photonics

MDPI

Article

Adaptive Resource Allocation for Emergency Communications with Unmanned Aerial Vehicle-Assisted Free Space Optical/Radio Frequency Relay System

Yuanmo Lin [1,2], Yuxun Ai [2], Zhiyong Xu [1,*], Jingyuan Wang [1] and Jianhua Li [1]

[1] College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China; linyuanmo@ptu.edu.cn (Y.L.); wjywjy2011@163.com (J.W.); ljhice@163.com (J.L.)
[2] College of Mechanical Electrical and Information Engineering, Putian University, Putian 351100, China; aiyuxun1492@ptu.edu.cn
* Correspondence: njxzy123@163.com

Abstract: This paper investigates the problem of coordinated resource allocation for multiple unmanned aerial vehicles (UAVs) to address the scarcity of communication resources in disaster-affected areas. UAVs carrying modules of free space optical (FSO) and radio frequency (RF) serve as relay nodes and edge offloading nodes, presenting an FSO/RF dual-hop framework. Considering the varying urgency levels of tasks, we assign task priorities and transform the proposed problem into distributed collaborative optimization problem. Based on the K-means algorithm and the multi-agent deep deterministic policy gradient (MADDPG) algorithm, we propose a UAV-coordinated K-means MADDPG (KMADDPG) to maximize the number of completed tasks while prioritizing high-priority tasks. Simulation results show that KMADDPG is 5% to 10% better than the benchmark DRL methods in convergence performance.

Keywords: unmanned aerial vehicles; emergency communication; task priority; free space optical/radio frequency; multi-agent DDPG

check for updates

Citation: Lin, Y.; Ai, Y.; Xu, Z.; Wang, J.; Li, J. Adaptive Resource Allocation for Emergency Communications with Unmanned Aerial Vehicle-Assisted Free Space Optical/Radio Frequency Relay System. *Photonics* **2024**, *11*, 754. https://doi.org/10.3390/photonics11080754

Received: 9 July 2024
Revised: 5 August 2024
Accepted: 6 August 2024
Published: 13 August 2024

1. Introduction

Natural disasters often result in severe damage to ground communication infrastructure, such as base stations, leading to significant communication challenges and hindering emergency response efforts. These events create a critical need for efficient and rapid deployment of temporary communication networks to facilitate rescue operations and mitigate losses [1]. Infrastructures in affected areas may be destroyed, causing traditional communication networks to malfunction. A primary issue caused by natural disasters is resource scarcity, which also applies to communication networks. The efficient utilization of network resources is top priority. To prevent the losses from network collapse, researchers are focusing on rapidly establishing temporary networks [2–4].

Due to their flexibility, low cost, and independence of ground conditions, UAVs are widely used in smart city applications such as urban sensing and delivery [5,6]. UAV-assisted communication networks do not rely on ground infrastructure and can be quickly established as temporary networks using UAVs and FSO communication. FSO communication offers high bandwidth, high speed, and the advantage of no spectrum licensing by transmitting data through the air using visible or infrared light [7]. Its immunity to electromagnetic interference and highly focused beam transmission ensures communication security [8]. The rapid deployment capability as well as low installation and maintenance costs of FSO systems make them perform excellently in emergency communication recovery and complex geographical environments, particularly for urban network interconnection, temporary communication facilities, military communication, and post-disaster emergency

communication. Therefore, under bandwidth constraints, adopting FSO technology significantly improves the resource utilization of communication systems [9]. It gradually becomes indispensable in network construction.

However, under adverse weather conditions, the FSO backhaul will be severely affected, reducing the reliability of FSO links [10]. In contrast, RF technology can be used in various environments, unaffected by weather, and is suitable for both indoor and outdoor communication scenarios [11]. To handle rare foggy conditions, it is imperative to quickly build emergency temporary networks instead of constructing expensive permanent parallel RF links [12]. Thus, UAVs can be employed to serve some users, reducing the load on GBSs. We therefore considered a dual-hop FSO/RF model, constructing two channels: from mobile users (MUs) to UAVs and from UAVs to GBSs.

When UAV-assisted communication networks operate in disaster areas, providing equal service to all user devices is unreasonable [13]. When users make communication requests and emergency rescue requests, their urgency levels are different. A more reasonable strategy is for UAVs to prioritize higher-urgency tasks, so we need to incorporate priority levels to distinguish different tasks [14].

To further address the limited terminal computing power and meet MU's Quality-of-Service (QoS) requirements, we consider using Mobile Edge Computing (MEC) technology in emergency communication. In this mode, MUs can offload computationally intensive tasks to nearby GBSs, reducing processing delays and saving the energy consumption of user devices [15,16]. Specifically, to prevent the base station from being overloaded with tasks, in proposed application scenarios, MUs can offload tasks to UAVs for computation, leveraging UAVs' mobility to mitigate GBS load [17].

Considering the implementation of UAVs autonomously performing edge offloading and resource allocation, we introduce deep reinforcement learning (DRL) technology. Deep multi-agent reinforcement learning has become a key role in the field of multi-UAV-assisted communications [18,19]. With the increase in the number of UAV and task complexity, traditional control and optimization methods struggle to cope with complex environments and dynamic demands. DRL enables multiple UAVs to act as agents for collaborative learning and decision-making, achieving adaptive and efficient communication network management in dynamic and uncertain environments [20,21].

This study explores the challenges of edge offloading and resource allocation within UAV-supported emergency communication systems. To enhance convergence speed, we incorporate the K-means algorithm into multi-agent deep reinforcement learning. This approach allows each agent to share locally observed user data during the training phase of action neural networks, jointly processing it and moving to a better positiaon. The key contributions of this research are outlined as follows:

(1) A mathematical model for emergency communication scenarios is formulated, where several UAVs were used as offloading nodes for MEC or as access relays, facilitating the connection between MUs and GBSs. Specifically, we described the scenario, RF model, FSO model, and delay model of emergency communication.
(2) We propose a UAV-assisted resource allocation method known as K-means MADDPS (KMADDPG), which aims to maximize the number of successful tasks while prioritizing high-priority ones. The proposed algorithm builds on the MADDPG algorithm, integrating it with the K-means algorithm to handle the high data dimensionality in the above-mentioned scenarios. For different tasks with different urgency levels, we incorporate a priority mechanism for mobile users.
(3) We examine the time complexity of the proposed algorithm and assess its performance in emergency communication settings through simulation studies. The results indicate that the proposed KMADDPG effectively optimizes communication resource allocation in regions affected by disasters with compromised communication infrastructures. Additionally, extensive simulations reveal that KMADDPG surpasses several baseline methods regarding convergence speed and the number of successful tasks.

The structure of the rest of this paper is as follows: Section 2 covers a review of related work. Section 3 describes the system model. In Section 4, we introduce the KMADDPG for resource allocation and edge offloading. Section 5 provides and discusses the simulation results. Finally, conclusions are drawn in Section 6.

2. Related Work

The FSO/RF dual-hop model is a feasible approach to establishing a stable, cost-effective, and rapid heterogeneous network. Known for its extensive transmission range and high bandwidth, the FSO/RF dual-hop model has garnered significant research attention in recent years. For instance, Pang et al. [22] initially proposed an optical IRS-assisted dual-hop hybrid FSO and RF system for cloud radio access networks (C-RANs). They introduced polarization codes in the FSO link to mitigate signal fading and achieve optimal data rates for the RF link. Wang et al. [23] employed gamma–gamma distribution to characterize the atmospheric turbulence effects on the FSO link from UAV to HAP and proposed a novel RIS-assisted UAV secure multi-user FSO/RF system. Lee et al. [24] explored the forwarding of packets between ground terminals and backhaul through multi-HAPS relays in a dual-hop FSO/RF network.

The allocation of scarce communication resources is a critical issue that merits discussion. Che et al. [25] enhanced system-level energy efficiency by jointly designing FSO and RF links and optimizing UAV altitude using power allocation techniques. Qi et al. [26] proposed a UAV-assisted vehicular communication network, utilizing DRL for resource allocation to enhance UAV energy efficiency and ensure QoS. MEC significantly enhances the computational performance of MUs with constrained resources. In the study by Jiang et al. [27], a new distributed DRL framework utilizing multi-agent systems is introduced to reduce latency and energy usage in solving optimization challenges within large-scale MEC systems.

Most studies have explored mechanisms for task prioritization. For instance, Qin et al. [28] addressed the issue of time-varying priorities for reconnaissance tasks, examining task selection and scheduling within UAV-enabled multi-access edge computing for reconnaissance. Liu et al. [29] focused on prioritizing industrial equipment and developed a dynamic priority multi-channel access algorithm using DRL. However, they have yet to apply the integration of prioritization and DRL within the domain of UAV.

Researchers have introduced multi-agent reinforcement learning and integrated it with UAV-assisted networks to enable UAV autonomous decision-making [30]. For instance, a UAV-assisted communication network with edge offloading capabilities was developed, and the MADDPG algorithm was proposed to minimize service delay in internet of vehicle (IoV) task processing [31]. In contrast to previous optimization goals, a similar model was constructed, and the MADDPG algorithm was utilized to optimize UAV trajectory design and offloading strategy in a 3D environment [32]. Lee et al. [33] investigated the use of deep reinforcement learning to maximize communication efficiency in the context of hybrid FSO/RF communication models. Guan et al. [34] studied the cooperative trajectory optimization of multiple UAVs in the hybrid FSO/RF communication model to maximize the achievable rate for mobile users.

The above research has made significant contributions to the field of UAV-assisted communication networks. Studies such as [22–24] primarily focus on constructing FSO/RF models. We have incorporated reinforcement learning to train UAVs for autonomous decision-making within the FSO/RF system model. In contrast to [25,26], our focus is on resource allocation for bandwidth and CPU frequency. Our objective, distinct from the aforementioned studies [31–34], is to maximize both the number of successful tasks and the number of high-priority tasks. In this article, we apply a multi-agent reinforcement learning scheme to achieve resource allocation, offloading control, and task prioritization in MEC, enabling UAVs to learn cooperative behaviors.

3. System Model

3.1. Scenario

This paper proposes an application scenario of a multi-UAV-assisted heterogeneous wireless network. As shown in Figure 1, we define that there are J MUs scattered on the plane, following a Poisson distribution, denoted by a set $\mathcal{J} = \{1, \cdots, J\}$. A fleet of UAVs hover above the MUs, serving as aerial base stations for relay or as nodes for edge computing, receiving data transmitted by MUs. The height of the UAVs is denoted by H and the UAVs are represented by a set $\mathcal{I} = \{1, \cdots, I\}$. GBSs are located at the edges of the plane, represented by a set $\mathcal{G} = \{1, \cdots, G\}$. The main notations are listed in Table 1. This model can be considered as a two-layer model, since we considered the transmission channel between the UAV and the MU is different from the transmission channel to the GBSs. Communication between UAVs and MUs uses traditional RF communication, while communication between UAVs and GBSs uses FSO communication. The positions of MUs and UAVs are represented by $Q_j^{MU} = [x_j, y_j, 0]^T$ and $Q_i^{UAV} = [x_j, y_j, H]^T$, respectively.

Table 1. List of main notations.

Notation	Definition
$i/\mathcal{I}$	The index/set of UAVs
$j/\mathcal{J}$	The index/set of MUs
$g/\mathcal{G}$	The index/set of GBSs
d_{ij}	Distance between UAV i and MU j
Q_{u_i}/Q_{m_j}	The position of UAV i / MU j
Γ_{ij}^{avg}	Average path loss between UAV i and MU j
θ_{ij}^{RF}	Transmission rate between UAV i and MU j
θ_i^{FSO}	Backhaul rate between UAV i and its GBS
d_{ig}	Distance between UAV i and GBS g
R_{ij}	Load rate between UAV i and MU j
B_{ij}	Bandwidth allocated to MU j by UAV i
f_{ij}	CPU frequency allocated to MU j by UAV i
T_{ij}^{RF}/T_{ig}^{FSO}	The RF/FSO transmit time between UAV i and GBS g
σ	Additive white Gaussian noise
pr_i	The weight of priority
S	State space
A	Action space
$r(t)$	Reward
$J(\theta_i)$	Obeject function
$Q^{\mu}(s,a)$	Value function

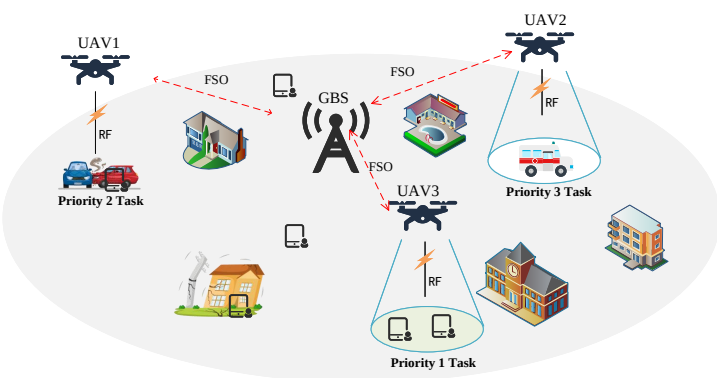

Figure 1. The model of UAVs cooperate to connect GBSs and MUs for emergency communication.

3.2. Air-to-Ground Channel

In the proposed channel model, due to the high complexity of urban environments and dynamic environmental changes, UAVs cannot obtain sufficient information. Therefore, directly applying the ideal free space model is impractical. We introduce a probabilistic model for light-of-sight (LoS) and non-light-of-sight (NLoS), which depends on environmental parameters, elevation angle, and the positions of MUs and UAVs.

The 3D distance between the MU and UAV can be expressed as

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + H^2}. \tag{1}$$

Similarly, according to [35], the probability of the channel being LoS can be considered as

$$p_{ij}^{LoS}(\varphi_{ij}) = \frac{1}{1 + C\exp\left(-D(\varphi_{ij} - C)\right)}, \tag{2}$$

where C and D are coefficients depending on environmental factors such as terrain, atmospheric conditions, lighting conditions, etc. The elevation angle φ_{ij} between UAV i and MU j is given by

$$\varphi_{ij} = \frac{180^\circ}{2\pi}\arcsin(H/d_{ij}). \tag{3}$$

According to [36], the LoS and NLoS path losses can be defined as

$$\Gamma_{ij}^{LoS} = \eta^{LoS} + \gamma^{LoS}\log d_{ij} + G, \tag{4}$$

$$\Gamma_{ij}^{NLoS} = \eta^{NLoS} + \gamma^{NLoS}\log d_{ij} + G, \tag{5}$$

where η^{LoS} and η^{NLoS} present the path losses when the reference distance d_{ij} is equal to 1 m. γ^{LoS} and γ^{NLoS} indicate the path loss exponents of the LoS and NLoS transmissions, respectively; G represents a Gaussian random variable with standard deviation. The average value of the path loss can be determined by

$$\Gamma_{ij}^{avg} = p_{ij}^{LoS}\Gamma_{ij}^{LoS} + (1 - p_{ij}^{LoS})\Gamma_{ij}^{NLoS}. \tag{6}$$

Between the UAV and the MU, the Signal-to-Interference-plus-Noise Ratio (SINR) of the RF channel can be expressed as

$$SINR = \frac{P_{\text{signal}}}{P_{\text{noise}} + P_{\text{intf}}}, \tag{7}$$

$$= \frac{p_i^{RF} 10^{\frac{-\Gamma_{ij}^{avg}}{10}}}{\sigma^2 + \left(\sum\limits_{a=j+1}^{\Phi_i} p_a^{RF} 10^{\frac{-\Gamma_{ia}^{avg}}{10}}\right)}, \tag{8}$$

where P_{signal} is the transmission power; P_{intf} and P_{noise} are the co-channel interference and noise, respectively. The latter is modeled as additive white Gaussian noise (AWGN) σ. To improve the communication quality, we employ Non-Orthogonal Multiple Access (NOMA) technology to reduce co-channel interference among multiple devices served by the same UAV. The set of tasks received by UAV Φ_i is defined and sorted in descending order of signal strength. When a UAV processes offloading tasks, it prioritizes tasks with stronger signals, reducing interference for subsequent tasks.

The achievable rate can be defined by the SINR as

$$\theta^{RF} = B_{ij} * \log_2 \left(1 + \frac{p_i^{RF} 10^{\frac{-\Gamma_{ij}^{avg}}{10}}}{\sigma^2 + \sum_{a=j+1}^{\Phi_i} p_a^{RF} 10^{\frac{-\Gamma_{ia}^{avg}}{10}}} \right), \tag{9}$$

where B_{ij} is the bandwidth allocated by UAV i to MU j.

3.3. Air-to-Air Channel

We consider that the computing ability of the GBS, which is the endpoint of the FSO channel for task offloading, is sufficiently abundant. Therefore, the computing time for edge offloading tasks at the GBS can be neglected. We only need to focus on optimizing the transmission rate of the FSO channel between the UAV and the GBS. Due to the typical high-altitude operation range of UAVs, atmospheric conditions affecting the FSO channel must be considered. According to [37], we introduce a size distribution of scattering particles based on weather conditions, defined as

$$\xi = \begin{cases} 1.6, & \nu > 50 \text{ km}, \\ 1.3, & 6 \text{ km} < \nu < 50 \text{ km}, \\ 0.16\nu + 0.34, & 1 \text{ km} < \nu < 6 \text{ km}, \\ \nu - 0.5, & 0.5 \text{ km} < \nu < 1 \text{ km}, \\ 0, & \nu < 0.5 \text{ km}, \end{cases} \tag{10}$$

where ν is the atmospheric visibility per kilometer. The atmospheric attenuation coefficient can be represented as [38]

$$\rho = \frac{3.91}{\nu} \left(\frac{\lambda}{550 \times 10^{-9}} \right)^{-\zeta}, \tag{11}$$

where λ is the transmission wavelength. Utilizing the atmospheric attenuation coefficient ρ and the distance between the UAV and the GBS d_{ig}, the atmospheric transmittance at the laser transmitter wavelength $\varkappa_{atm}$ can be expressed as [34]

$$\varkappa_{atm} = 10^{\frac{-\rho d_{ig}}{10}}. \tag{12}$$

The achievable rate of the FSO channel between the UAV and the GBS can be represented as [34]

$$\theta_i^{FSO} = \frac{p^{FSO} \varkappa_t \varkappa_{atm} \zeta^2}{\pi (v_t/2)^2 d_{ig} E_p N_b}, \tag{13}$$

where p^{FSO} is the transmission power of the UAV on the FSO channel; $\varkappa_t$ is the optical efficiency of the receiver and transmitter. ζ represents the free-space beam diameter at the UAV receiver. v_t denotes the divergence angle of the transmitter. $E_p = hc/\lambda_c$ represents the photon energy, where h denotes the constant of Planck, c is the speed of light, and λ_c is the wavelength. N_b represents the average receiver sensitivity. Summarizing these considerations, we establish the high-altitude FSO channel model.

3.4. Computing Delay Model

In the proposed system model, computational tasks of user devices can be processed in three ways: local computing, partial offloading to UAVs via RF channels, and further offloading from UAVs to the GBSs via FSO channels. These methods are detailed as follows:

(1) **Local Computing:** When the computing power of the device of an MU can satisfy the task requirements, the MU can choose local processing. The processing time of a task can be expressed as $T_{ij}^{total} = T_{ij}^{local}$, where T_{ij}^{local} can be expressed as

$$T_j^{local}[l] = \frac{(1 - R_{ij}[l])D_j[l]s}{f_{m_i}[l]}, \tag{14}$$

where l indicates the time slot, $R_{ij}[l]$ represents the pre-assigned offload ratio of UAVs for the task of MU j. D_j is the size of the task submitted by user j. f_{m_j} is the CPU frequency of the user's device. s represents the CPU cycles required per bit.

(2) **Offloading to UAVs:** When the computing power of the device of an MU cannot meet the task requirements, the MU can choose to offload part of the task to the UAVs for concurrent execution. T_{ij}^{total} can be represented as

$$T_{ij}^{total} = \max\left(T_{ij}^{local}, T_{ij}^{RF} + T_{ij}^{comp}\right), \tag{15}$$

where T_{ij}^{RF} and T_{ij}^{comp} represent the time of transmission and computing when the user transmits data to UAV i by the RF channel, respectively. T_{ij}^{RF} and T_{ij}^{comp} can be expressed as

$$T_{ij}^{RF}[l] = \frac{R_{ij}[l]D_j[l]}{\theta^{RF}}, \tag{16}$$

and

$$T_{ij}^{comp}[l] = \frac{R_{ij}[l]D_j[l]s}{f_{u_i}[l]}. \tag{17}$$

(3) **Offloading to BS via UAV:** When the computing power of both the user's device and the UAV cannot meet the task requirements, the UAV will forward the received task data to the GBS via the FSO channel. T_{ij}^{total} can be represented as

$$T_j^{total} = \max(T_j^{local}, T_{ij}^{RF} + T_{ig}^{FSO}), \tag{18}$$

where T_{ig}^{FSO} is the transmission time from UAV i to the GBS g via the FSO channel, T_{ig}^{FSO} can be expressed as

$$T_{ig}^{FSO}[l] = \frac{R_{ij}[l]D_j[l]}{\theta^{FSO}}. \tag{19}$$

3.5. Problem Formulation

In scenarios with limited system resources and environmental obstructions, our objective is to maximize the completion rate of tasks, with a particular emphasis on prioritizing high importance tasks. We introduce a binary parameter λ_{ij} to represent the matching relationship between user j and UAV i. If $\lambda_{ij} = 1$, this indicates that user J is served by UAV i, and vice versa. We can formalize the system optimization problem as

$$P1: \max_{x_i[l], y_i[l], z_i[l]} \sum_{j}^{J} \mathbb{I}(T_{ij}^{total} \le t_j^{delay}) \cdot pr_j,$$
$$\text{s.t.} \begin{cases} C1: \sum_i \lambda_{ij} \le 1, \quad \forall i \in \mathcal{I}, \lambda \in [0,1] \\ C2: SINR - \delta \le 0, \\ C3: \sum_i \lambda_{ij} B_{ij} < B, \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \\ C4: \sum_i \lambda_{ij} f_{ij} < F, \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \\ C5: 0 \le R_{ij} \le 1, \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \\ C6: x_i[l] \in [X_{min}, X_{max}], \quad \forall i \in \mathcal{I}, \\ C7: y_i[l] \in [Y_{min}, Y_{max}], \quad \forall i \in \mathcal{I}, \\ C8: z_i[l] \in [Z_{min}, Z_{max}], \quad \forall i \in \mathcal{I}. \end{cases} \tag{20}$$

In this context, P1 presents the goal of formulating the proposed system model problem pr_j, indicating the priority weights of tasks. $\mathbb{I}(T_{ij}^{total} \leq t_j^{delay})$ is an indicator function, and we have $\mathbb{I}(T_{ij}^{total} \leq t_j^{delay}) = 1$ if $T_{ij}^{total} \leq t_j^{delay}$, otherwise, $\mathbb{I}(T_{ij}^{total} \leq t_j^{delay}) = 0$. C1 represents that for user device j, no more than one UAV can serve it. C2 specifies that when users transmit data through the channel, if the SINR is below a certain threshold, it will be deemed a failure, where δ is the minimum SINR threshold. C3 and C4 indicate that the bandwidth and CPU frequency allocated by UAV i to user device j cannot exceed the total resource of the UAV. C5 states that the offloading ratio cannot be negative and cannot exceed 1. C6–8 specify that the coordinates of the UAV cannot exceed the scope of the application scenario.

4. KMADDPG-Based UAV Resource Allocation

4.1. Overview

In multi-agent scenarios, traditional reinforcement learning algorithms such as Deep Q-Network (DQN) face significant challenges. The primary issue is that during training, each agent treats other agents as part of the environment. As other agents change, the environment becomes unstable, violating the Markov assumption essential for Q-learning algorithms. Consequently, DQN struggles to converge in multi-agent environments.

Conversely, the MADDPG algorithm employs centralized training and decentralized execution strategies. This approach allows each agent to optimize its actions by considering the strategies of other agents, leading to more stable convergence and significantly enhancing the overall performance of the multi-agent system. Thus, MADDPG exhibits greater adaptability and efficiency in multi-agent contexts. Given the limited system resources, UAVs functioning as edge offloading nodes must carefully allocate resources to users.

4.2. Markov Decision Analysis

The proposed multi-UAV environment is typically considered a discrete partially observable Markov decision process (POMDP), as there are multiple interacting agents, each with its local observations and decision space, and the agents' partial observations of the environment may be incomplete. The future state of a UAV depends only on the current state and the actions of other UAVs, not on past states. They are regarded as distributed agents capable of making resource allocation and task offloading decisions. The decisions made by UAVs are decentralized, but their actions affect other agents, necessitating coordination with other UAVs to achieve the objectives. The observations, states, actions, and rewards of distributed UAVs at time t are defined as $o(t), s(t), a(t)$ and $r(t)$:

(1) State $s(t)$: In the proposed system model, the information obtained by agents through observations at time slot t is defined as $s(t)$. Due to each agent having its local observation and local decision space, $s(t)$ is typically incomplete. It can be composed of $\{s_i(t), s_{m_0}(t), \cdots, s_{m_J}(t), s_g(t)\}$, where $s_g(t)$ represents the state information of the GBS. $s_i(t)$ denotes the state information of UAV i at time slot t, which can be expressed as

$$s_i(t) = \{x_i, y_i, z_i, B_i, f_i\}, \tag{21}$$

where B_i and f_i represent the bandwidth and CPU computing power held by the UAV, respectively. $s_{m_j}(t)$ represents the state information of user equipment, and its components are as

$$s_{m_j}(t) = \{x_j, y_j, t_j^{delay}, f_j, D_j, pr_j\}, \tag{22}$$

where t_j^{delay} represents the delay requirement for the task delivered by user j; pr_j indicates the priority weight of the current task.

(2) Observation $o(t)$: At time slot t, all agents' local observation information $s(t)$ can be combined to form a global observation $o(t)$. It can be expressed as

$$o(t) = \{s_0(t), \cdots, s_I(t), s_{m_0}(t), \cdots, s_{m_J}(t), s_g(t)\}. \tag{23}$$

(3) Action Space $a(t)$: The action space of KMADDPG mainly consists of three parts: decisions of task matching, edge offloading, and resource allocation. The components of $a(t)$ are as

$$a_i(t) = \{a_i^{mat}(t), a_i^{load}(t), a_i^{res}(t)\}, \tag{24}$$

where $a_i^{mat}(t)$ denotes the decision of UAV matching with users, and $a_i^{mat}(t) = \{\lambda_{i0}, \cdots, \lambda_{iJ}\}$. $a_i^{offload}(t)$ denotes the decision of task offloading rate, and $a_i^{res}(t)$ indicates the actions of resource allocation.

(4) Reward Function $r(t)$: $r(t)$ is the sum of rewards completed by all UAVs. In the proposed scenario, the priority weights of user requests differ; it should focus on handling high-priority urgent requests. Therefore, the reward function is related to priority weights and completion status, which can be expressed as

$$R(t) = \sum_{i}^{I} \{\gamma_{ij} pr_j \omega + (1 - \gamma_{ij})\xi\}, \tag{25}$$

where γ_{ij} is a binary coefficient indicating the completion status of the task. If the UAV can satisfy the user's delay requirement, then $\gamma_{ij} = 1$ and the task is considered completed. The task reward can be expressed as $R(t) = \gamma_{ij} \times pr_j \times \omega$. Otherwise, $\gamma_{ij} = 0$ and the task reward can be expressed as $R(t) = \xi$, where ω is the reward coefficient for task completion, and ξ is the penalty coefficient for task failure.

4.3. MADDPG-Based Resource Allocation and Task Offloading

In the proposed multi-agent algorithm, we adopt a centralized training and decentralized execution (CTDE) strategy. This approach maximizes rewards while minimizing interactions between agents to enhance convergence speed. As show in Figure 2, we propose to use reinforcement learning combined with the K-means algorithm to optimize the initial positions of the agents, to achieve better optimization results.

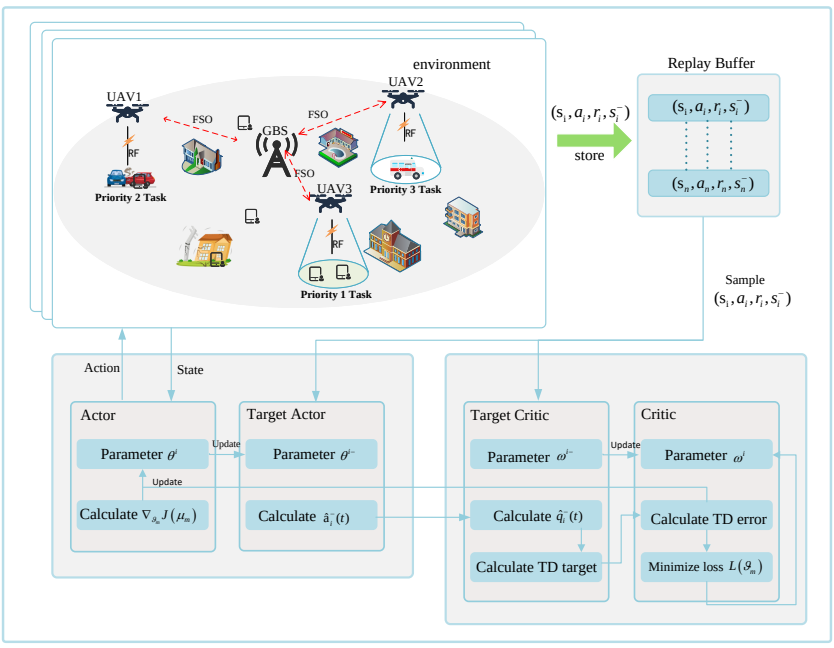

Figure 2. The KMADDPG algorithm in the UAV-assisted emergency communication network.

(1) **Joint K-Means Algorithm:** We propose a joint K-means algorithm for initializing the positions of UAVs. According to the path loss formula, path loss is positively correlated with the distance between UAVs and MUs. During centralized training, the system obtains the positions of all agents. We define sets $\hat{Q}^{UAV}$ and $\hat{Q}^{MU}$ to

represent the positions of UAVs and users, respectively. Here, $\hat{Q}$ is different from the Q mentioned earlier; it only represents the 2D coordinates of UAVs and users. The purpose of the K-means algorithm is to set I cluster centers $\{C_0, \cdots, C_I\}$, assign J user devices to the nearest cluster centers C_i, and finally, the positions of the cluster centers will be the target positions for the UAVs. The optimization objective function of the K-means algorithm can be expressed as

$$\arg\min_{C} \sum_{i=1}^{I} \sum_{j \in C_i} \|\hat{Q}_j^{mu} - \hat{Q}_i^{UAV}\|^2. \tag{26}$$

The updating process of K-means can be expressed as

$$\hat{Q}_i^{UAV}(t+1) = \frac{1}{|C_j(t)|} \sum_{i \in C_I(t)} \hat{Q}_j^{mu}. \tag{27}$$

(2) **Deep Reinforcement Learning Algorithm:** To avoid the additional load and delay caused by information exchange between the central controller and UAVs, this paper proposes a distributed joint resource management algorithm based on MADDPG. We consider a centralized training decentralized execution DRL mechanism. In the CTDE strategy, the collaborative operation of agents can be viewed as an optimization problem that maximizes expected rewards. This optimization problem can be expressed as

$$J = \arg\max_{\mu} \mathbb{E}\left[\gamma^{t-1} r(t)\right], \tag{28}$$

where $\mu = \{\mu_0, \cdots, \mu_I\}$ represents the policy network of the agent. γ^{t-1} denotes the discounted return at time slot t. We set the parameters of the policy network as $\theta = \{\theta_0, \cdots, \theta_I\}$, then the expected return gradient of the UAV can be written as

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}_{s,a \sim D}\left[\nabla_{\theta_i} \log \mu_i(a_i \mid o_i) Q_i^{\mu}(*) \mid a_i = \mu_i(o_i)\right], \tag{29}$$

where D is the experience replay buffer that stores experience tuples $\{S, a, r, S'\}$. $Q_i^{\mu}(^*)$ is the value function. The loss function of the Critic network for UAV i can be expressed as

$$L(\theta_i) = \mathbb{E}_{S,a,r,S}\left[\left(Q_i^{\mu}(S, a_1, \cdots, a_I) - \hat{y}\right)^2\right], \tag{30}$$

where $\hat{y}$ is the TD error, which can be expressed as

$$\hat{y} = r_i + \gamma Q_i^{\mu^-}(S, a_1, \cdots, a_I)\Big|_{a_i = \mu^-(o_i^-)}, \tag{31}$$

where $Q_i^{\mu^-}$ and $\mu_i^-(*)$ are the target Critic network and target Actor network, respectively. The specific neural network update process is shown in Algorithm 1.

(3) **Complexity Analysis:** The complexity analysis of the KMADDPG algorithm can be divided into two main components: the K-means algorithm and the neural network. For the K-means algorithm, each iteration requires the position coordinates of each UAV and MU, resulting in a time complexity of $O(I \times J)$. The number of iteration is denoted as N_k. After iterating, the total time complexity for the K-means algorithm is given by

$$O(2 \times N_k \times I \times J). \tag{32}$$

Since the dimensionality of the 2D coordinates and the number of iterations are generally constants, this complexity simplifies to $O(I \times J)$. For the neural network part, the efficiency of the proposed algorithm is further evaluated through its time

complexity. Both offline training and online execution involve mapping states to actions using deep neural networks. The computational complexity of the deep neural network, denoted as $O(NN)$, is expressed as follows [39]:

$$O(NN) = O(\sum_{n}^{N^{layer}} l_n l_{n+1}), \tag{33}$$

where N^{layer} is the number of layers, and l_n represents the number of neurons in the n-th layer. The dimensions of input and output layers of the deep neural network depends on the size of state space and action space. For N_{ep} episodes, S steps, and N_{exp} experiences, the algorithm's time complexity is

$$O(\sum_{n}^{N^{layer}} l_n l_{n+1} \times N^{exp} \times N_{ep}^{S}). \tag{34}$$

Algorithm 1 KMADDPG Algorithm

Initialize: Max_episode, Max_ep_len;

for episode from 1 to Max_episode **do**
  Initialize state: S; local observations: o_i for any agent
  $i \in \mathcal{I}$; cluster $\{C_0, \cdots, C_I\}$ centroids randomly;
  for each MU $\hat{Q}_j^{mu}$ **do**
    Calculate the distance between UAV and MU;
    Find the closest centroids $\hat{C}_i = \arg\min_i \hat{d}_{ij}$;
    Assign $\hat{Q}_j^{mu}$ to closet centroids C_i;
  endfor
  Update centroids position:
    $\hat{Q}_i^{UAV} = \frac{1}{|C_j(t)|} \sum_{i \in C_I(t)} \hat{Q}_j^{mu}$;
  for t = 1 to Max_ep_len **do**
    For each UAV, move to Q^{UAV} from K-Means;
    UAV select action a_i;
    Get the reward r_i by environment;
    Store (S, a, r, S') in the replay buffer $\mathcal{D}$;
    $S \leftarrow S'$
    for agent i = 1 to I **do**
      Sample randomly a batch of N samples
      $\left(S_j, a_j, r_j, S_j'\right)$ sampled from $\mathcal{D}$;
      Calculate TD error according to (31);
      Update critic network using minimizing loss $L(\vartheta_i)$;
      Update actor network through gradient $\nabla_{\vartheta_i} J(\mu_i)$;
    endfor
    Update the target network parameters for agent i;
      $\omega_{new}^{i-} \leftarrow \tau \omega_{new}^{i} + (1-\tau)\omega_{now}^{i-}$;
      $\vartheta_{new}^{i-} \leftarrow \tau \vartheta_{new}^{i} + (1-\tau)\vartheta_{now}^{i-}$;
  endfor
endfor

5. Experimental Design and Analysis

5.1. Simulation Parameters

In this section, we evaluate the proposed KMADDPG-based UAV resource allocation and edge offloading optimization scheme. Specifically, we consider a disaster area of 200 m × 200 m. We consider a GBS set up at the center of the disaster area, establishing an FSO communication link with the UAVs. Meanwhile, the UAVs serve as relay BS or edge

offloading nodes, providing radio communication services or computational resources to MUs. To simplify the system model, we set the penalty coefficient to −1 and the reward coefficient to 1. The reward is the cumulative reward of tasks completed by all UAVs in each time slot. We used 8 MUs, 3 UAVs, and 1 GBS as the experimental subjects. The detailed parameters are listed in Table 2. The algorithms to be evaluated are as follows:

- **MeanResource MADDPG (MeanRes)**: An algorithm where the neural network selects the matching objects, but the offloading ratio and resource allocation ratio are evenly distributed.
- **NearbyMatch MADDPG (NearbyMatch)**: An algorithm where the neural network selects the offloading and resource ratios, and the service objects are the nearest users.
- **Base MADDPG (Base)**: An algorithm where the neural network selects the matching objects, offloading, and resource ratios.
- **Proposed KMADDPG (Proposed)**: An algorithm that integrates K-means into the Base MADDPG algorithm.

Table 2. List of main simulations parameters.

Parameters	Value
Learning rate of actor	2×10^{-5}
Learning rate of Critic	5×10^{-4}
The number of Episode	15,000
Episode length	10
Channel bandwidth (B)	50 MHz
Cpu frequency of UAV (f_{ij})	1.5×10^{9}
Cpu frequency of MU (f_{local})	10^{9}
Size of task (D_{ij})	0.1 M–1.5 M
Requested delay of task (t_j^{delay})	12 ms–15 ms
LoS additional path loss (η^{LoS})	1 dB
NLoS additional path loss for (η^{NLoS})	20 dB
LoS path loss exponent (β^{LoS})	2.09
NLoS path loss exponent (β^{NLoS})	3.75
Noise power (σ)	−95 dBm
Operating altitude of UAV (h_t)	100 m
Transmit power of UAV i (p^{RF})	30 dBm
Transmission power of FSO (p^{FSO})	200 mW
Efficiencies of optic $(\varkappa_t)$	0.8
Receiver diameter of FSO channel (ς)	0.06 m
Transmitter divergence of FSO channel (v_t)	2.07×10^{-4} rad
Wavelength of FSO channel (λ)	1550 mm
Sensitivity of receiver (N_b)	100 photons/bit

5.2. Convergence Analysis

Figure 3 illustrates the trend of average system rewards obtained by each UAV per episode, and Figure 4 presents the tendency of total success rate obtained by each UAV per episode. This experiment evaluates the convergence of the proposed algorithm in comparison to other algorithms. It can be observed that the rewards for KMADDPG, Base, NearbyMatch, and MeanResource decrease in that order. The MeanRes algorithm exhibits significant fluctuations in performance, performs the worst, and fails to converge. This is due to the algorithm's inability to customize based on user demand differences, resulting in a significant waste of computational resources.

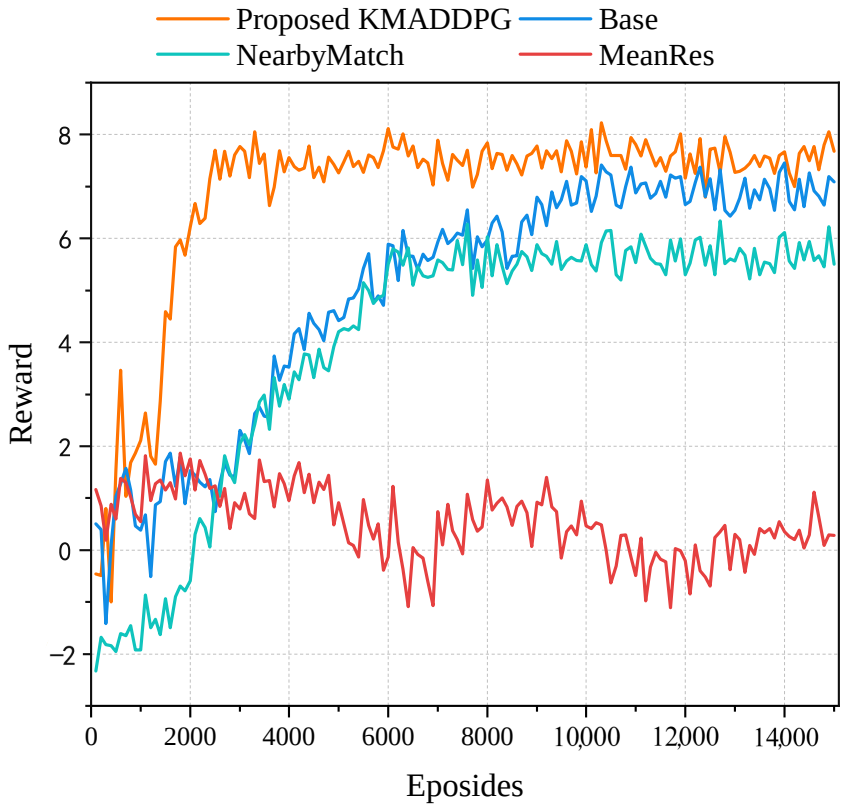

Figure 3. Convergence performance of reward under different algorithms.

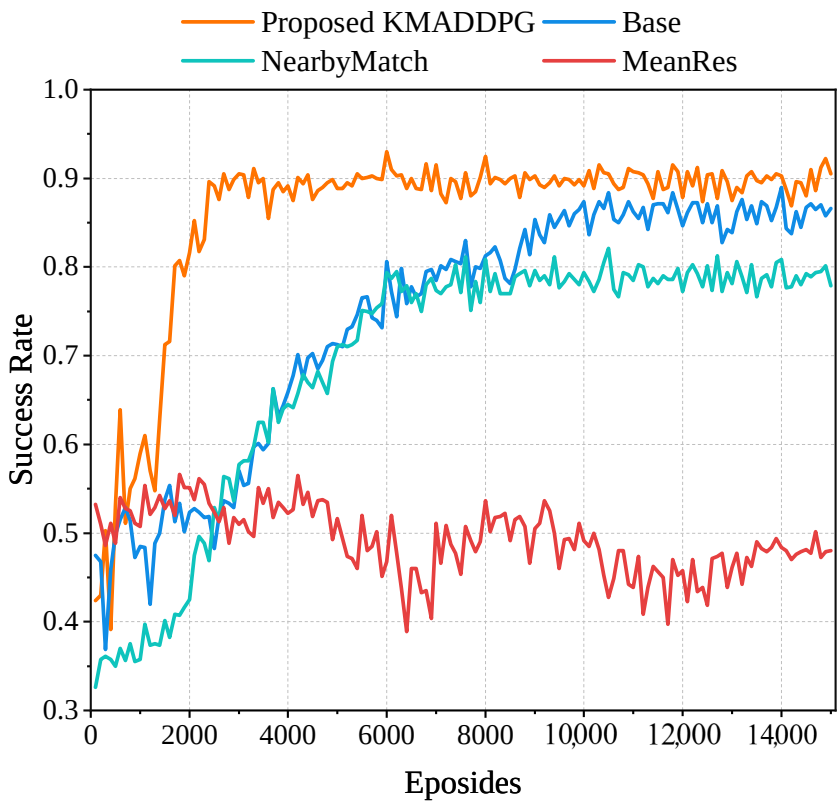

Figure 4. Convergence performance of total success rate under different algorithms.

The NearbyMatch algorithm performs somewhat better, with average rewards and average total success rates of 5.8 and 78% converges when the episode is equal to 8000,

as it devises strategies based on user demand. However, the limitation of NearbyMatch algorithm often causes it to fall into local optima. This limitation is that during training, MUs may cluster around a single UAV and offload task to it, leading to an inability of that UAV to satisfy all demands of MUs while the remaining UAVs remain idle.

The Base algorithm demonstrates better performance, with average rewards and average total success rates of 6.8 and 86% converges when the episode is equal to 10,000. The proposed KMADDPG algorithm addresses these shortcomings by integrating the KMeans algorithm to optimize UAV positioning, reducing path loss, and achieving better convergence performance compared to Base MADDPG. Consequently, it achieved the best results, with average rewards and average total success rates of 7.8 and 90% converges when the episode is equal to 3000. The proposed algorithm is capable of converging under minor constraints and has a faster convergence rate than the other algorithm.

5.3. Priority Analysis

Figures 5 and 6 depict the performance of various algorithms in terms of task completion within a single episode and their ability to handle tasks of different priority levels, respectively. Figure 5 shows the number of completed tasks within one episode. The KMADDPG algorithm completed 719 tasks, which is 8%, 12%, and 41% more than those completed by Base, NearbyMatch, and MeanResource, respectively. This indicates that the proposed algorithm exhibits superior task scheduling and processing capabilities. Figure 6 illustrates the performance of different algorithms in handling tasks with varying priorities, where P1, P2, and P3 represent low-, medium-, and high-priority tasks, respectively. It is evident from Figure 6 that all four algorithms perform better on higher-priority tasks, indicating that high-priority tasks receive more resource allocation. The proposed algorithm outperforms the other three algorithms across all three priority levels, with an increase in task completion by 12.5%, 5%, and 6% for low, medium, and high priorities, respectively, compared to Base. This demonstrates that the proposed algorithm not only maximizes the number of successful tasks but also exhibits superior handling capabilities for high-priority tasks.

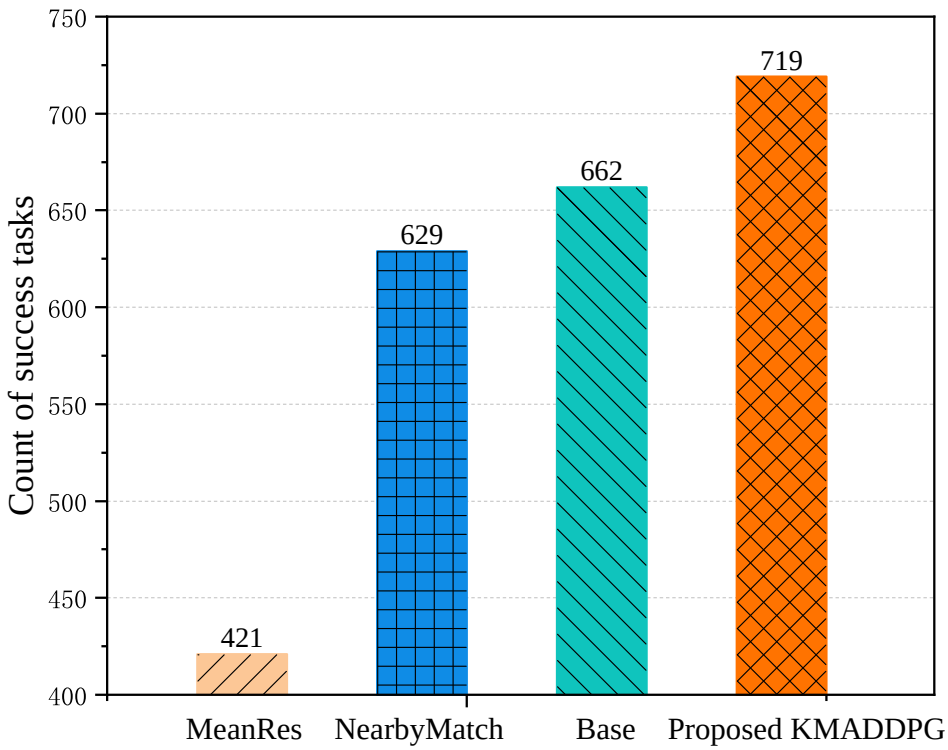

Figure 5. The distribution of success tasks under different algorithms.

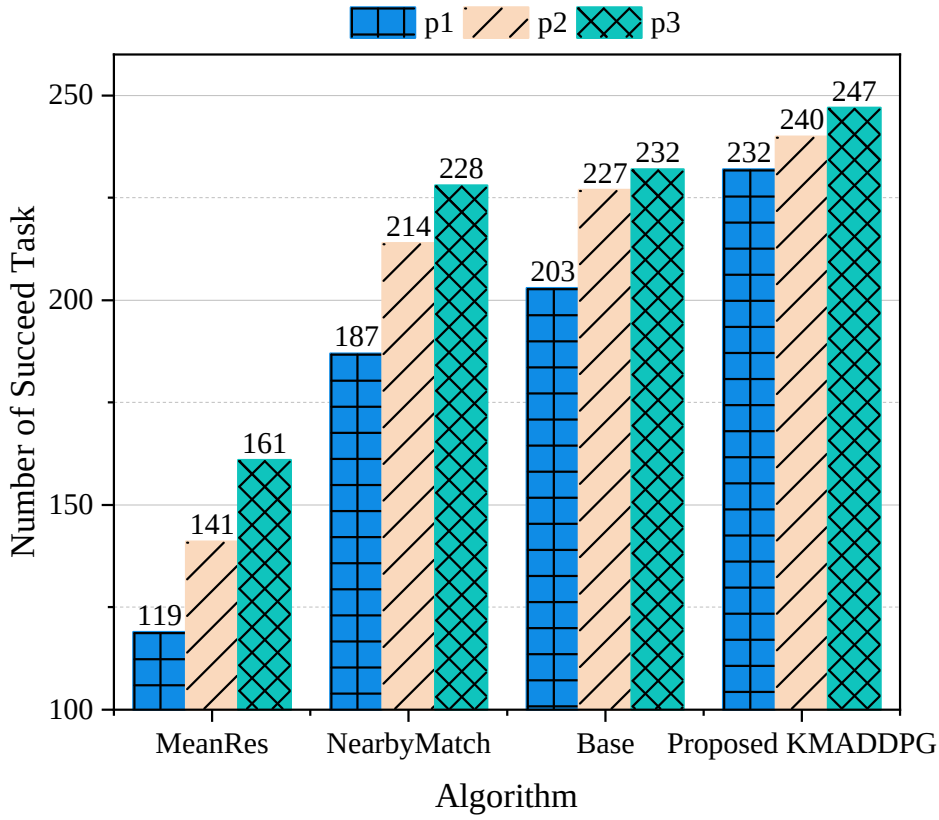

Figure 6. The distribution of successful tasks under different algorithms according to priority.

5.4. Convergence under Different Proportions

Figures 7 and 8 illustrate the trends of the average system reward and total success rate, respectively. In the proposed system model, as the number of UAVs and users increases, resources increase proportionally, making the environment more complex. Particularly, the increase in the number of agents means that each agent's decision can more frequently affect the states of other agents, making it harder for the results to converge. We compare different combinations of ratios to demonstrate that the proposed algorithm can still perform well even as the number of agents increases, though inevitably at a slower convergence rate. The reason for increasing resources proportionally is to avoid task failures due to insufficient resources. This experiment evaluated the convergence speed of various algorithms under different ratios of UAV and mobile user (MU) combinations. With the increase in the number of UAVs and users, resources increase proportionally, and the environment becomes more complex. To verify that the proposed algorithm can achieve good convergence results in complex situations, we used a UAV ratio and MU of 3:8 (simplified to (3, 8)) as the baseline, adjusting their numbers proportionally. From Figure 7, it is evident that the convergence rewards vary with different UAV combinations: the (2, 4) combination converges around 4, the (6, 16) combination converges around 15, and the (9, 24) combination converges around 22. This variation is due to the different number of users each combination serves. The convergence speed decreases sequentially because, as the number of agents and users increases, the system environment becomes more complex, necessitating more training for the neural network to achieve optimal results. However, as shown in Figure 8, even as the environment complexity increases, the final convergence effect still reaches 90%, demonstrating that the proposed algorithm performs well in more complex environments.

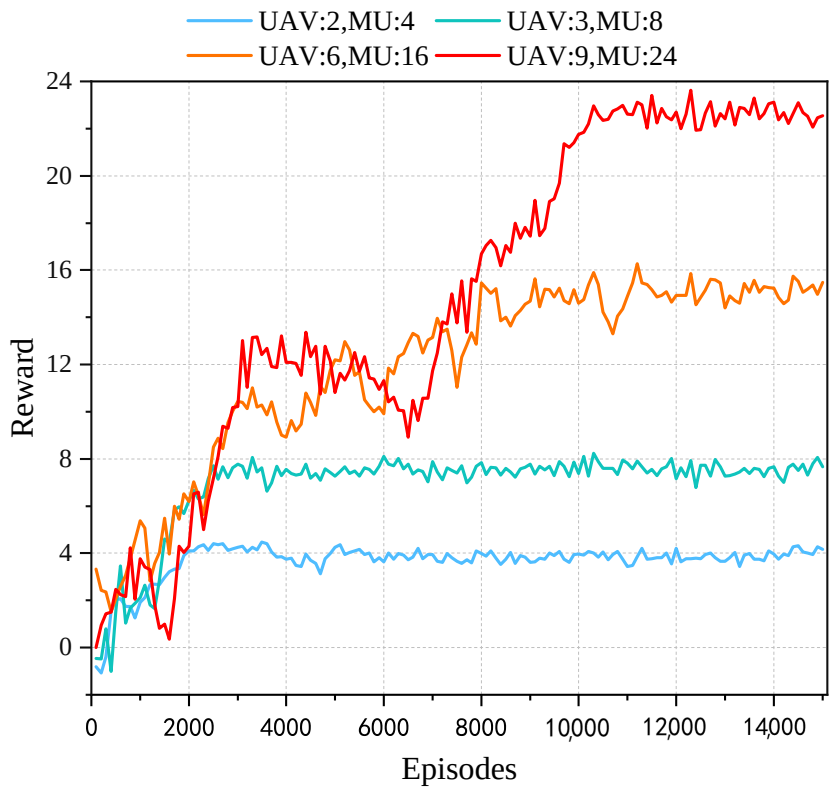

Figure 7. Convergence performance of reward under different couples.

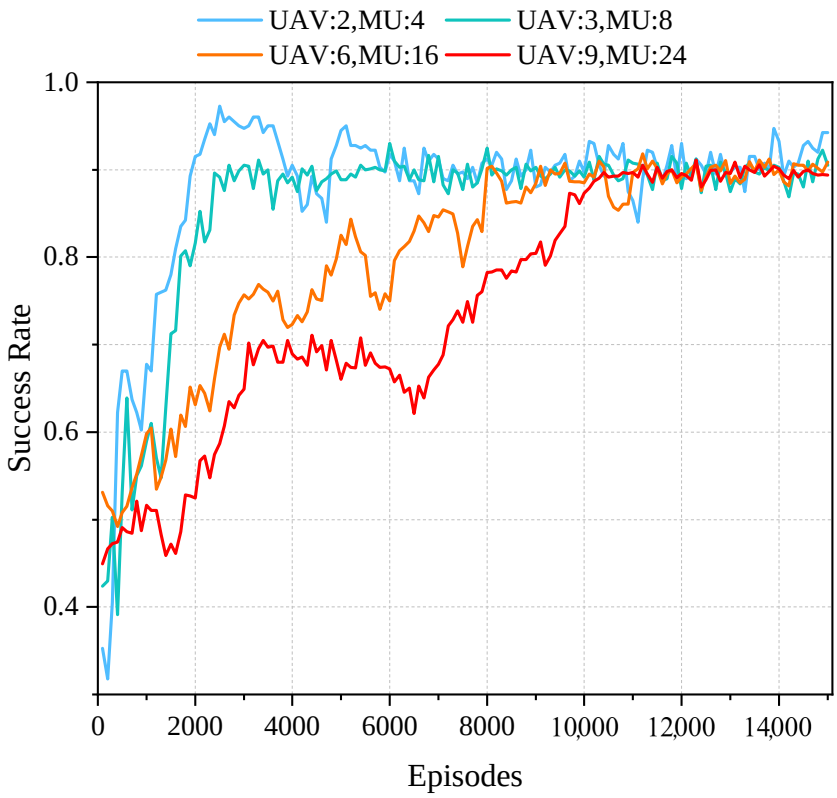

Figure 8. Convergence performance of total success rate under different couples.

6. Conclusions

To address the challenges faced by multi-UAV rescue systems utilizing a hybrid FSO/RF dual-hop model, we introduce a priority-driven resource allocation strategy. This

innovative approach is designed to optimize the completion rate of tasks, with a particular emphasis on ensuring that high-priority tasks are efficiently managed and expedited. This paper transforms the problem into a discrete partially observable Markov decision process and develops a KMADDPG algorithm to improve convergence speed. Simulation results indicate that the KMADDPG algorithm outperforms other DRL algorithms in enhancing system performance. In future work, we will explore resource allocation strategies for UAVs in terms of energy efficiency and transmission power, and incorporate dynamic priority mechanisms.

Author Contributions: Conceptualization, Y.L., Y.A. and Z.X.; Data curation, Y.A., Z.X. and J.W.; Formal analysis, Y.L., Y.A. and J.L.; Funding acquisition, Y.L., J.W. and J.L.; Investigation, Y.L., Y.A., Z.X. and J.W.; Methodology, Y.L., Z.X., J.W. and J.L.; Project administration, Y.L., Z.X., J.W. and J.L.; Resources, Y.A., Z.X. and J.W.; Software, Y.L., Y.A. and Z.X.; Supervision, Z.X. and J.L.; Validation, Y.A.; Visualization, J.W.; Writing—original draft, Y.L. and Y.A.; Writing—review and editing, Y.L., Y.A. and Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by the National Natural Science Foundation of China (62171463, 62271502), Natural Science Foundation of Fujian Province, China (2021J011112), and Science and Technology Bureau of Putian, Fujian Province, China (2022GZ2001ptxy11).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest

References

1. Wang, D.; Cao, Y.; Lam, K.Y.; Hu, Y.; Kaiwartya, O. Authentication and Key Agreement Based On Three Factors and PUF for UAVs-Assisted Post-Disaster Emergency Communication. *IEEE Internet Things J.* **2024**, *11*, 20457–20472. [CrossRef]
2. Yao, Z.; Cheng, W.; Zhang, W.; Zhang, H. Resource allocation for 5G-UAV-based emergency wireless communications. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 3395–3410. [CrossRef]
3. Wu, J.; Chen, Q.; Jiang, H.; Wang, H.; Xie, Y.; Xu, W.; Zhou, P.; Xu, Z.; Chen, L.; Li, B.; et al. Joint Power and Coverage Control of Massive UAVs in Post-Disaster Emergency Networks: An Aggregative Game-Theoretic Learning Approach. *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 3782–3799. [CrossRef]
4. Tang, X.; Chen, F.; Wang, F.; Jia, Z. Disaster Resilient Emergency Communication With Intelligent Air-Ground Cooperation. *IEEE Internet Things J.* **2023**, *11*, 5331–5346. [CrossRef]
5. Gao, J.; Wang, Q.; Li, Z.; Zhang, X.; Hu, Y.; Han, Q.; Pan, Y. Towards Efficient Urban Emergency Response Using UAVs Riding Crowdsourced Buses. *IEEE Internet Things J.* **2024**, *11*, 22439–22455. [CrossRef]
6. Zhou, M.; Chen, H.; Shu, L.; Liu, Y. UAV-assisted sleep scheduling algorithm for energy-efficient data collection in agricultural Internet of Things. *IEEE Internet Things J.* **2021**, *9*, 11043–11056. [CrossRef]
7. Bekkali, A.; Fujita, H.; Hattori, M. New generation free-space optical communication systems with advanced optical beam stabilizer. *J. Light. Technol.* **2022**, *40*, 1509–1518. [CrossRef]
8. Bekkali, A.; Hattori, M.; Hara, Y.; Suga, Y. Free Space Optical Communication Systems FOR 6G: A Modular Transceiver Design. *IEEE Wirel. Commun.* **2023**, *30*, 50–57. [CrossRef]
9. Guo, Z.; Gao, W.; Ye, H.; Wang, G. A location-aware resource optimization for maximizing throughput of emergency outdoor–indoor UAV communication with FSO/RF. *Sensors* **2023**, *23*, 2541. [CrossRef] [PubMed]
10. Yahia, O.B.; Erdogan, E.; Kurt, G.K.; Altunbas, I.; Yanikomeroglu, H. A weather-dependent hybrid RF/FSO satellite communication for improved power efficiency. *IEEE Wirel. Commun. Lett.* **2021**, *11*, 573–577. [CrossRef]
11. Aboelala, O.; Lee, I.E.; Chung, G.C. A survey of hybrid free space optics (FSO) communication networks to achieve 5G connectivity for backhauling. *Entropy* **2022**, *24*, 1573. [CrossRef] [PubMed]
12. Nafees, M.; Huang, S.; Thompson, J.; Safari, M. Backhaul-aware user association and throughput maximization in UAV-aided hybrid FSO/RF network. *Drones* **2023**, *7*, 74. [CrossRef]
13. Li, Y.; Zhang, W.; Wang, C.X.; Sun, J.; Liu, Y. Deep reinforcement learning for dynamic spectrum sensing and aggregation in multi-channel wireless networks. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 464–475. [CrossRef]
14. Zhu, G.; Lyu, Z.; Jiao, X.; Liu, P.; Chen, M.; Xu, J.; Cui, S.; Zhang, P. Pushing AI to wireless network edge: An overview on integrated sensing, communication, and computation towards 6G. *Sci. China Inf. Sci.* **2023**, *66*, 130301. [CrossRef]

15. Song, F.; Xing, H.; Wang, X.; Luo, S.; Dai, P.; Xiao, Z.; Zhao, B. Evolutionary multi-objective reinforcement learning based trajectory control and task offloading in UAV-assisted mobile edge computing. *IEEE Trans. Mob. Comput.* **2022**, *22*, 7387–7405. [CrossRef]
16. Yang, Y.; Song, T.; Yang, J.; Xu, H.; Xing, S. Joint Energy and AoI Optimization in UAV-Assisted MEC-WET Systems. *IEEE Sensors J.* **2024**, *24*, 15110–15124. [CrossRef]
17. Guo, S.; Zhao, X. Multi-Agent Deep Reinforcement Learning Based Transmission Latency Minimization for Delay-Sensitive Cognitive Satellite-UAV Networks. *IEEE Trans. Commun.* **2023**, *71*, 131–144. [CrossRef]
18. Xiong, Z.; Zhang, Y.; Lim, W.Y.B.; Kang, J.; Niyato, D.; Leung, C.; Miao, C. UAV-assisted wireless energy and data transfer with deep reinforcement learning. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *7*, 85–99. [CrossRef]
19. Qin, Z.; Liu, Z.; Han, G.; Lin, C.; Guo, L.; Xie, L. Distributed UAV-BSs Trajectory Optimization for User-Level Fair Communication Service With Multi-Agent Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2021**, *70*, 12290–12301. [CrossRef]
20. Westheider, J.; Rückin, J.; Popović, M. Multi-uav adaptive path planning using deep reinforcement learning. In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 649–656.
21. Orr, J.; Dutta, A. Multi-agent deep reinforcement learning for multi-robot applications: A survey. *Sensors* **2023**, *23*, 3625. [CrossRef] [PubMed]
22. Pang, W.; Wang, P.; Han, M.; Li, S.; Yang, P.; Li, G.; Guo, L. Optical intelligent reflecting surface for mixed dual-hop FSO and beamforming-based RF system in C-RAN. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 8489–8506. [CrossRef]
23. Wang, D.; Wu, M.; Wei, Z.; Yu, K.; Min, L.; Mumtaz, S. Uplink secrecy performance of RIS-based RF/FSO three-dimension heterogeneous networks. *IEEE Trans. Wirel. Commun.* **2023**, *23*, 1798–1809. [CrossRef]
24. Lee, J.H.; Park, K.H.; Ko, Y.C.; Alouini, M.S. Spectral-efficient network design for high-altitude platform station networks with mixed RF/FSO system. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 7072–7087. [CrossRef]
25. Che, Y.L.; Long, W.; Luo, S.; Wu, K.; Zhang, R. Energy-efficient UAV multicasting with simultaneous FSO backhaul and power transfer. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1537–1541. [CrossRef]
26. Qi, W.; Song, Q.; Guo, L.; Jamalipour, A. Energy-efficient resource allocation for UAV-assisted vehicular networks with spectrum sharing. *IEEE Trans. Veh. Technol.* **2022**, *71*, 7691–7702. [CrossRef]
27. Jiang, F.; Dong, L.; Wang, K.; Yang, K.; Pan, C. Distributed resource scheduling for large-scale MEC systems: A multiagent ensemble deep reinforcement learning with imitation acceleration. *IEEE Internet Things J.* **2021**, *9*, 6597–6610. [CrossRef]
28. Qin, Z.; Wang, H.; Wei, Z.; Qu, Y.; Xiong, F.; Dai, H.; Wu, T. Task selection and scheduling in UAV-enabled MEC for reconnaissance with time-varying priorities. *IEEE Internet Things J.* **2021**, *8*, 17290–17307. [CrossRef]
29. Liu, X.; Xu, C.; Yu, H.; Zeng, P. Deep reinforcement learning-based multichannel access for industrial wireless networks with dynamic multiuser priority. *IEEE Trans. Ind. Inform.* **2021**, *18*, 7048–7058. [CrossRef]
30. Seid, A.M.; Boateng, G.O.; Anokye, S.; Kwantwi, T.; Sun, G.; Liu, G. Collaborative computation offloading and resource allocation in multi-UAV-assisted IoT networks: A deep reinforcement learning approach. *IEEE Internet Things J.* **2021**, *8*, 12203–12218. [CrossRef]
31. Liu, Y.; Lin, P.; Zhang, M.; Zhang, Z.; Yu, F.R. Mobile-Aware Service Offloading for UAV-Assisted IoVs: A Multi-Agent Tiny Distributed Learning Approach. *IEEE Internet Things J.* **2024**, *11*, 21191–21201. [CrossRef]
32. He, Y.; Gan, Y.; Cui, H.; Guizani, M. Fairness-based 3D multi-UAV trajectory optimization in multi-UAV-assisted MEC system. *IEEE Internet Things J.* **2023**, *10*, 11383–11395. [CrossRef]
33. Lee, J.H.; Park, J.; Bennis, M.; Ko, Y.C. Integrating LEO satellites and multi-UAV reinforcement learning for hybrid FSO/RF non-terrestrial networks. *IEEE Trans. Veh. Technol.* **2022**, *72*, 3647–3662. [CrossRef]
34. Guan, Y.; Zou, S.; Peng, H.; Ni, W.; Sun, Y.; Gao, H. Cooperative UAV trajectory design for disaster area emergency communications: A multi-agent PPO method. *IEEE Internet Things J.* **2023**, *11*, 8848–8859. [CrossRef]
35. Ali, M.A.; Jamalipour, A. UAV-aided cellular operation by user offloading. *IEEE Internet Things J.* **2020**, *8*, 9855–9864. [CrossRef]
36. Liu, C.; Ding, M.; Ma, C.; Li, Q.; Lin, Z.; Liang, Y.C. Performance analysis for practical unmanned aerial vehicle networks with LoS/NLoS transmissions. In Proceedings of the 2018 IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, USA, 20–24 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
37. Nistazakis, H.E.; Tsiftsis, T.A.; Tombras, G.S. Performance analysis of free-space optical communication systems over atmospheric turbulence channels. *IET Commun.* **2009**, *3*, 1402–1409. [CrossRef]
38. Kim, I.I.; Koontz, J.; Hakakha, H.; Adhikari, P.; Stieger, R.; Moursund, C.; Barclay, M.; Stanford, A.; Ruigrok, R.; Schuster, J.J.; et al. Measurement of scintillation and link margin for the TerraLink laser communication system. In *Wireless Technologies and Systems: Millimeter-Wave and Optical*; SPIE: Bellingham, WA, USA, 1998; Volume 3232, pp. 100–118.
39. Wang, S.; Lv, T. Deep reinforcement learning based dynamic multichannel access in HetNets. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019 IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.