IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Digital Object Identifier

Towards Energy-Efficiency: Integrating DRL and ze-RIS for Task Offloading in UAV-MEC Environments

MUHAMMAD NAQQASH TARIQ [1], JINGYU WANG [1], (Senior Member, IEEE), SAIFULLAH MEMON [1], MOHAMMAD SIRAJ [2], (Senior Member, IEEE), MAJID ALTAMIMI [2], (Member, IEEE), and MUHAMMAD AYZED MIRZA[3].

[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications.
[2]Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh 11543, Saudi Arabia.
[3]School of Computer Science and Information Engineering, Qilu Institute of Technology, Jinan, Shandong 250200 P.R, China.

The authors are grateful to the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunication (BUPT), Beijing, China, for providing the Laboratory Services. The authors extend their appreciation to the Deputyship for Research and Innovation, 433 Ministry of Education, Saudi Arabia for funding this research (IFKSUOR3-459).

ABSTRACT Unmanned aerial vehicles (UAVs) play an important role within mobile edge computing (MEC) networks in improving communications for ground users during emergency situations. However, sustaining high-quality service for extended periods is challenging because of constraints on battery capacity and computing capabilities of UAVs. To address this issue, we leverage zero-energy reconfigurable intelligent surfaces (ze-RIS) within UAV-MEC networks and introduce a comprehensive strategy that combines task offloading and resource sharing. A deep reinforcement learning (DRL) driven energy efficient task offloading (DEETO) scheme is presented. The primary objective is to minimize UAV energy ingestion. DEETO aims to enhance task offloading decision mechanism, computing and communication resource allocation, while adopting hybrid task offloading mechanism with intelligent RIS phase-shift control. We begin by modeling it as a DRL problem, structuring it as a Markov decision process (MDP), and subsequently resolving it effectively through the use of the advantage actor-critic (A2C) algorithm. Our simulation results highlight the superiority of the DEETO scheme compared to alternative approaches. DEETO excelled by achieving a notable energy savings of 16.98% from the allocated energy resources, coupled with the highest task turnover rate of 94.12%, all achieved within a shorter learning time frames per second (TFPS) and yielding higher rewards.

INDEX TERMS DRL, MEC, Task offloading, UAV, ze-RIS.

I. INTRODUCTION

The mobile edge computing (MEC) networks have evolved as a very potential technological boost the quality of service (QoS) for tasks that demand significant computational power and low latency [1]. When MEC networks incorporate unmanned aerial vehicles (UAVs) into their infrastructure, they can harness the flexibility of UAV deployment [2]. This means that they can quickly expand the communication coverage and also guarantee the connectivity of the mobile users, particularly in cases of the destruction of the terrestrial communication infrastructures or a sudden influx of large numbers of users [3].

However, deploying additional UAVs in MEC networks comes with substantial costs and increased power consumption caused by the constraints in UAVs' energy and computational resources [4]. Moreover, UAV-assisted MEC systems encounter several hurdles. For instance, in urban settings, direct links between UAVs and IoT devices may encounter obstacles such as buildings, leading to significant deterioration in channel conditions. Similarly, since UAV is energy limited, it becomes critical to co-optimize UAV trajectory and computation energy to improve the overall energy efficiency. To take full advantage of UAV integration benefits in MEC and to further enhance the IoT task offloading performance, there is increasing interest in exploiting a new paradigm which is referred to as reconfigurable intelligent surfaces (RIS). RIS has emerged as an energy-efficient alternative capable of enhancing network capacity [5].

With the integration of a energy-efficient, cost-effective and high-gain metasurface, wireless propagation environ-

ment can be modified by RIS. This reconfiguration can minimize the energy consumption of the MEC networks and enhances communication capacity [6]. Recent research, such as studies in [7] through [8], has explored the applicability of RIS in MEC systems. Lately, numerous studies have delved into the theoretical exploration of leveraging RIS for network optimization. For instance, a notable work [9] focuses on integrating RIS into MEC systems to facilitate device task offloading. It seeks to enhance computational performance by maximizing the summation of computational bits. This is achieved through a joint optimization approach involving time allocation, transmit power and CPU frequency for computational offloading, while adjusting phase shifts of the RIS. The RIS beamforming and sensing mechanism for near field communications is investigated in [10]. Additionally, research by [11] delves into joint resource allocation and user grouping aided by IRS. Furthermore, [12] offers valuable insights into the resource trading mechanisms within MEC enabled UAV networks. Subsequently, in [13], the focus shifts on enhancing the total completed task-input bits within RIS-assisted MEC networks, a problem initially tackled by using an algorithm called block coordinate descending (BCD). A framework of deep learning is devised which can support the BCD algorithm and its online implementation in order to mitigate its computational complexity. Moreover, work in [14] explores the integration of wireless energy transfer (WET) into RIS-assisted MEC networks to meet both computation and energy supply demands of IoT devices.

In the context of RIS-assisted UAV networks, RIS deployment serves to mitigate blockages and improve achievable rates. In [15], the focus lies on maximizing the sum rate of users in the down-link through a joint optimization approach. This involves optimizing UAV trajectory, power control, THz sub-band allocation and RIS phase shifts. To ensure secure communication between the UAV and legitimate users, [16] employs RIS deployment to enhance legitimate transmission quality while compromising eavesdropping. To achieve the maximum average secrecy rate an iterative algorithm was proposed. Likewise, [17] explores secure transmission in UAV and RIS-assisted mmWave networks, achieving near-optimal RIS and UAV positions through exhaustive search methods. Moreover, [18] introduces a decaying deep Q-network (D-DQN) algorithm to address the dynamic stochastic environments in RIS-assisted UAV networks. This algorithm aims to minimize UAV energy consumption by optimizing RIS phase shift, power allocation, UAV trajectory and decoding order. Results of the simulation demonstrate that D-DQN algorithm effectively balances training speed acceleration, convergence to local optima, and oscillation avoidance. In [19], an active RIS-aided MEC system was the subject of attention. Acknowledging the fact that there are some tasks that cannot be partitioned, the author presented a multi-user MEC system with RIS assistance, which operates based on both binary and partial offloading policy [20]. Meanwhile, in [21], the author proposed an RIS-assisted iterative algorithm for MEC downlink communication system, aiming to minimize transmission power through alternating optimization (AO) and penalty-based optimization.

In studies [22] and [23], an innovative communication concept was introduced, featuring a UAV-based base station transmitting data to the users on ground with the simultaneous support of a reflecting and transmitting reflecting RIS (STAR-RIS). In [24], an approach was developed to enhance the energy efficacy of the system by jointly augmenting the UAV's flight trajectory in three-dimensional space and designing RIS phase shifts. Authors in [25] introduced an innovative MEC framework powered by RIS and UAV relays, specifically targeting a maximum-minimum computation capability issue.

The previously discussed networks utilizing RIS predominantly focus on two areas: RIS-assisted MEC networks, where user equipment (UE) computational capabilities benefit from access point (AP) resources, and RIS-assisted UAV architectures. This enhances the UE communication rates via UAV trajectory adaptability. However, scant attention has been given, to our knowledge, to the performance enhancement potential of RIS in UAV-enabled MEC environments. In studies [26], despite the use of a UAV-mounted RIS to aid ground user communication with an MEC server, the UAV's computational capacity remains unaddressed. Task-input data offloading by users to the MEC server via orthogonal multiple access (OMA) protocol is employed, which fails to fully exploit time and frequency resources. Subsequently, the authors in [27] introduces a non-orthogonal multiple access (NOMA) into the RIS-assisted UAV-MEC system, aiming to optimize UAV computational capacity. Simulation outcomes demonstrate the superior performance of the NOMA scheme compared to OMA. In association to it, a waveform design and detection mechanism in NOMA systems is also presented in [28]. Work presented in [29] also considered a IRS based NOMA assisted MEC mechanism for energy efficiency with queue stability option. A Lyapunov-function-based mixed integer DDPG (LMIDDPG) algorithm is proposed for centralized learning and a heterogeneous MA-LMIDDPG algorithm is proposed for distributed learning, considering system throughput, power consumption, and queue stability for computing decisions, power allocation, and phase shifts optimizations.

In an other work [30] a DDQN based resource allocation scheme, DTORA, is introduced. DTORA optimizes task offloading and resource allocation decision using double DQN algorithm. These decisions target to mitigate UAV energy consumption through optimizing computing and communication resources and RIS phase shifts. Additionally, work presented in [31] utilized another concept of RIS-equipped UAVs to facilitate MEC. A DDPG based energy-efficient solution mechanism is proposed for MEC network with RIS-equipped UAVs, considering energy conservation for ground users and UAVs. The optimization problem maximizes system energy efficiency by jointly optimizing ground user transmission power, UAV trajectory, and RIS phase shifts. The works presented in [32] explored a multi-UAV assisted

TABLE 1: Contrast of our work and related studies, emphasizing disparities and distinctive contributions.

	Offloading Mechanism		RAT		RIS Usage	ZED Association	Relaying		Objective (Minimization)		Solution Type	
	Binary	Partial	Cellular	NR			UAV	RIS	Energy	Delay	Other	DRL
This Work	✓	✓	–	✓	✓	✓	✓	✓	✓	✓	–	A2C
[20]	✓	✓	–	✓	✓	✓	✗	✓	✓	✓	–	A2C
[21]	✗	✓	✓	–	✓	✗	✗	✓	✓	✗	Optimization	–
[24]	✓	✗	✓	–	✓	✗	✗	✓	✓	✗	–	DQN
[29]	✗	✓	✓	–	✓	✗	✗	✓	✓	✓	–	LMIDDPG
[30]	✓	✗	✓	–	✓	✗	✓	✓	✓	✓	–	DDQN
[31]	✓	✗	✓	–	✓	✗	✓	✓	✓	✗	–	DDPG
[32]	✗	✓	✓	–	✓	✗	✗	✓	✓	✓	–	MA-TD3

MEC system with RIS. A multi agent DRL based MA-TD3 algorithm is proposed while aiming the optimization of computation offloading, UAV trajectories, and RIS phase shifts.

Motivated by the promising features of RIS and the performance limitations encountered by UAVs operating in MEC based networks, we intend to harness RIS to enhance UAV communications within MEC based networks. The RISs offer opportunities to reduce energy expenditures, boost UAVs' channel capacity, improve transmission reliability in MEC based networks, and expand wireless coverage. However, there is a notable scarcity of research concerning RIS-assisted UAVs in the field of mobile edge computing networks. While previously discussed studies have laid the groundwork for a novel system framework that incorporates RIS-assisted UAV-MEC, limited consideration has been devoted to investigating the energy ingestion aspect of RIS-assisted UAV-MEC systems.

Besides, while the offloading users intend to use UAV's resources, but given the limited resources and UAV's energy constraints, it's infeasible for users to execute their computation tasks concurrently on the UAV. Hence, there's a need to mutually augment task offloading and resource allocation to efficiently manage UAV resources and decrease energy consumption [13], [14], [21], [24]–[26]. However, the challenge lies in the complex, non-convex nature of this joint optimization problem, making it impractical to swiftly attain the optimal solution using conventional methods. Providentially, DRL has recently emerged as a powerful instrument for tackling numerous optimization problems within wireless communication networks. DRL offers the ability to derive optimal solutions for wide-reaching, intricate, and non-convex problems without necessitating complete and precise network information [18], [20], [24], [33].

The study outlined in [20] focused on minimizing energy and delay with ze-RIS considerations, albeit limited to vehicular networks and lacking the UAV relaying mechanism. Similarly, the research in [21] solely addressed partial offloading mechanism without considering ze-RIS or UAV relaying, while also being a non-DRL solution, potentially increasing complexity overhead. Additionally, [24] exclusively explored binary offloading mechanism without incorporating ze-RIS and only focused on energy optimizations. Moreover, [29] and [30] examined partial offloading mechanisms without considering UAV relaying or ze-RIS. Both [30] and [31] only delved into binary offloading mechanisms without any association with ze-RIS, with the latter not considering delay optimizations. The work in [32] is limited to partial offloading mechanisms without a UAV relaying mechanism.

Taking these factors into account, this paper introduces a hybrid resource allocation and task offloading scheme tailored for ze-RIS-assisted UAVs operating within MEC networks. ze-RIS offers energy efficiency, making it suitable for power-constrained environments [34]. The ze-RIS and conventional passive RIS differs in their power requirements. ze-RIS operates without external power sources, relying on energy harvesting or wireless power transfer. In contrast, conventional passive RIS requires a continuous power supply. While both enhance wireless communication by reflecting and manipulating radio waves [35], [36].

The scheme formalizes the objective of minimizing energy ingesting and leverages the DRL-based A2C algorithm to efficiently address the optimization problem. A specified comparison of related studies and our work is presented in the Table 1, and the key contributions of this work are as follows:

- The proposed DRL driven energy efficient task offloading (DEETO) scheme is based on B5G technological infrastructure. It incorporates an innovative RIS-assisted UAV-MEC system that not only equips the UAV with computation capabilities but also empowers it for relay transmissions. Within the proposed framework, we conduct a comprehensive joint optimization of several critical parameters, including the allocation of the UAV's computational resources, working environment communication resource allocation, task offloading decisions, and phase shifts for the RIS. The primary objective of this optimization is to minimize the energy consumption of the UAV.
- Given the non-convex nature of the objective function, our approach involves an initial problem formulation using a Markov decision process (MDP). In this context, it's important to highlight that both of the MDP spaces, the state space and the action space are continuous. To efficiently tackle this challenge, we employ the A2C algorithm.
- Our simulation outcomes demonstrate the notable energy efficiency improvements achieved by our RIS-

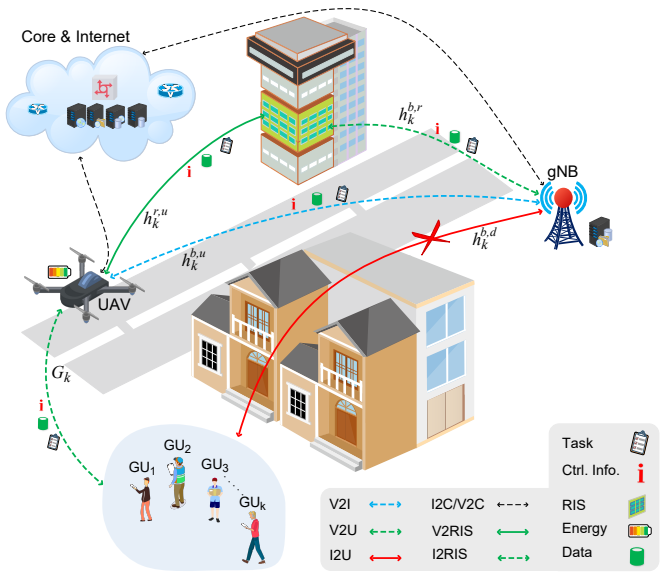


FIGURE 1: System model, demonstrating the conceptual framework.

aided UAV-MEC system, which employs the A2C algorithm. When compared to alternative benchmark schemes, our approach substantially reduces the overall energy consumption of UAVs.

II. SYSTEM AND OFFLOADING MODEL

A. SYSTEM MODEL

Recognizing the capabilities and advancements in 5G-NR communications, we have explored an RIS-enhanced UAV-MEC system, as depicted in Fig. 1. The UAV serves as a spatial base station, equipped with computational capabilities and the ability to function as a relay station. Thus, enabling the transmission of ground user's computational tasks to the next generation evolved node B (gNB) type base transceiver station (BTS). The BTS is supposed to be provided with multi CPU MEC server. We have integrated zero-energy reconfigurable intelligent surface (ze-RIS) elements into the architecture, thoughtfully positioned on the building's facade. We have followed the work proposed in [34], [35], and [36] for ze-RIS, and the zero energy device (ZED) based network configurations. The ze-RIS elements play a pivotal role in redirecting signals when needed. When users need to offload their computational tasks, they can choose between transmitting them directly to the UAV or using the UAV as a relay to send them to a distant small cell BTS for processing. The remarkable contribution of ze-RIS is its capacity to assist in signal relay from the UAV to the BTS, particularly in some situations where a direct communication is created amongst users and the BTS is hindered by physical barriers or obstructions.

Within the scheme, we have gu_m ground users, a UAV av_i, an RIS r_j equipped with k reflecting elements and a 5G small cell gNB type BTS b_l connected to MEC infrastructure. Here $i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}, l \in \mathcal{L}, m \in \mathcal{M}$ takes on values from the set $\{1, 2, \ldots, \mathcal{N}\}$, and $\mathcal{N}$ belongs to a natural number. To manage the RIS's reflection capabilities, we employ a reflection factor matrix denoted as $\Theta = \text{diag}[\theta_1(n), \theta_2(n), \ldots, \theta_k(n)]^T$, which embodies the reflecting beam-forming functionalities of the RIS. Here, $\theta_k(n)$ signifies the reflection factor of the k-th reflecting component during time slot n. The phase shift, $\phi_k(n)$, is contained within the interval $[0, 2\pi)$, and λ_k is the amplitude reflection coefficient, residing within the range $[0, 1]$. To optimize signal reflection, we typically set $\lambda_k = 1$, as previously done in [33]. We then divide total time period, T, into n distinct time periods. The variable n belongs to the set $\mathcal{N}$, and each of these time slots has a duration of δ. Users, during each of these time slots, offload their intended tasks to the in-range UAV following orthogonal frequency-division multiple access (OFDMA).

The channel gains for the ground user gu_m to UAV (gu_m–av_i link), from the UAV to BS (av_i–b_l link), UAV to RIS (av_i–r_j link), and from RIS to BS (r_j–b_l link) during time period n can be mathematically signified as $G_m(n)$, $h_m^{i,l}(n)$, $\mathbf{h}_m^{i,j}(n)$, and $\mathbf{h}_m^{j,l}(n)$, respectively. The gu_m–av_i link, av_i–b_l link, and av_i–r_j link are categorized as line-of-sight (LoS) links, and modeled using the free-space path loss model, and expressed as:

$$G_m(n) = \sqrt{\beta_0 {d_{m,i}}^{-2}(n)} \;:\; G_m(n) \in gu_m - av_i - \text{link}, \tag{1}$$

$$h_m^{i,l}(n) = \sqrt{\beta_0 {d_{i,l}}^{-2}(n)} \;:\; h_m^{i,l}(n) \in av_i - b_l - \text{link}, \tag{2}$$

$$\mathbf{h}_m^{i,j}(n) = \sqrt{\beta_0 d_{i,j}^{-2}(n)} \left[1, e^{-j\frac{2\pi}{\lambda} d\varphi_{i,j}(n)}, \ldots, e^{-j\frac{2\pi}{\lambda} d(K-1)\varphi_{i,j}(n)}\right] \;:\; \mathbf{h}_m^{i,j}(n) \in av_i - r_j - \text{link}, \tag{3}$$

In the given context, $d_{m,i}(n)$, $d_{i,l}(n)$, and $d_{i,j}(n)$ signify the distances for the gu_m–av_i, av_i–b_l, and av_i–r_j links in time slot n, respectively. The path-loss at reference distance of one meter is represented by β_0, d represents the inter RIS elements distance, and $\varphi_{i,j}(n) = (x_j - x_i)/(d_{i,j})$ denotes the cosine of the angle of arrival (AoA) of the signal. It's important to note that, for simplicity, the channel from RIS to BTS is considered to be mainly influenced by the Rician fading channel, expressed as:

$$\mathbf{h}_k^{j,l}(n) = \sqrt{\beta_0 d_{j,l}^{-2}(n)} \left[\sqrt{\frac{\beta_1}{1+\beta_1}} \mathbf{h}_{j,l}^{LoS}(n) + \sqrt{\frac{1}{\beta_1 + 1}} h_{j,l}^{NLoS}(n)\right] \;:\; \mathbf{h}_m^{j,l}(n) \in r_j - b_l - \text{link}, \tag{4}$$

$$\mathbf{h}_{j,l}^{LoS}(n) = \left[1, e^{-j\frac{2\pi}{\lambda} d\varphi_{j,l}(n)}, \ldots, e^{-j\frac{2\pi}{\lambda} d(K-1)\varphi_{j,l}(n)}\right] \tag{5}$$

where β_1 represents the Rician factor, $d_{j,l}(n)$ corresponds to the distance for the r_j–b_l link., and $\mathbf{h}_{j,l}^{LoS}(n)$ is the LOS

component. $\varphi_{j,l}(n) = (x_l - x_j)/(d_{j,l})$ denotes the AoD signal's cosine. For the NLoS component, $\mathbf{h}_{j,l}^{NLoS}(n)$ follows a Gaussian distribution with zero mean and unit variance, i.e., $\mathbf{h}_{j,l}^{NLoS}(n) \sim \mathcal{CN}(0,1)$.

B. OFFLOADING AND PROCESSING MODEL

In our study, we categorize the processing of user computation tasks into three specific phases. Initially, there is the task offloading from the ground users to the UAV, followed by the computation processing at the UAV, and finally, the relaying of tasks to the BTS. It is important to note that the BTS, being grounded and potentially equipped with multi-CPU MEC servers. These servers are expected to have a considerably higher computation rate compared to the UAV. Therefore, we can safely disregard the computing latency of BTS within all time slot.

1) UAV offloading

Given the constraints of finite computational power then the battery size for the ground user, the decision is made to transfer all of the computation tasks to UAV. In this context, the computation tasks of the m-th user in the n-th time slot are denoted as $t_k(n)$. Each $tk(n)$ is characterized as a 3-tuple: $tk_m(n) = \{s_m, c_m, t_m, e_m\}$, where s_m represents the task size in bits, c_m indicates the required CPU cycles for 1 bit computation, t_m, and e_m are the time and energy threshold units allocated for task completion. Since all the tasks are pre-decided to be offloaded, the transmission rate $R_m^{ofld}(n)$ of the m-th user in the n-th time slot while offloading task $tk_m(n)$ is given by:

$$R_m^{ofld}(n) = B_0 \log_2\left(1 + \frac{p_m(n)\ ||G_m(n)||^2}{\sigma^2}\right), \quad \text{and} \quad B_0 = B/m, \tag{6}$$

Here, B is the total bandwidth, distributed into m even channels as. $p_m(n)$ is the transmission power of m-th user in the n-th time slot. σ^2 denotes the power of additive Gaussian white noise. The offloading delay for the m-th user, when offloading their computation extensive tasks to the in range UAV, is calculated while following Eq. (7). Additionally, offloading a given task must follow the constraint $t_m^{ofld} < t_m$.

$$t_m^{ofld}(n) = s_m(n)/R_m^{ofld}(n). \tag{7}$$

2) UAV Computation

Since all of the user's computational tasks are transferred to UAV. The UAV dynamically determines the decisions of offloading depending on the incoming task count from the ground users. By employing a partial task offloading strategy, the UAV handles the ζ portion of the task, while the β segment is relayed to the BTS. The decision-making process involves the UAV determining the allocation of the m-th user's computation tasks, deciding how much of the task should be processed on the UAV and how much should be relayed to the BTS. These decisions adhere to the constraint $\zeta + \beta = 1$. The decision parameters ζ_m and β_m are binary in nature. In the case of adopting a partial task processing model, both of these parameters are set to 1. When exclusively UAV processing is preferred, ζ_m is set to 1, and β_m is set to 0. Conversely, when only relaying is favored, β_m is set to 1, and ζ_m is set to 0. The time t_m^{av} and energy e_m^{av} consumed in processing task $t_k(n)$ during the n-th time slot at UAV av_i can be expressed as follows:

$$t_m^{av}(n) = \Big((1-\zeta_m(n)(1-\beta_m(n))\Big)\left\{\frac{\zeta s_m(n)\ c_m(n)}{f_m^{av}(n)}\right\} \tag{8}$$

$$e_m^{av}(n) = (1 - d_m(n))\Big\{\zeta s_m(n)\ c_m(n)\ \varphi^{av}\ \left(f_m^{av}(n)\right)^2\Big\} \tag{9}$$

where $f_m^{av}(n)$ in Eq. (8) signifies the CPU resources allocated to the task t_k in n-th time slot, and φ^{av} in Eq. (9) is the UAV chip specific effective capacitance coefficient, determined by architecture of the chip.

3) UAV–BTS Relaying

For the decisions $(\zeta_m = \beta_m = 1)$ or $(\zeta_m = 0 \wedge \beta_m - 1)$ the task is to be relayed to the BTS for processing, either in parts or as a whole. Therefore, the transmission rate and time for the UAV–BTS link can be calculated while following Eq. (10), and Eq. (11), while the consumed energy for relaying can be calculated while following Eq. (12).

$$R_m^{rel} = B_0 \log_2\left(1 + \frac{p_m(n)\ ||h(n)||^2}{\sigma^2}\right), \quad \text{where}\ \ h(n) = h_m^{i,l}(n) + \mathbf{h}_m^{i,j}(n)\ \Theta \mathbf{h}_k^{j,l}(n) \tag{10}$$

$$t_m^{rel}(n) = \Big((1 - \zeta_m(n)(1 - \beta_m(n))\Big)\left\{\frac{\beta\ s_m(n)}{R_m^{rel}}\right\} \tag{11}$$

$$e_m^{rel}(n) = \Big((1-\zeta_m(n)(1-\beta_m(n))\Big)\left\{\frac{P_m^{rel}\beta\ s_m(n)}{R_m^{rel}}\right\} \tag{12}$$

III. OBJECTIVE AND SOLUTION METHODOLOGY

In this segment, we present our approach to addressing the challenge of reducing energy ingesting while enhancing the computational efficacy of the UAV based MEC network. The key objective is to fine-tune the use of computation resources, while aiming to maximize the proportion of total processed bits to the energy consumed by UAVs. In line with our goal, we can articulate the average energy ingesting and time consumed while offloading, computing, and relaying the task m is expressed in Eq. (13) and Eq. (14).

$$\mathcal{T}_m = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \left(t_m^{ofld}(n) + t_m^{av}(n) + t_m^{rel}(n)\right) \tag{13}$$

$$\mathcal{E}_m = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \left(e_m^{av}(n) + e_m^{rel}(n)\right) \tag{14}$$

A. DEXTERITY FACTOR

To ensure fair and impartial assessments of efficiency, we have introduced a parameter known as the dexterity factor. This factor is designed to incorporate standard benchmarks for time and energy consumption into the evaluation process. Specifically, we establish t_m^m as the average of the maximum and minimum time thresholds assigned to tasks, representing the benchmark for delay, and e_m^m as the average of the maximum and minimum energy thresholds allocated to tasks, serving as the benchmark for energy. These benchmarks, derived from the range of allowable time and energy values, provide a reference point against which the actual performance of the UAV can be compared.

In a dynamic RIS-assisted MEC network, the performance of a UAV in terms of time and energy can vary significantly across different time slots. To ensure unbiased evaluations of efficiency, we have introduced a dexterity factor. This factor is designed to incorporate standard benchmarks for time and energy consumption into the evaluation process. We define $t_m^m = \text{mean}(t_m^{max}, t_m^{min})$ as the benchmark for delay and $e_m^m = \text{mean}(e_m^{max}, e_m^{min})$ as the benchmark for energy. Here, t_m^{max} and t_m^{min} signify the maximum and minimum threshold times allocated to the tasks, while e_m^{max} and e_m^{min} signify the maximum and minimum threshold energy units allocated to the tasks, respectively. These benchmarks provide reference points against which the actual performance of the UAV can be compared, allowing for a more accurate evaluation of its efficiency. The dexterity factor is then formulated as:

$$\mathcal{D}_m = \{((t_m + t_m^m) - \mathcal{T}_m) + ((e_m + e_m^m) - \mathcal{E}_m)\} \quad (15)$$

It is worth noting that a larger value of $\mathcal{D}_m$ corresponds to a greater amount of saved energy and reduced delay. Since the dexterity factor $\mathcal{D}_m$ is associated with single task tk_m , the average dexterity factor for all M tasks can be expressed as:

$$\mathcal{D}_a = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \{((t_m + t_m^m) - \mathcal{T}_m) + ((e_m + e_m^m) - \mathcal{E}_m)\} \quad (16)$$

Building upon these foundational concepts, we establish our joint objective to optimize $\mathcal{D}_m$. This encompasses decisions associated with offloading, UAV processing and execution, and the relaying mechanism.

$$\underset{\zeta,\, \beta,\, \Theta,\, P^{rel},\, f^{av}}{\text{maximize}} \quad \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left(\frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \mathcal{D}_m \right)(n),$$

s.t.

$$C1 : \mathcal{E}_m(n) \leq e_m(n), \quad (17)$$

$$C2 : \mathcal{T}_m(n) \leq t_m(n), \quad (18)$$

$$C3 : 0 \leq f_m^{av}, \text{ and}$$
$$\sum_{m=1}^{\mathcal{M}} (1 - \zeta_m(n)(1 - \beta_m(n)) f_m^{av}(n) \leq F_{max}(n), \quad (19)$$

$$C4 : 0 \leq P_m^{rel}, \text{ and}$$
$$\sum_{m=1}^{\mathcal{M}} (1 - \zeta_m(n)(1 - \beta_m(n))\, P_m^{rel}(n) \leq P_{max}^{rel}(n), \quad (20)$$

$$C5 : 0 \leq \phi_k < 2\pi, \quad (21)$$

$$C6 : \forall\, \zeta_m \in \{0, 1\}, \text{ and } \forall\, \beta_m \in \{0, 1\}. \quad (22)$$

Constraint C1 is directly associated with energy and stipulates that the energy consumed during task processing and relaying must be less than or equal to the allocated energy. Constraint C2 pertains to user experience, ensuring that the time taken for task offloading, processing, and relaying must be less than or equal to the allocated threshold time. Constraints C3 and C4 signify the limitations on UAV computation and energy resources, respectively. Constraint C5 denotes the phase shift constraints for RIS elements. Constraint C6 signifies the parameters governing task offloading decisions.

The objective of the optimization problem described in Eq. (17), i.e., to maximize the D_m, corresponds to minimize the overall energy consumption of the UAV during its operational tasks. This involves a comprehensive approach to optimizing various aspects, including the task offloading decisions ζ_m and β_m, the phase shift Θ of RIS elements, the power consumption P^{rel} during relay transmissions, and the allocation of UAV computational resources f^{av}.

Given the dynamic nature of problem Eq. (17) and its inherent involvement in a multidimensional spaciotemporal environment, employing traditional optimization approaches proves challenging due to significant difficulty and high-order complexity. To address these challenges, we have chosen to employ a DRL based methodology. This decision is driven by the adaptive and learning-centric nature of DRL, which is particularly well-suited for handling dynamic and complex optimization problems in dynamic environments.

B. MDP FORMULATION

In the scope of this paper, the UAV plays the role of an agent that actively interact with the environment, with the primary objective of collecting comprehensive state information and formulating an optimum policy to exploit the cumulative reward. The designations for the MDP components, including the state, action spaces, and reward function, are explicitly outlined as follows:

1) State Space

The state space s_t in our MDP at the t time slot is characterized by various UAV states, including UAV position state s_u^p, UAV velocity state s_u^v, UAV energy state s_u^e, and UAV processing capacity state s_u^c. Additionally, the state space also incorporates RIS configuration state s_r, BS status state s_b, MEC workload state s_m^w, MEC capacity state s_m^c, and task-related states such as task size state s_{tk}^s, task CPU cycles requirement state s_{tk}^c, task threshold time state s_{tk}^t, and task energy threshold state s_{tk}^e. The comprehensive definition of the state space is as follows:

$$S \triangleq \{s|s = (s_u^p,\ s_u^v,\ s_u^e,\ s_u^c,\ s_r,\ s_b,\ s_m^w,\ s_m^c,\ s_{tk}^s, s_{tk}^c,\ s_{tk}^t,\ s_{tk}^e)\} \tag{23}$$

2) Action Space

The action state a_t in our MDP at t time slot encompasses UAV flight related actions including UAV speed a_u^s, UAV direction a_u^d, and UAV altitude a_u^a. Additionally, it also includes actions related to RIS phase shifts adjustments a_r, local task processing actions a_{tk}^l, task offloading decision action a_{tk}^o, and task relaying actions a_{tk}^r. In a comprehensive view, the action space can be articulated as follows:

$$A \triangleq \{a|a = (a_u^s,\ a_u^d,\ a_u^a,\ a_r,\ a_{tk}^l,\ a_{tk}^o,\ a_{tk}^r)\} \tag{24}$$

3) Reward Function

The reward function $R(s_t, a_t, s_{t+1})$ in our MDP encapsulates the goal of minimizing UAV energy consumption, ensuring task completion, and promoting efficient utilization of the ze-RIS. We have formulated the reward function as specified in Eq. (25), incorporating various factors to effectively train the DRL agent for optimal alignment with our objectives.

$$R(s_t, a_t, s_{t+1}) \triangleq \{r|r = (r_{df} + r_{ee} + r_{tc} + r_{ru})\} \tag{25}$$

The rewards r_{df}, r_{ee}, r_{tc}, and r_{ru} in Eq. (25) are directly related to the objective function. r_{df}, dexterity factor reward, this factor represents the agility and flexibility of the system in handling tasks. In the objective function, this could be associated with constraints related to task completion C2 and energy consumption C1. Maximizing r_{df} encourages the system to efficiently complete tasks while managing energy resources effectively. r_{ee}, the energy efficiency factor reward, this factor focuses on minimizing energy consumption, which directly aligns with constraint C1 in the objective function. By maximizing r_{ee}, the system aims to optimize energy usage while achieving its objectives. r_{tc}, the task completion reward, this factor is binary, indicating whether tasks are successfully completed $r_{tc} = 1$ or not $r_{tc} = 0$. In the objective function, this corresponds to the task completion constraint C2. Maximizing r_{tc} encourages the system to prioritize completing tasks efficiently. r_{ru}, the RIS utilization reward, this factor promotes the optimal utilization of RIS elements to enhance communication and energy consumption, aligning with constraints C4 and C5 in the objective function. Maximizing r_{ru} encourages the system to leverage RIS effectively to achieve its goals.

In order to train the DRL agent and guiding that towards higher rewards we have added weights with these factors as these can be observed from Eq. (26)

$$R(s_t, a_t, s_{t+1}) \triangleq \{r|r = (r_{df} + (-w_1 \cdot r_{ee}) + (w_2 \cdot r_{tc}) + (w_3 \cdot r_{ru}))\} \tag{26}$$

Here, w_1 weight is used to prioritize the importance of UAV energy consumption. The weight w_2 is associated with task completion reward, and w_3 is a measure indicating how effectively the RIS is utilized.

C. THE A2C BASED DEETO ALGORITHM

The DEETO scheme includes the collaboration among the DRL agent and the RIS-assisted MEC based UAV task offloading environment. The algorithm 1 integrates the Advantage Actor-Critic (A2C) algorithm with the UAV task offloading and relaying scenario for obtaining optimal task offloading decisions policy. The DEETO algorithm starts by initializing the policy network π_θ and the value function V_ϕ, along with the learning rate α and discount factor γ hyperparameters. Then algorithm iterates through episodes, representing sequences of interactions between the UAV and the environment. Each episode captures a series of time steps. At each time step within an episode, the algorithm the policy network π_θ samples an action a_t based on the current state s_t. Later, the sampled action a_t is executed in the environment, resulting in a reward r_t and a new state s_{t+1}. Then, based on the current state and connectivity information, the algorithm determines the communication scenario. This scenario can be a direct link $gu_m - av_i$ between the mobile user and the UAV or relay link $gu_m - av_i - b_l$ from mobile user to UAV to BS, or an RIS assisted link $gu_m - av_i - r_j - b_l$ an RIS-assisted link. Depending on the communication scenario, the state of the mobile user, BS, RIS, and UAV is updated to reflect changes in parameters such as task status, bandwidth, energy levels, delays, phase shifts, channel conditions, and resource availability. After updating, the transition tuple (s_t, a_t, r_t, s_{t+1}) is stored in a trajectory buffer for later use in policy and value function updates. Then the advantage A_t is calculated using the reward r_t and the value functions V_ϕ for the current and next states s_t and s_{t+1}, respectively. Next the policy parameters θ and value function parameters ϕ are updated using the A2C algorithm. The policy parameters are updated in the direction that increases the likelihood of selecting actions that yield higher advantages, while the value function parameters are updated to reduce the difference between the estimated value of the current state and the estimated value of the next state. After the specified number of episodes, the algorithm outputs the optimal policy π_θ and value function V_ϕ, which can then be used to make task offloading and relaying decisions in the UAV environment.

Algorithm 1 A2C assisted DEETO Algorithm

Input: Initial state s_0, policy network π_θ, value function V_ϕ, gu_m, av_i, r_j, b_l parameters
Initialize policy network π_θ and value function V_ϕ, learning rate α, discount factor γ
for each episode **do**
  Initialize environment initial state s_0
  for each time step t **do**
    Take sample offloading action $a_t \sim \pi_\theta(\cdot|s_t)$
    Execute offloading action a_t, observe reward r_t and new environment state s_{t+1}
    Determine communicating node locations and connectivity
    Select communication link ($gu_m - av_i$, $gu_m - av_i - b_l$, $gu_m - av_i - r_j - b_l$)
    if $gu_m - av_i$ **then**
      Update location, trajectory, computation and energy states of av_i
    else if $gu_m - av_i - b_l$ **then**
      Update location, trajectory, computation and energy states of av_i,
      Update b_l states
    else if $gu_m - av_i - r_j - b_l$ **then**
      Update location, trajectory, computation and energy states of av_i,
      Update phase shifts, channel conditions, energy states of r_j,
      Update b_l states
    end if
    Update gu_m, av_i, r_j, and b_l state parameters
    Update communication scenario used,
    Update energy consumed,
    Update delay incurred,
    Update available resources
    Store transition (s_t, a_t, r_t, s_{t+1}) in trajectory buffer
    Calculate advantage $A_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)$
    Update policy parameters θ and value function parameters ϕ using A2C:

$$\theta_{\text{new}} = \theta + \alpha \nabla_\theta \log \pi_\theta(a_t|s_t) \cdot A_t$$

$$\phi_{\text{new}} = \phi + \alpha \nabla_\phi \left(0.5(V_\phi(s_{t+1}) - V_\phi(s_t))^2\right)$$

  end for
end for
Output: Optimal offloading policy π_θ and value function V_ϕ

Throughout the algorithm, mathematical notations such as π_θ, V_ϕ, α, γ, ∇_θ, and ∇_ϕ are used to represent the policy network, value function, hyperparameters, and gradients for policy and value function updates. The algorithm provides a rigorous framework for delay and energy optimizing task offloading and relaying decisions in the UAV environment.

TABLE 2: Simulation Parameters

Simulation Parameter	*Value*
CPU frequency of av_i, and MEC	2MHz–2GHz, 2GHz–5GHz
b_l antennas, RIS elements	4, 50
Maximum p_m, P_m^{rel}, β_0 and σ^2	0.5W, 1W, -50dB, and -90dBm
$av_i - r_j$ link Rician factor	10dB
$r_j - b_l$ link Rician factor	10dB
$gu_m - av_i$ link Rician factor	3dB
$r_j - b_l$ link path loss exponent	2dB
$gu_m - av_i$ link path loss exponent	3.5dB
$av_i - r_j$ link path loss exponent	2.8dB
Cellular Bandwidth	10MHz
A2C Algorithm Parameter	
Time steps	100k
Update window size per step	100
γ, GAE λ, and Learning rate	0.99,1.0, 0.0005
Value function coefficient	0.5

IV. SIMULATION SETUP, RESULTS AND DISCUSSION

In this section, we present the experimental setup and outcomes of our study on optimizing task offloading and relaying in a UAV-assisted MEC system while minimizing delays and energy consumption. The underlying subsection contains the details regarding simulation parameters, and configurations. The subsequent section contains the achieved results, discussing the DEETO algorithm's performance.

A. SIMULATION SETUP

We consider a UAV-assisted MEC based communication environment with RIS enhancement. The network is equipped with BTS having the communication range radius of 200m with ze-RISs support. Similarly, the UAV have a communication range radius of 500m. The positions of gu_m users are circularly distributed in an area around the UAV with 100m radius, randomly. We configure the channel parameters by equipping the ze-RIS with a uniform planar array (UPA), organized in an $M_x \times M_y$ grid layout. Employing the Rician fading model, our channel model encompasses both line-of-sight (LoS) and non-LoS components for the links between the $gu_m - av_i$, $av_i - b_l$, $av_i - r_j$, and $r_j - b_l$. The path loss is characterized by $PL(\alpha, d) = \beta_0(d/d_0) - \alpha$, where d represents the transmission distance, β_0 denotes the path loss at the reference distance $d_0 = 1$m, and is set to $-40dB$. The α signifies the path loss exponent. The task attributes, including size and processing time thresholds, are selected at random from predefined intervals. Task sizes vary from 1 to 50Mb, while processing thresholds range from 200 to 500ms, and 1–50 joules of energy units to process it, accordingly. Additionally, our system incorporates a sub-task feature, enabling each task to comprise a flexible number of sub-tasks, ranging from 1 to 100. The simulations were executed on a M1-MacBook, using Python 3.7, PyTorch 2.0, Spyder, and Jupyter Notebook IDEs were employed for implementation. The DRL algorithms adhered to OpenAI standards. Other simulation parametric values and hyperparameters are given in Table 2.

For thorough analysis, we conducted convergence tests on

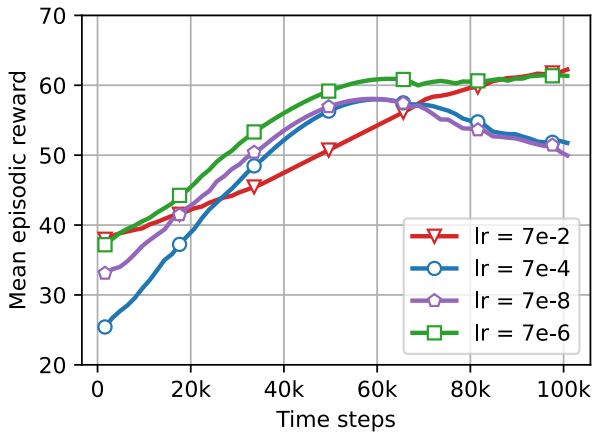


FIGURE 2: Convergence analysis of A2C based DEETO scheme for different learning rates.

the proposed DEETO scheme using various A2C DRL agent learning rates. Specifically, we evaluated the A2C algorithm's performance with learning rates set at $7e^{-2}$, $7e^{-4}$, $7e^{-6}$, and $7e^{-8}$. The convergence and reward collection analysis, depicted in Fig. 2, revealed distinct patterns. Notably, results indicated that the $7e^{-4}$ and $7e^{-8}$ learning rates achieved earlier convergence compared to $7e^{-2}$ and $7e^{-6}$, albeit experiencing declines in performance over additional time steps. While the convergence pattern for $7e^{-2}$ appeared superior to $7e^{-4}$ and $7e^{-8}$, it exhibited lower reward collection than $7e^{-6}$. Remarkably, the A2C algorithm demonstrated notable efficacy with DEETO at a learning rate of $7e^{-6}$, exhibiting enhanced convergence and higher rewards. Consequently, we have selected $7e^{-6}$ as the learning rate for subsequent experimental analyses.

B. RESULTS AND DISCUSSION

The efficiency of the DEETO scheme is systematically examined in dynamical scenarios, comparing it with other candidate DRL-based offloading techniques. we have evaluated and analyzed these schemes in terms of rewards, time frames per second (TFPS), task completion rate, energy consumption rate, and dexterity factors. Confirming an adequate comparison, all schemes are assessed using consistent simulation parameters. we have compared the DREEO scheme with the following contender schemes:

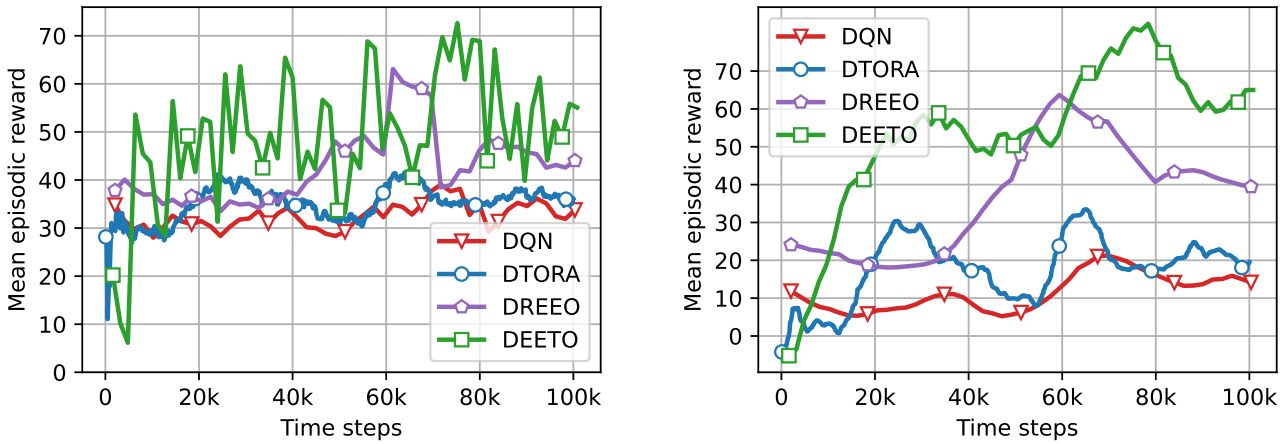
- *DEETO*: This is the proposed scheme.
- *DREEO*: This is a DRL based energy optimizing task offloading scheme. DREEO is an RIS assisted scheme, comprising hybrid task offloading scheme including binary and partial offloading mechanisms [20].
- *DTORA*: This is an RIS enabled UAV based MEC task offloading scheme. The DTORA optimizes task offloading and resource allocation decision using double DQN algorithm. These decisions target to mitigate UAV energy consumption through optimizing computing and communication resources and RIS phase shifts [30].
- *DQN*: This is an other DRL driven scheme to optimize task offloading delays and energy consumption. The DQN algorithm is adopted to optimizing task offloading decisions.

Several assessments have been made to verify and affirm the efficacy of our proposed DEETO scheme in a dynamic UAV-centric edge computing network environment. Fig. 3 provides a comprehensive comparison between DEETO and other schemes concerning rewards, a key metric in DRL-based approaches. Fig. 3(a) illustrates raw episodic rewards, while Fig. 3(b) depicts smoothed episodic rewards across time steps. Here, the DRL agent interacts with the environment, receiving rewards or penalties based on its actions. As the iteration with the environment progresses, the DRL agent adapts its policy to optimize its reward value. Favorable actions corresponding to the current environmental state yield higher rewards, whereas unfavorable actions result in reduced or negative rewards (penalties).

Besides, in the context of evaluating DRL based algorithms, apart from the comparison of rewards, time frames per seconds (TFPS) serves as a crucial performance metric. TFPS refers to the rate at which an algorithm processes the environment state steps within a given timeframe. A lower TFPS indicates that the algorithm can process each time frame more efficiently, meaning it can accomplish its tasks or decision-making processes in less time. Therefore, achieving a lower TFPS over time signifies improved performance and efficiency. Fig. 4 provides an illustration of the TFPS for all the contender approaches.

The observations indicate that the proposed DEETO scheme initially shows a higher TFPS, which gradually decreases over time and eventually converges to the lowest rate compared to all competing schemes. DEETO scheme achieves the lowest average TFPS rate of 870 among all other schemes. The DREEO scheme closely follows the characteristics of the DEETO scheme in terms of TFPS rate, yet it concludes with a slightly higher average TFPS of 910. Contrarily, the DTORA scheme starts with a higher TFPS, experiences a temporary decrease, but eventually shows an increasing trend over time, and finally converging at an average TFPS of 992. In contrast, the DQN-based scheme performs the poorest in terms of TFPS, obtaining an average TFPS rate of 1048.

The DRL characteristics shown in Fig. 3 and Fig. 4 illustrate the performance of each scheme in terms of rewards and TFPS, respectively. However, evaluating both features simultaneously unveils optimal algorithmic performance when higher rewards are achieved at lower TFPS rates collectively. It is understandable that the proposed DEETO scheme outperforms all contestant schemes while adhering to this measure. This performance gap between DEETO and other schemes can be accredited to several advantages of employing the A2C algorithm. A2C's simplicity, combining policy gradients and value-based methods, along with lower variance, renders it more robust. These attributes make it



(a) Reward Raw

(b) Reward Smoothed

FIGURE 3: Evaluation of mean episodic rewards across different DRL based offloading schemes.

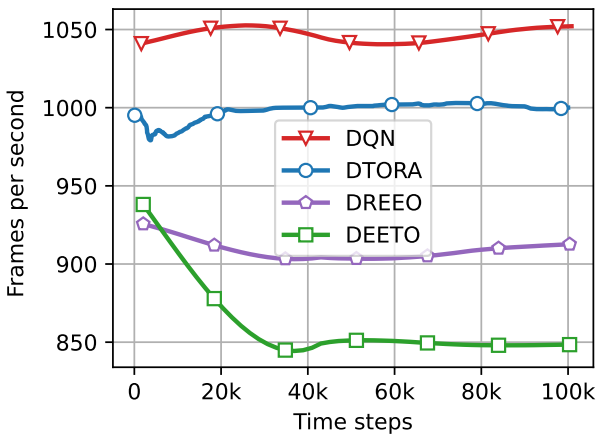


FIGURE 4: Evaluation of time frames per second across completing DRL based offloading schemes.

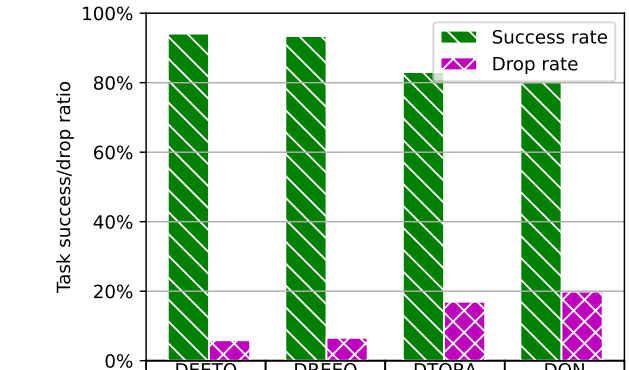
predominantly suitable for real-time learning scenarios, enabling firmer convergence and compliance to the dynamic communication environments.

Apart from the DRL characteristics, it is necessary to consider the interacting environmental factors in which the DRL agent operates. Therefore, to assess the effectiveness of these task offloading schemes we have analyzed the task turnover rate and energy consumption rate, as shown in Fig. 5(a) and Fig. 5(b). The analysis illustrates how accurately and effectively task offloading decisions are made, balancing task completion and energy optimization considerations. This decision involves determining whether to execute tasks locally, on the UAV, or to relay them to the edge server. For the DEETO and DREEO schemes, there exists an additional dimension of decision-making, i.e., whether task processing/offloading occurs in a binary or partial manner. Fig. 5(a) illustrates the analysis, representing that all schemes performed worthily in terms of task success and drop rates, although with slight variations. However, the DEETO scheme achieved the highest task success rate of 94.12% and the lowest drop rate of 05.87%. The DREEO scheme closely followed, achieving a 93.42% task success rate and a 06.58% drop rate. Conversely, the DTORA and DQN schemes achieved task success rates of 83.04% and 80.13%, with corresponding task drop rates of 16.96% and 19.87%, respectively.

Similarly, Fig. 5(b) depicts the percentage of energy consumed and saved during task processing and relaying. All schemes successfully conserved energy within the allocated resources while observing to delay and task completion constraints. The DREEO scheme demonstrated the highest energy savings, conserving 16.98% of the allocated energy resources. Moreover, DREEO remained a close competitor to the DEETO scheme, achieving energy savings of 12.51%. In comparison, the DTORA and DQN schemes saved 12.22% and 8.32% of the allocated energy resources, respectively.

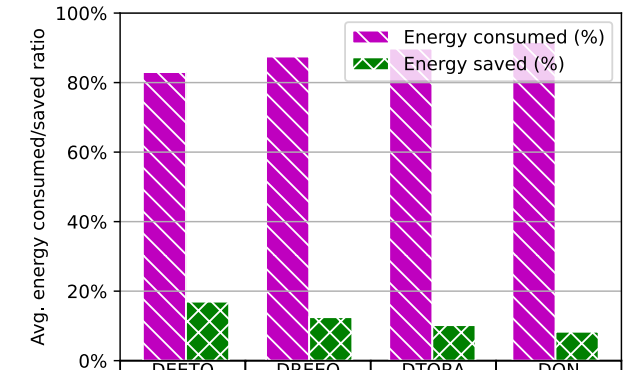
In addition to the aforementioned analyses, we conducted a comparative investigation involving the dexterity factor across varying numbers of tasks, task threshold times, and task sizes. The dexterity factor is composed as the sum of differences between consumed time and benchmark time, as well as consumed energy and benchmark energy.

Fig. 6a illustrates the comparative performance of all offloading schemes as the number of tasks increases. It is evident that an increase in the number of tasks leads to higher delay and energy consumption. However, the dexterity factor also increases concurrently, as it represents the average sum of saved energy and time. Notably, both the DREEO and DEETO schemes exhibit a similar pattern of energy and time savings, although with differing scales. This similarity can be attributed to the shared utilization of a hybrid task offloading mechanism. However, DEETO achieves the highest average dexterity factor, reaching 239.07 against 209.79 of



	DEETO	DREEO	DTORA	DQN
Success rate	94.12 %	93.42 %	83.04 %	80.13 %
Drop rate	05.87 %	06.58 %	16.96 %	19.87 %

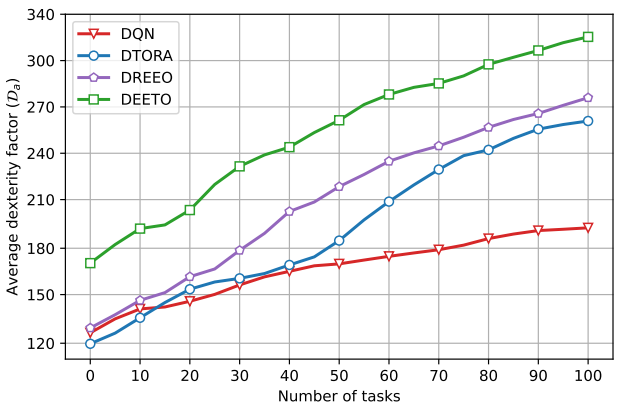
(a) task rate



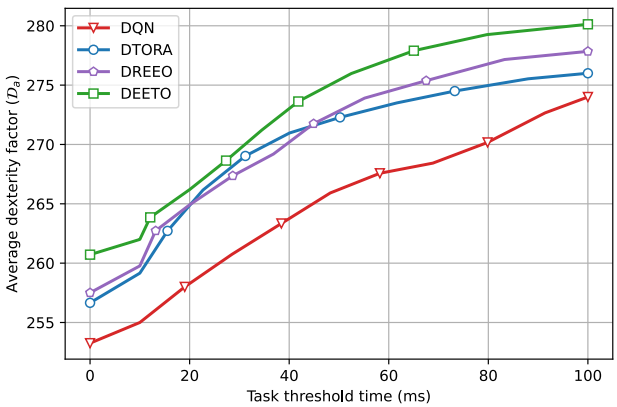
	DEETO	DREEO	DTORA	DQN
Consumed	83.02 %	87.49 %	89.78 %	91.68 %
Saved	16.98 %	12.51 %	10.22 %	08.32 %

(b) energy rate

FIGURE 5: Task completion and energy consumption of various DRL based schemes.



(a) Dexterity factor vs increasing number of tasks



(b) Dexterity factor vs increasing threshold time (t_{max})

FIGURE 6: Evaluation of dexterity factor against increasing number of tasks and the task threshold time

DREEO, 195.55 of DTORA, and 178.79 of DQN scheme. The superior decision-making mechanism of DEETO results in a higher dexterity factor, reflecting the cumulative savings in energy and time. This further validates the efficacy of relaying through ze-RIS via efficient phase shift mechanisms in reducing UAV energy consumption.

Expanding the analysis, we examined the performance of the DEETO scheme against the other schemes while increasing the threshold time, as depicted in Fig. 6b. It is evident that at lower threshold times, DEETO, DREEO, and DTORA exhibit similar performance levels, but as the threshold time increases, DEETO emerges as the superior scheme. However, the DQN scheme demonstrates the least efficiency. The proposed DEETO scheme achieved an average dexterity factor of 270.89, closely followed by DREEO with an average dexterity factor of 268.90. DTORA also closely trails DREEO with an average dexterity factor of 266.81, while the DQN scheme attained an average dexterity factor of 264.45.

The Fig. 7 illustrates the relationship of dexterity factor against increasing task size. it is obvious that as the task size increases the computation and communication over head also increases. the results show that DEETO and the DREEO schemes again has the same pattern of achieving dexterity factor, but the DEETO scheme has a smooth non linear trend as the task size increases and tends to convergence. However all the other schemes has a near linear sharp decreasing trend of dexterity factor with the increasing task size. The DEETO scheme achieved the highest average dexterity factor of 282.32, while DREEO, DTORA, and DQN achieved average dexterity factor of 277.65, 273.26, and 271.84, respectively.

The DREEO scheme emerged as the decent performer among all contender schemes due to its intelligent optimization of ze-RIS phase shifts coupled with a strategic hybrid task offloading mechanism. DREEO consistently remained a close competitor of DEETO across all evaluation results, leveraging the A2C algorithm alongside its hybrid task offloading approach, however with a random ze-RIS phase shift mechanism. In contrast, the DTORA scheme utilized the DDQN algorithm along with an intelligent RIS phase shift

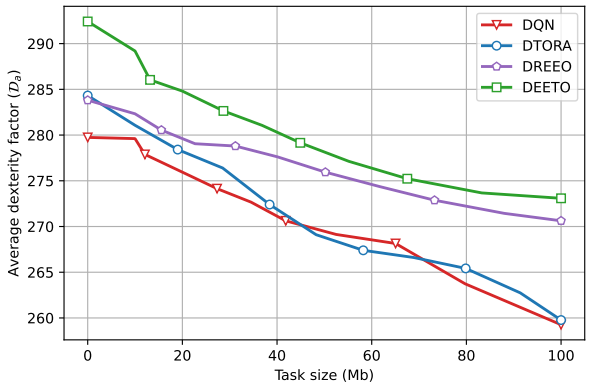


FIGURE 7: Dexterity factor vs increasing task size

mechanism but lacked a hybrid task offloading mechanism, resulting in slightly inferior performance. On the other hand, the DQN scheme, employing a deep Q-Network and employing a binary task offloading mechanism with random RIS phase shifts put it poorer scheme against all the competitor task offloading schemes.

V. CONCLUSIONS

This study addresses the challenge of sustaining high-quality communication services over extended periods in mobile edge computing (MEC) networks, particularly during emergencies, by leveraging UAVs and RIS. We introduced a comprehensive strategy that integrates ze-RIS into UAV-MEC networks, incorporating task offloading and resource allocation mechanisms. Our proposed DRL-driven energy-efficient task offloading (DEETO) scheme minimizes UAV energy consumption by optimizing task offloading decisions, UAV computing resource allocation, communication resource allocation, and RIS phase shift control, employing a hybrid task offloading mechanism. We modeled this problem as a DRL problem and structured it as an MDP and employed A2C algorithm due to its simplicity and less complexity. Simulation results demonstrate that the DREEO scheme emerged as a strong performer, utilizing intelligent ze-RIS phase shift optimization and a strategic hybrid task offloading mechanism. DEETO scheme showcased a significant energy savings of 16.98% from allocated resources, alongside the highest task turnover rate of 94.12%. Notably, DEETO scheme achieved these outcomes within shorter average learning time frames per second (TFPS) of 870 while yielding higher rewards of 57.38, underscoring its effectiveness over alternative approaches.

REFERENCES

[1] M. Y. Akhlaqi and Z. B. M. Hanapi, "Task offloading paradigm in mobile edge computing-current issues, adopted approaches, and future directions," Journal of Network and Computer Applications, vol. 212, p. 103568, 2023.

[2] Y. Bai, H. Zhao, X. Zhang, Z. Chang, R. Jäntti, and K. Yang, "Towards autonomous multi-uav wireless network: A survey of reinforcement learning-based approaches," IEEE Communications Surveys & Tutorials, 2023.

[3] M. N. Kharil, M. S. Johal, F. Idris, and N. Hashim, "Uav-enabled communications using noma for 5g and beyond: research challenges and opportunities," Indonesian Journal of Electrical Engineering and Computer Science, vol. 31, no. 3, pp. 1420–1432, 2023.

[4] F. Pervez, A. Sultana, C. Yang, and L. Zhao, "Energy and latency efficient joint communication and computation optimization in a multi-uav assisted mec network," IEEE Transactions on Wireless Communications, 2023.

[5] T. Sharma, A. Chehri, and P. Fortier, "Reconfigurable intelligent surfaces for 5g and beyond wireless communications: A comprehensive survey," Energies, vol. 14, no. 24, p. 8219, 2021.

[6] F. Naaz, A. Nauman, T. Khurshaid, and S.-W. Kim, "Empowering the vehicular network with ris technology: A state-of-the-art review," Sensors, vol. 24, no. 2, p. 337, 2024.

[7] M. Ahmed, A. Wahid, S. S. Laique, W. U. Khan, A. Ihsan, F. Xu, S. Chatzinotas, and Z. Han, "A survey on star-ris: Use cases, recent advances, and future research challenges," IEEE Internet of Things Journal, 2023.

[8] F. Xu, T. Hussain, M. Ahmed, K. Ali, M. A. Mirza, W. U. Khan, A. Ihsan, and Z. Han, "The state of ai-empowered backscatter communications: A comprehensive survey," IEEE Internet of Things Journal, 2023.

[9] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for internet of things," IEEE Wireless Communications Letters, vol. 10, no. 3, pp. 619–623, 2020.

[10] Y. Jiang, F. Gao, M. Jian, S. Zhang, and W. Zhang, "Reconfigurable intelligent surface for near field communications: Beamforming and sensing," IEEE Transactions on Wireless Communications, vol. 22, no. 5, pp. 3447–3459, 2023.

[11] Y. Gao, Q. Wu, W. Chen, Y. Liu, M. Li, and D. B. Da Costa, "Irs-aided overloaded multi-antenna systems: Joint user grouping and resource allocation," IEEE Transactions on Wireless Communications, pp. 1–1, 2024.

[12] M. Liwang, Z. Gao, and X. Wang, "Let's trade in the future! a futures-enabled fast resource trading mechanism in edge computing-assisted uav networks," IEEE Journal on Selected Areas in Communications, vol. 39, no. 11, pp. 3252–3270, 2021.

[13] X. Hu, C. Masouros, and K.-K. Wong, "Reconfigurable intelligent surface aided mobile edge computing: From optimization-based to location-only learning-based solutions," IEEE Transactions on Communications, vol. 69, no. 6, pp. 3709–3725, 2021.

[14] S. Mao, N. Zhang, L. Liu, J. Wu, M. Dong, K. Ota, T. Liu, and D. Wu, "Computation rate maximization for intelligent reflecting surface enhanced wireless powered mobile edge computing networks," IEEE Transactions on Vehicular Technology, vol. 70, no. 10, pp. 10 820–10 831, 2021.

[15] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Uav-assisted and intelligent reflecting surfaces-supported terahertz communications," IEEE Wireless Communications Letters, vol. 10, no. 6, pp. 1256–1260, 2021.

[16] X. Pang, N. Zhao, J. Tang, C. Wu, D. Niyato, and K.-K. Wong, "Irs-assisted secure uav transmission via joint trajectory and beamforming design," IEEE Transactions on Communications, vol. 70, no. 2, pp. 1140–1152, 2021.

[17] G. Sun, X. Tao, N. Li, and J. Xu, "Intelligent reflecting surface and uav assisted secrecy communication in millimeter-wave networks," IEEE Transactions on Vehicular Technology, vol. 70, no. 11, pp. 11 949–11 961, 2021.

[18] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in uav-ris wireless networks," IEEE Journal on Selected Areas in Communications, vol. 39, no. 7, pp. 2042–2055, 2020.

[19] Y. Li, Y. Zou, H. Hui, J. Zhu, and B. Ning, "Improving computing capability for active ris assisted noma-mec networks," IEEE Wireless Communications Letters, 2024.

[20] M. A. Mirza, J. Yu, M. Ahmed, S. Raza, W. U. Khan, F. Xu, and A. Nauman, "Drl-driven zero-ris assisted energy-efficient task offloading in vehicular edge computing networks," Journal of King Saud University-Computer and Information Sciences, vol. 35, no. 10, p. 101837, 2023.

[21] X. Qin, Z. Song, T. Hou, W. Yu, J. Wang, and X. Sun, "Joint resource allocation and configuration design for star-ris-enhanced wireless-powered mec," IEEE Transactions on Communications, vol. 71, no. 4, pp. 2381–2395, 2023.

[22] J. Zhao, Y. Zhu, X. Mu, K. Cai, Y. Liu, and L. Hanzo, "Simultaneously transmitting and reflecting reconfigurable intelligent surface (star-ris) assisted uav communications," IEEE Journal on Selected Areas in Communications, vol. 40, no. 10, pp. 3041–3056, 2022.

[23] A. V. Savkin, C. Huang, and W. Ni, "Joint multi-uav path planning and los communication for mobile-edge computing in iot networks with riss," IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2720–2727, 2022.

[24] L. Wang, K. Wang, C. Pan, and N. Aslam, "Joint trajectory and passive beamforming design for intelligent reflecting surface-aided uav communications: A deep reinforcement learning approach," IEEE Transactions on Mobile Computing, 2022.

[25] Y. Xu, T. Zhang, Y. Liu, D. Yang, L. Xiao, and M. Tao, "Computation capacity enhancement by joint uav and ris design in iot," IEEE Internet of Things Journal, vol. 9, no. 20, pp. 20 590–20 603, 2022.

[26] Z. Zhai, X. Dai, B. Duo, X. Wang, and X. Yuan, "Energy-efficient uav-mounted ris assisted mobile edge computing," IEEE Wireless Communications Letters, vol. 11, no. 12, pp. 2507–2511, 2022.

[27] Y. Xu, T. Zhang, Y. Zou, and Y. Liu, "Reconfigurable intelligence surface aided uav-mec systems with noma," IEEE Communications Letters, vol. 26, no. 9, pp. 2121–2125, 2022.

[28] C. Wang, M. Pang, G. Cui, X. Chang, F. Jiang, Y. Yao, and W. Wang, "Joint waveform design and detection in symbiotic ambient backscatter noma systems," IEEE Internet of Things Journal, vol. 10, no. 22, pp. 19 507–19 517, 2023.

[29] J. Yu, Y. Li, X. Liu, B. Sun, Y. Wu, and D. Hin-Kwok Tsang, "Irs assisted noma aided mobile edge computing with queue stability: Heterogeneous multi-agent reinforcement learning," IEEE Transactions on Wireless Communications, vol. 22, no. 7, pp. 4296–4312, 2023.

[30] Q. Liu, J. Han, and Q. Liu, "Joint task offloading and resource allocation for ris-assisted uav for mobile edge computing networks," in 2023 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2023, pp. 1–6.

[31] L. Linpei, Z. Chuan, S. Yu, H. Jiahao, H. Yao, and L. Haojin, "Energy-efficient computation offloading assisted by ris-based uav," The Journal of China Universities of Posts and Telecommunications, vol. 31, no. 1, p. 37, 2024.

[32] S. Wang, X. Song, T. Song, and Y. Yang, "Fairness-aware computation offloading with trajectory optimization and phase-shift design in ris-assisted multi-uav mec network," IEEE Internet of Things Journal, 2024.

[33] H. Mei, K. Yang, Q. Liu, and K. Wang, "3d-trajectory and phase-shift design for ris-assisted uav systems using deep reinforcement learning," IEEE Transactions on Vehicular Technology, vol. 71, no. 3, pp. 3020–3029, 2022.

[34] D. Tyrovolas, S. A. Tegos, V. K. Papanikolaou, Y. Xiao, P.-V. Mekikis, P. D. Diamantoulakis, S. Ioannidis, C. K. Liaskos, and G. K. Karagiannidis, "Zero-energy reconfigurable intelligent surfaces (zeris)," arXiv preprint arXiv:2305.07686, 2023.

[35] Y. Zheng, S. A. Tegos, Y. Xiao, P. D. Diamantoulakis, Z. Ma, and G. K. Karagiannidis, "Zero-energy device networks with wireless-powered riss," IEEE Transactions on Vehicular Technology, 2023.

[36] S. Naser, L. Bariah, S. Muhaidat, and E. Basar, "Zero-energy devices empowered 6g networks: Opportunities, key technologies, and challenges," TechRxiv, 2023.

MUHAMMAD NAQQASH TARIQ obtained his Bachelor's degree (BEE) from The University of Faisalabad (TUF), Faisalabad in 2012, and his Master's degree in Electronics & Communication Engineering from the Beijing University of Posts and Telecommunications, Beijing, China in 2017. Currently, he is a Ph.D. candidate at the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include Unmanned Ariel vehicular communication networks, Software Defined Networking, fog and edge computing, internet of things, machine learning and deep reinforcement learning, and big data analytics.

JINGYU WANG (Senior Member, IEEE) received the Ph.D. degree from the Beijing University of Posts and Telecommunications, in 2008. He is currently a Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. He has published more than 100 articles in international journals, including IEEE Communications Magazine, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MULTIMEDIA, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. His research interests include the IoV and AIoT, SDN, overlay networks, and traffic engineering.

DR. SAIFULLAH MEMON received a master's degree from Quaid-e-awam University of Engineering, Science and Technology, (QUEST) in 2015. Onward he acquired his Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), China in 2023 in the field of communication engineering. He is the author of more than 30 research articles and is currently working as an Assistant Professor at QUEST. His research interests include wireless communication, sensors, wireless body area networks (WBAN), QoS issues in sensor and ad-hoc networks, routing algorithms, and network security.

MOHAMMAD SIRAJ (Senior Member IEEE) received his Ph.D. from Universiti Technologi Malaysia. He holds Master of Engineering degree in Computer Technology and Applications from Delhi College of Engineering, Delhi, India and Bachelor of Engineering degree in Electronics and Communication Engineering from Jamia Millia Islamia, New Delhi. He has worked as a Scientist in Defense Research and Development Organization, India. Currently he is working as an Assistant Professor in Electrical Engineering, King Saud University. His research interests are Cognitive Wireless Networks, Wireless Mesh Networks, Sensor Networks, Internet of Things, Cloud Computing and Telecom Optical Networks. He has numerous peered publications in well-known international journals and Conferences. He is a Senior Member IEEE and reviewer of many well-known International Journals and Conferences.

MAJID AL TAMIMI (Member, IEEE) received the B.Sc.Eng. degree (Hons.) in electrical engineering from King Saud University, Saudi Arabia, in 2004, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Canada, in 2010 and 2014, respectively. In 2015, he joined the Department of Electrical Engineering, King Saud University, as an Assistant Professor. He served as a TA with the Department of Electrical and Computer Engineering, University of Waterloo, in 2013, and the Department of Electrical Engineering, King Saud University, from 2004 to 2006. His current research interests include analyzing the energy cost for wireless handheld devices and cloud computing architecture, integrating mobile computing with cloud computing, and studying and designing green ICT solutions.

MUHAMMAD AYZED MIRZA received his BSCS degree from Govt. College University, Faisalabad in 2008 and his MSCS degree in Computer Science from the National Textile University, Faisalabad in 2016. He worked as a lecturer at Punjab College University Campus Faisalabad for 4 years and at National Textile University, Faisalabad from 2016 to 2018. He completed his Ph.D. in 2023 from the School of Electronic Engineering at Beijing University of Posts and Telecommunications, Beijing, China. Currently, he is working as Associate Professor at the School of Computer Science and Information Engineering, Qilu Institute of Technology, Jinan, Shandong 250200 P.R, China. His research focuses on vehicular communications and networks, the IoT/IoV, fog and edge computing, and task/data offloading techniques.