zhang, "Adaptive deployment for uav-aided communication networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 9, pp. 4531–4543, 2019.

[9] P. Yang, X. Cao, X. Xi, W. Du, Z. Xiao, and D. Wu, "Three-dimensional continuous movement control of drone cells for energy-efficient communication coverage," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 6535–6546, 2019.

[10] B. Zhang, C. H. Liu, J. Tang, Z. Xu, J. Ma, and W. Wang, "Learning-based energy-efficient data collection by unmanned vehicles in smart cities," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1666–1676, 2018.

[11] C. H. Liu, Z. Chen, and Y. Zhan, "Energy-efficient distributed mobile crowd sensing: A deep learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1262–1276, 2019.

[12] C. Li, L. Guan, H. Wu, N. Cheng, Z. Li, and X. S. Shen, "Dynamic spectrum control-assisted secure and efficient transmission scheme in heterogeneous cellular networks," *Engineering*, vol. 17, pp. 220–231, 2022.

[13] L. Wang, H. Yang, Y. Lin, S. Yin, and Y. Wu, "Explainable and safe reinforcement learning for autonomous air mobility," 2022.

[14] X. Yang and P. Wei, "Autonomous free flight operations in urban air mobility with computational guidance and collision avoidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5962–5975, 2021.