



Research paper

# Data collection in deep reinforcement learning-enhanced reconfigurable intelligent surface-assisted wireless networks

 Idris Ertas <sup>a</sup>, Halil Yetgin <sup>a,b</sup>\*,

<sup>a</sup> Department of EEE, Bitlis Eren University, Bes Minare Mah, Bitlis, 13100, Turkiye

<sup>b</sup> Department of Computer Science, Middlesex University, The Burroughs, London, NW4 4BT, UK


## ARTICLE INFO

## Keywords:

Deep reinforcement learning  
Data collection  
Unmanned aerial vehicle  
Internet of Things  
Reconfigurable intelligent surfaces

## ABSTRACT

In the evolving sixth generation (6G) landscape, the integration of reconfigurable intelligent surfaces (RIS) with unmanned aerial vehicles (UAVs) offers a revolutionary opportunity to optimize data collection in the Internet of things (IoT) through deep reinforcement learning (DRL) and improve energy efficiency and network performance. This paper aims to study how reconfigurable intelligent surfaces and deep reinforcement learning can help increase throughput and energy efficiency in unmanned aerial vehicle-controlled Internet of things networks. The focus is on improving the capabilities of unmanned aerial vehicles to efficiently collect data in different regions and ensure safe landings. Divided into two phases, the study first improves the directional capacity and flexibility of unmanned aerial vehicles and then evaluates the integration of reconfigurable intelligent surface technology. We introduce two deep reinforcement learning models, namely the directional capacity and flexible reconnaissance (DCFR) model and the reconfigurable intelligent surface model, and compare them with a benchmark model. We found significant improvements in communication and data collection efficiency. The simulation results show an 8.18% increase in data collection performance and a 6.92% increase in collected data per unit energy when using reconfigurable intelligent surfaces, with a 10.59% increase in collection performance and a 22.64% increase in energy efficiency. Furthermore, an unmanned aerial vehicle optimized with the double deep Q-network algorithm effectively identified optimal trajectories for data collection, confirming the significant benefits of reconfigurable intelligent surfaces in unmanned aerial vehicle-controlled Internet of things networks.

## 1. Introduction

In the era of ubiquitous connectivity, the integration of Internet of things (IoT) networks with unmanned aerial vehicles (UAVs) offers unprecedented opportunities for the intelligent collection and dissemination of data across vast and diverse areas. The dynamic nature of UAVs, coupled with their ability to quickly cover and adapt to different geographical areas, makes them an ideal carrier for IoT applications that require large-scale data interaction. However, the inherent limitations of UAV communication capabilities, such as throughput, range and reliability, require innovative solutions to improve signal quality and network performance. IoT networks constitute a communication infrastructure wherein billions of devices interconnect, enabling data sharing and interaction across various sectors such as health-care, energy, agriculture, and industrial automation, thereby impacting numerous facets of daily life (Atzori et al., 2010). However, the performance of IoT networks may be constrained due to limited resources and challenges such as energy consumption by devices within the

network. Consequently, managing IoT networks efficiently necessitates innovative solutions to enhance network performance (Gubbi et al., 2013; Zanella et al., 2014).

UAV hold significant potential for enhancing the performance and efficiency of IoT networks. For example, they can be an ideal solution for tasks such as data collection, routing, and enhancing communication capabilities within IoT networks (Mozaffari et al., 2019). UAVs are characterized by their low cost, rapid mobility, and expansive coverage capabilities, rendering them effective tools for deployment within IoT networks (Mozaffari et al., 2019).

In this context, reconfigurable intelligent surfaces (RIS) are proving to be a transformative technology that has the potential to significantly increase the efficiency of wireless communication networks. By intelligently manipulating electromagnetic waves, RIS can improve signal propagation and extend communication links between IoT devices and UAVs. In particular, RIS can facilitate the reduction of signal losses in communication, thereby enabling more secure communication between

\* Corresponding author at: Department of Computer Science, Middlesex University, The Burroughs, London, NW4 4BT, UK.  
E-mail address: [h.yetgin@mdx.ac.uk](mailto:h.yetgin@mdx.ac.uk) (H. Yetgin).

transmitters and receivers, particularly in scenarios where direct line-of-sight communication is obstructed (Huang et al., 2019; Mnih et al., 2013). This paper explores the role of RIS in extending the capabilities of IoT networks with UAVs, with a focus on how RIS can be used to optimize data collection processes.

There is a notable lack of research on RIS-based data collection using UAVs in combination with deep reinforcement learning (DRL) techniques. (Almasoud, 2023) develop a robust framework for data collection in IoT environments using UAVs and RIS enhanced by landing platforms for increased anti-jamming resilience. The study applies an ant colony optimization algorithm to refine the routing of UAVs and clustering of IoT devices. This introduces a novel method to improve the flexibility and robustness of the network. However, the focus remains on ant colony algorithms and resilience rather than dynamic adaptability. Chu et al. (2023) studied resource allocation and power distribution in a wirelessly operated IoT network supported by statically configured RIS elements to increase communication efficiency, but it lacked the integration of UAVs and adaptive learning algorithms, which revealed a gap in adapting to dynamic environments. Chen et al. (2023) addressed the energy constraints in IoT systems by utilizing active RIS for multiple access configurations, focusing on energy efficiency and improving signal quality, but omitting UAVs and deep learning strategies for real-time decision making. In contrast, Fan et al. (2023) successfully combined RIS with UAVs in urban environments to optimize data freshness using a DRL model, focusing primarily on Age of Information (AoI) rather than maximizing throughput. Similarly, the study conducted by Qi et al. (2024) reduces AoI in vehicular networks by integrating UAVs with RIS, using a multi-step duelling double deep Q-network (MSD3QN) to fine-tune UAV trajectories, RIS phase adjustments and spectrum allocation, significantly improving data timeliness and network coverage. Despite the progress, existing research generally overlooks the combination of throughput maximization, RIS-assisted data collection via UAVs and leveraging reinforcement learning.

Our research aims to comprehensively integrate all the aforementioned elements, as portrayed in Fig. 1, where users communicate with UAV through RIS in the interest of maximizing throughput. Therefore, UAV's trajectory decisions for efficient data collection becomes prominent. The aim of this paper is to design and develop a novel method for data collection by incorporating RIS-assisted throughput maximization in UAV-powered IoT networks considering DRL. The main contributions of this paper are as follows:

- We design and develop the directional capacity and flexible reconnaissance (DCFR) model to improve the performance of the Bayerlein model proposed in Bayerlein et al. (2020). The motivation behind developing the new DCFR model is to prevent the agent from getting stuck in a certain state.
- We provide a method to integrate RIS into an existing IoT network to improve the energy efficiency and data collection performance of the UAV. To this end, we have developed the RIS model to demonstrate the additional performance improvement in data collection and energy efficiency.
- We introduce a new metric, *collected data per unit energy*, to measure the energy efficiency performance of the RIS model which integrates the RIS technology into the existing IoT network. Together with the data collection ratio, this new metric is very useful to determine whether a model is energy efficient.
- We conducted a comprehensive analysis of the overall performance for each model, i.e. the Bayerlein model, the DCFR model and the RIS model. We also uncovered specific scenarios where the Bayerlein model stalls the agent and provided the DCFR solution to these challenges by analysing each scenario individually.

The rest of the paper is structured as follows. Section 2 discusses the related work while Section 3 elaborates on the proposed system model along with RIS system model. Section 4 presents the generic solution

approaches that are leveraged, whereas Section 5 discusses the DCFR model along with the integration of RIS as an extension to the DCFR model. Section 6 comprehensively evaluates all the proposed models. Finally, Section 7 concludes the paper.

## 2. Related work

Unmanned Aerial Vehicles (UAVs) have emerged as a pivotal technology in the realm of IoT networks, offering versatile and efficient solutions for data collection in diverse and challenging environments. Their mobility, adaptability, and cost-effectiveness have positioned them as essential tools in applications such as disaster management, smart agriculture, and environmental monitoring (Haider et al., 2022).

UAVs serve as mobile nodes capable of traversing complex terrains and collecting data dynamically from IoT devices. Unlike traditional static base stations, UAVs can adapt their trajectories in real time to optimize data collection based on environmental and network conditions. Their ability to establish line-of-sight (LoS) communication links further enhances their suitability for data collection tasks (Wang et al., 2024; Haider et al., 2022).

In IoT networks, UAVs facilitate efficient data gathering from sensors deployed in remote or inaccessible regions. Their utility in agriculture is particularly noteworthy, where they assist in monitoring crop health, optimizing irrigation schedules, and assessing soil conditions. UAVs enable timely data collection, which is critical for precision agriculture applications. By employing multi-objective optimization techniques, UAVs balance energy consumption, data rate, and latency, ensuring the sustainability and effectiveness of data collection operations (Haider et al., 2022; Liu et al., 2024).

Despite their advantages, UAV-based data collection faces significant challenges, including energy constraints, communication interference, and the need for timely data delivery. Limited onboard energy restricts the operational duration of UAVs, necessitating efficient energy management strategies. Advanced trajectory planning and clustering techniques address this issue by minimizing unnecessary movements and optimizing data collection routes (Wang et al., 2024; Haider et al., 2022).

Additionally, environmental factors such as obstacles and channel interference can hinder communication reliability. The integration of RIS has been shown to mitigate these issues by enhancing signal propagation and establishing reliable communication links in obstructed environments (Haider et al., 2022). Ensuring data timeliness is another critical concern, particularly for time-sensitive applications. Metrics such as Age of Information (AoI) are employed to evaluate and optimize data freshness during transmission (Wang et al., 2024; Mondal et al., 2024).

We categorize the related work into two subsections followed by the critical evaluation of these relevant works, the first one is concerned with the IoT data collection using DRL, whereas the second one focuses on the integration of RIS to enhance the data collection performance of UAVs.

### 2.1. Utilizing UAVs for IoT data collection

Chen et al. (2019) focus their attention on enhancing the efficiency of data collection in large-scale UAV-supported wireless sensor networks. UAVs are deployed for collecting sensor data to improve the effectiveness of the data collection process. The study divides the target area into clusters, and cluster head selection and data transmission rules for each cluster are determined based on the information value and power levels of the nodes. While sensor nodes collect environmental information in response to events, UAV serves as a mobile data collection centre. A direct future prediction (DFP) model is utilized to plan the trajectory of the UAV, maximizing the total information value collected while achieving multiple target tasks with low power consumption. Simulations demonstrate that the DFP model outperforms

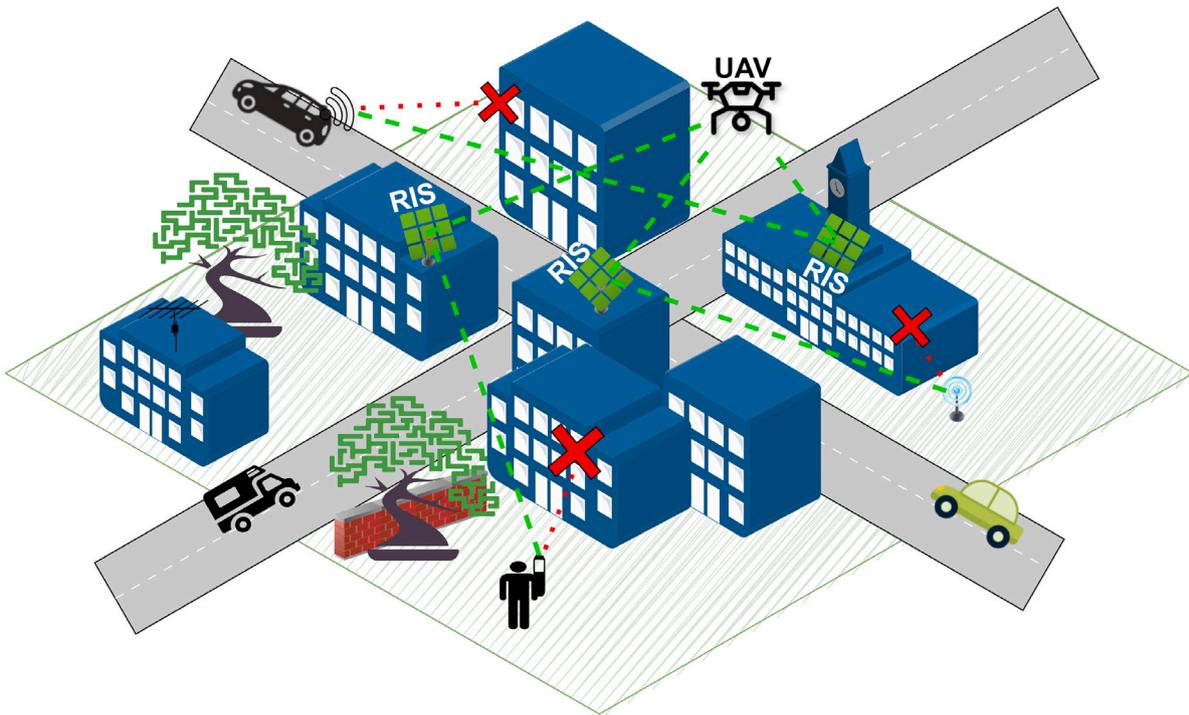


Fig. 1. Users to communicate with UAV through RIS for efficient data collection.

traditional methods. The approach demonstrates advantages such as efficient clustering and trajectory planning, which significantly enhance data collection efficiency and power consumption. However, the computational overhead in DFP implementation and potential scalability challenges for larger networks are notable disadvantages.

Bayerlein et al. (2020) address route planning and data collection in UAV-supported IoT networks using DRL-based models. Specifically, in a data collection scenario from IoT devices involving challenges, such as limited flight time and obstacle avoidance, regulatory restrictions, the learned control policy using double deep Q-network (DDQN) and experience replay stands out for its ability to adapt extensively to changes in environmental parameters. The method provides high adaptability to dynamic environments and effectively manages complex constraints such as obstacle avoidance and limited flight time. However, its disadvantages may include limited generalizability beyond trained scenarios and high training time and resource requirements for DRL models.

Cicek (2021), the aim is to enhance the efficiency of using UAV for data collection from sensors in IoT networks. The paper presents an approach addressing current issues, such as limited UAV battery and deterministic operating times by enabling UAV to swap batteries at stations where battery replacement is conducted. Additionally, in a network with sensors where data upload times are uncertain, the UAV's journey planning and battery replacement times are jointly optimized to minimize total data loss. The proposed reinforcement learning algorithm outperforms other methods in the computations conducted and achieves a significant reduction in data loss. This approach is advantageous for effective battery management and significant reduction in data loss through optimized journey planning. However, it relies heavily on the availability of battery swap stations and has limited flexibility for real-time adjustments.

Benmad et al. (2022) examine the utilization of multiple UAVs and the determination of optimal trajectories in the data collection process of wireless sensor networks (WSNs). The rapid deployment capabilities and access to challenging terrains of UAV provide significant advantages to the data collection process. However, energy constraints are a significant concern for both WSN and UAV. Therefore, this study propose to jointly optimize trajectories in order to optimize

energy transfer for all UAV and minimize data collection time. These approaches utilize intuitive methods such as nearest insertion algorithm (NIA) and greedy algorithm (GA). It is concluded that the proposed methods effectively facilitate the efficient utilization of UAV in the data collection process and can enhance the performance of WSNs. The study demonstrates advantages in optimizing energy transfer and enabling faster data collection in challenging terrains. However, it faces limitations in the scalability of heuristic methods and potential inefficiencies in highly dynamic environments.

Li et al. (2022) investigate the trajectory design problem for UAV used in data collection systems for large-scale time-sensitive IoT services. Although UAV have advantages such as automatic maneuverability and flexible mobility, ensuring the freshness of collected data under the constraint of limited flight energy is a challenging task. Therefore, a new metric called age of information (AoI) is utilized to measure information freshness, and the problem of jointly optimizing power control and trajectory design to minimize average AoI is addressed. To solve this complex problem, an approach based on DRL-based multi-agent method is proposed by decomposing power control and trajectory design into independent subtasks. Simulation results demonstrate that the proposed method outperforms other methods in terms of performance and stability. The proposed method offers advantages in improving data freshness using the AoI metric and achieving superior performance and stability through multi-agent DRL approaches. However, the complexity in implementing multi-agent DRL and the resource intensity required for real-time operations are significant drawbacks.

In a nutshell, these studies highlight the integration of UAVs in IoT data collection and the use of DRL models to improve efficiency and performance in wireless sensor networks. Chen et al. (2019) show how UAVs guided by a DFP model optimize data collection by dynamically planning trajectories based on the information value and energy supply of the nodes, significantly outperforming conventional methods. Bayerlein et al. (2020) use DDQN to navigate UAVs through complex environments with constraints such as limited flight time and obstacles, demonstrating the adaptability of the model. Cicek (2021) focus on minimizing data loss by optimizing UAV flight paths and battery management, while Benmad et al. (2022) investigate energy-efficient

trajectory planning for multiple UAVs to speed up data collection in difficult terrain. Finally, Li et al. (2022) address the timeliness of data in IoT services using an “Age of Information” metric and a multi-agent DRL approach, achieving superior performance and stability in trajectory and energy management. Overall, these studies highlight the central role of DRL in revolutionizing UAV-based data collection strategies in IoT frameworks, promising significant improvements in network efficiency, data relevance and operational autonomy.

## 2.2. UAV data collection enhanced by RIS

Reconfigurable intelligent surfaces (RIS) are increasingly recognized as a key technology for future sixth generation (6G) networks to improve network capacity, coverage, efficiency and security while minimizing energy and hardware costs. However, the integration of RIS into current infrastructures significantly complicates network management, especially when controlling numerous RIS elements. Efficient optimization strategies are crucial to fully exploit the capabilities of RIS. To this end, Zhou et al. (2023) provide a comprehensive analysis of the AI-based RIS assisted communication networks, where problem formulations with different objectives and constraints are investigated, a comparison of optimization strategies in terms of stability, robustness and optimality is presented and their applications to 6G networks are discussed, also highlighting future challenges.

Bjornson et al. (2020) evaluate the performance of RIS compared to traditional Decode-and-Forward (DF) relaying to improve the efficiency of wireless networks. Through rigorous analytical and numerical analyses, the study shows that despite its potential to manipulate signal propagation via software-controlled metasurfaces, RIS requires a significantly larger number of elements to outperform DF relaying in terms of energy efficiency and signal strength. The results emphasize that while RIS can theoretically reduce transmission power and improve energy efficiency, especially at higher data rates, its practical use requires large-scale implementation, which may limit its utility in scenarios where compact and less resource-intensive solutions are preferable. This article highlights the critical balance between the size of the RIS element and its effectiveness in real-world wireless communication environments.

Al-Hilo et al. (2023) investigate the integration of RIS with UAVs to optimize data collection in IoT networks under time pressure. Using DRL for trajectory planning and block coordinate descent for RIS configuration, the study shows that such integration can significantly improve the connectivity and energy efficiency of UAVs in data collection. Detailed simulations show that the method is more than 50% superior to traditional approaches. This emphasizes the potential of RIS to transform the handling of IoT data in urban environments where line of sight is often obstructed. The method demonstrates significant advantages in improving connectivity and energy efficiency, particularly under time-constrained conditions. However, the reliance on complex configurations and trajectory planning algorithms may present computational challenges.

Mondal et al. (2022) focus on the treatment of inter-node interference in IoT networks through RIS. They develop optimization techniques to maximize spectrum and energy efficiency using phase shifts of RIS elements and power allocation between IoT devices. The study addresses non-convex optimization problems by dividing them into subproblems that are solved alternately and employing advanced algorithms such as conjugate gradients on Riemannian manifolds and pricing-based techniques. The numerical results show significant efficiency gains compared to traditional networks and highlight the effectiveness of RIS in improving the performance of IoT networks while keeping computational complexity manageable. The approach offers advantages in significantly improving spectrum and energy efficiency while keeping computational complexity manageable. However, its reliance on solving complex non-convex problems may limit its scalability to larger networks.

Yao et al. (2022) investigate the use of RIS on UAVs to increase the energy efficiency of wireless powered air-to-ground communication networks (WPCNs). A system is presented in which UAVs reflect energy via RIS to energy-constrained hybrid access points (HAPs), which then communicate with users in blind areas. The study models energy efficiency as a non-convex quadratic programming problem and proposes a joint optimization algorithm for power and phase shifts. The simulation results show that the proposed RIS-based model achieves an energy efficiency improvement of up to 160% compared to conventional amplify-and-forward relay systems. This emphasizes the potential of RIS to significantly improve the performance of air-to-ground communication networks. The proposed system achieves significant advantages in energy efficiency, with improvements of up to 160% compared to conventional systems. However, addressing the non-convex nature of the problem requires substantial computational resources.

Mei et al. (2022) employ RIS to enhance communication between UAVs and ground terminals by focusing signals between UAVs and ground terminals towards the intended users. DRL is utilized to optimize the 3D position of UAVs and the phase shift of RIS, demonstrating that DRL can significantly enhance the energy efficiency of RIS-enabled UAV systems.

Cao et al. (2021) propose a non-stationary geometry-based channel model for multiple-input multiple-output (MIMO) channels using RIS applied to UAVs. The adjustable reflection phases of RIS are optimized based on the received signal power. The study investigates the impact of the number of RIS reflective units, geometric area, and UAV speed over channel statistics. Results demonstrate that optimizing the phases of signals reflected by RIS can increase received signal power and mitigate the effects of multipath fading, whereas increasing the number of RIS reflective units significantly reduces spatial correlation.

Lian et al. (2023) propose a new channel model using RIS to enhance the performance of UAV communication systems. The proposed model considers the geometry and scattering characteristics of RIS reflective units. Additionally, the impact of the number and size of RIS reflective units on channel characteristics is examined. Results show that communication performance can be improved with RIS by increasing the number and size of reflective units. The model demonstrates advantages in improving communication performance by leveraging the size and number of RIS reflective units. However, the scalability and cost implications of increasing reflective unit size and numbers are potential drawbacks.

Wang and Zhang (2023) propose a novel approach to minimize energy consumption in UAV-based communication systems. By utilizing active RIS technologies, the power of communication signals between UAVs and ground users is effectively enhanced. Using a hierarchical DRL model, energy consumption of UAVs and ground users is concurrently minimized. Simulations demonstrate that active RIS usage can significantly reduce energy consumption of UAVs and ground users when the thermal noise power on RIS is substantially lower than that of UAVs and ground users.

Al-Jarrah et al. (2021) analyse the impact of RIS used for UAV communication on system capacity. In the examined system, RIS panels on some UAVs modify the phase of incoming waves before being reflected to the receiving UAV, while the effect of phase errors on capacity is also considered. Results indicate that phase errors affect capacity, but at high signal-to-noise ratio (SNR) and above a certain threshold, phase errors become negligible.

Park et al. (2022) proposes a novel framework to improve wireless connectivity in smart railways using an RIS-assisted UAV system. This system aims to maximize the minimum achievable rate of trains by jointly optimizing the trajectories of the UAVs and the phase shifts of the RIS. The challenges posed by non-convex problems are addressed by decomposing them into sub-problems, which are then solved using Binary Integer Linear Programming and Soft Actor-Critic methods. Extensive numerical simulations confirm the effectiveness of the proposed

model and show significant improvements in data rate and connectivity for intelligent orbital systems. The integration of RIS with UAVs offers a promising solution for overcoming obstacles and improving seamless communication in high-speed railway environments. The framework demonstrates significant advantages in enhancing data rates and connectivity in high-speed railway environments. However, addressing the non-convex challenges requires substantial computational resources, which may limit real-time deployment.

Wang et al. (2024) introduces a framework to address key challenges in UAV-assisted IoT networks, focusing on improving data timeliness and energy efficiency while mitigating the effects of jamming. By integrating RIS to enhance signal quality, the framework optimizes UAV trajectories using an innovative particle swarm algorithm and adjusts IoT device power through convex approximation techniques. The combined approach effectively balances energy consumption and communication reliability, showing significant improvements in data freshness and anti-jamming capabilities. While the framework offers advantages like enhanced connectivity and adaptability, it also faces challenges such as high computational demands, reliance on RIS placement, and scalability limitations in dense networks.

As the aforementioned literature highlights, the integration of RIS with UAVs and IoT systems significantly improves the capabilities of 6G networks. Techniques such as DRL can optimize the trajectories of UAVs and RIS setups and improving communication efficiency. RIS technology increases signal strength and reduces energy consumption, albeit it may increase the complexity due to non-convex optimization of the resources. The use of RIS in UAVs in urban areas increases the efficiency of data collection and shows that it is possible to solve line-of-sight problems and extend network coverage. In addition, developments in phase shift design and power optimization contribute to significant improvements in spectrum and energy efficiency. This enables RIS to fulfil the requirements of future technologies with high capacity and low latency.

### 2.3. Critical evaluation

It is clear from the aforementioned related work that the integration of UAVs and IoT networks has demonstrated remarkable potential in addressing data collection challenges in complex and dynamic environments. However, the inherent limitations of UAVs, such as restricted energy capacity and communication range, and the complexity introduced by non-convex optimization in RIS-assisted networks, have presented significant challenges. This section critically analyses key studies in this domain, discussing their contributions, advantages, and limitations, while positioning our proposed work as a novel contribution to overcome some of these shortcomings.

The performance comparison among existing works highlights distinct advantages and limitations that frame the context for this study. Chen et al. (2019) demonstrated efficient clustering and trajectory planning for large-scale networks, yet their approach faced scalability challenges and computational overhead, which our study addresses by employing DRL for dynamic trajectory optimization. Bayerlein et al. utilized DDQN for high adaptability in dynamic environments but encountered agent stalling, which is resolved in our work through the DCFR model's enhanced robustness. Cicek (2021) focused on effective battery management and data loss reduction but relied on static battery swap stations, limiting real-time flexibility. Our integration of DRL and RIS mitigates this limitation by enabling dynamic adaptability. Mondal et al. (2024) achieved notable spectrum and energy efficiency gains through RIS phase optimization, though scalability to larger networks was hindered by non-convex problem-solving, a challenge we overcome by introducing the novel metric *collected data per unit energy* and RIS-based optimization. Fan et al. (2023) effectively reduced the Age of Information (AoI) using DRL models but did not emphasize throughput maximization, which is a core focus of our framework. Almasoud (2023) leveraged ant colony optimization for anti-jamming resilience

but lacked adaptability in dynamic conditions, a gap we fill with DRL-enhanced trajectory optimization. Wang et al. (2024) integrated RIS with particle swarm optimization (PSO) and convex approximation for anti-jamming and energy efficiency, but their approach faced high computational demands and scalability challenges in dense networks. Our study addresses these gaps by introducing the novel metric *collected data per unit energy* and improving energy-efficient data collection in randomized IoT deployments using RIS integration along with DRL. This comparison underscores the distinct contributions of our study in addressing scalability, adaptability, and performance metrics while bridging the limitations of prior works. A summary of the critical evaluation is provided in Table 1.

## 3. System model

Bayerlein's research in Bayerlein et al. (2020) served as the foundational system model for this study, with modifications incorporated. The subsequent section provides a concise summary of the adapted system model.

### 3.1. Map, scheduling and communication models

*a- Map and system structure.* As in Bayerlein et al. (2020), we follow the 2D Manhattan grid map model that is illustrated at the top of Fig. 2, which also portrays an abstracted 3D view of the general system model encompassing IoT devices, UAV and RIS as well as NFZ and landing zone. UAV collects data from  $K$  number of IoT devices and each IoT device is located on the ground with a position  $u_k = [x_k, y_k, 0]^T \in \mathbb{R}^3$  and  $k \in [1, K]$ .

*b- Scheduling model.* The data collection task is performed over a specific time period  $T \in \mathbb{N}$ . Duration of UAV operation is divided into equally spaced mission time slots  $t \in [0, T]$ . The UAV's position  $[x(t), y(t), h]^T \in \mathbb{R}^3$  with constant altitude  $h$  and its 2D projection on the ground is determined by  $\mathbf{p}(t) = [x(t), y(t)]^T$ . Each mission time slot is divided into smaller communication time slots  $n \in [0, N]$ . Scheduling is arranged in a way so that node  $k$  initiates transmission towards the UAV at time slot  $n$ , which is represented by  $q_k(n) \in \{0, 1\}$ .

*c- Communication model.* The communication between the UAV and the IoT devices is modelled with line-of-sight (LoS)/Non-LoS point-to-point channels using log-distance path loss and shadowing. The information rate at time  $n$  for the  $k$ th device is defined as (Bayerlein et al., 2020)

$$R_k(n) = \log_2 \left( 1 + \underbrace{\left( \frac{P_k}{\sigma^2} \cdot d_k(n)^{-\alpha_l} \cdot 10^{\eta_l/10} \right)}_{SNR} \right), \quad (1)$$

where  $P_k$  is the transmit power, white Gaussian noise power at the receiver is  $\sigma^2$ , distance between UAV and IoT device is depicted as  $d_k(n)$ , path loss exponent is  $\alpha_l$  and  $\eta_l \sim \mathcal{N}(0, \sigma_l^2)$  is the Gaussian random variable. Note that SNR in Eq. (1) is averaged over small scale fading. An IoT devices is served by the UAV in each communication time slot  $n$  for uploading the remaining data given that the link with the highest  $SNR_k(n)$  value is selected by the aforementioned scheduling mechanism.

### 3.2. RIS model

We base our RIS model on the work conducted in Bjornson et al. (2020), which concludes that to beat decode-and-forward relaying scheme, RIS requires large reflecting elements and/or very high rates both in terms of minimizing the total transmit power and maximizing the energy efficiency.

We consider  $M$  number of RIS located at different positions to serve sensors as portrayed in Fig. 2 so that RIS can enable the reflection of the incident signal towards UAV, while each RIS is equipped with  $E$  discrete reflecting elements. The deterministic channels from sensor  $s$

**Table 1**  
Critical evaluation of this study compared to some of the existing literature.

Study	Advantages	Disadvantages	Gaps addressed in this study
Chen et al. (2019)	Efficient clustering and trajectory planning for large-scale networks.	High computational overhead and limited scalability.	Introduces DRL to optimize trajectories dynamically, addressing scalability and computational challenges.
Bayerlein et al. (2020)	High adaptability in dynamic environments using DDQN.	Limited generalizability.	Develops the DCFR model to overcome agent stalling, improving robustness in dynamic environments.
Cicek (2021)	Effective battery management and significant reduction in data loss.	Relies heavily on battery swap stations and lacks flexibility for real-time adjustments.	Combines DRL and RIS for real-time adaptability, minimizing reliance on static configurations.
Mondal et al. (2022)	Significant improvement in spectrum and energy efficiency using RIS phase optimization.	Non-convex problems limit scalability in larger networks.	Integrates RIS with DRL to enhance scalability and introduces a novel metric ( <i>collected data per unit energy</i> ).
Fan et al. (2023)	Effective reduction in Age of Information (AoI) using DRL.	Focuses on AoI, neglecting throughput maximization.	Maximizes throughput using DRL-enhanced UAV and RIS integration.
Almasoud (2023)	Anti-jamming resilience and robust static performance through ant colony optimization.	Static optimization lacks adaptability to dynamic conditions.	Employs DRL for dynamic UAV trajectory optimization, enabling real-time adaptability.
Wang et al. (2024)	Enhances signal quality and energy efficiency with RIS, mitigates jamming using PSO and convex approximation, focusing on anti-jamming.	High computational demands, reliance on RIS placement accuracy, and scalability issues in dense networks.	Introduces a novel metric ("data per unit energy") and improves energy efficiency and data collection in randomized IoT deployments using RIS with DRL.

through UAV  $u$  are expressed in order of  $[\mathbf{h}_{s_k r_k}, \mathbf{h}_{r_k u}] \in \mathbb{C}^N$ , where for example,  $\mathbf{h}_{s_k r_1}$  represents the channel from sensor  $k$  to the first RIS considering  $K$  number of RIS as each RIS is dedicated to one sensor and  $[\mathbf{h}_{(\cdot)_i}]_i$  denotes  $i$ th component of the concerned RIS. Noting that the size of each reflecting element is smaller than the wavelength, e.g.,  $\lambda/5 \times \lambda/5$  and  $\lambda/8 \times \lambda/8$  are reported in Bjornson et al. (2020), so that it can scatter the incoming signal using a near constant gain in all direction of interest (Bjornson et al., 2020) with adjustable amplitude and/or phase. For each RIS, signal model can be represented by a diagonal matrix  $\Theta_{r_1, \dots, r_K} = \alpha \text{diag}_{r_1, \dots, r_K}(e^{j\theta_1}, \dots, e^{j\theta_E})$  for a tunable phase-shift array of  $\theta^{r_1, \dots, r_K} = [\theta_1, \theta_2, \theta_3, \dots, \theta_E]_{\theta \in [0, \pi]}$ , respectively, where  $\alpha \in [0, 1]$  denotes the fixed amplitude reflection coefficient. Following the RIS models derived in Bjornson et al. (2020), the achievable rate at the UAV  $u$  given each sensor can be formulated as follows:

$$R_{r_k}(E) = \max_{\theta^{r_k, \forall k \in \{1, \dots, K\}}} \log_2 \left( 1 + \frac{p |h_{s_k u} + (\mathbf{h}_{s_k r_k}^T \Theta_{r_k} \mathbf{h}_{r_k u})|^2}{\sigma^2} \right),$$

$$= \log_2 \left( 1 + \underbrace{\frac{p (|h_{s_k u}| + \alpha \sum_{i=1}^E |[\mathbf{h}_{s_k r_k}]_i [\mathbf{h}_{r_k u}]_i|)^2}{\sigma^2}}_{SNR_{r_k}} \right), \quad (2)$$

where  $SNR_{r_k}$  represents the received SNR at the UAV from  $k$ th sensor via the  $k$ th RIS, depicted as  $r_k$ . For each task time slot  $t$ , RIS establishes  $\delta$  times communication to UAV via RIS model using the duration of communication time slot  $n$ . For a fair comparison to Bayerlein et al. (2020), we assume that our channel model incorporates shadow fading in addition to Bjornson et al. (2020). In Section 5.2, we provide more details on how RIS is adopted to our system.

### 3.3. Problem formulation

For one mission time slot  $t \in [0, T]$ , the achievable throughput is given by the sum of all achieved rates of the activated communication time slot  $n \in [0, N]$  over  $K$  ( $k \in [1, K]$ ) number of IoT devices and given by Bayerlein et al. (2020)

$$D(t) = \sum_{n=\delta t}^{\delta(t+1)-1} q_k(n) R_k(n), \quad (3)$$

where  $q_k(n) \in \{0, 1\}$  represents the activated communication link from which the UAV starts collecting data. We also introduce RIS to help

increase the data collection performance and throughput via RIS is formulated as follows,

$$\widehat{D}(t) = \sum_{n=\delta t}^{\delta(t+1)-1} q_k(n) R_{r_k}^E(n), \quad (4)$$

where,  $R_{r_k}^E$  depicts the achievable rate given  $E$  number of reflecting elements for  $k$ th RIS that assists communication at time slot  $n$ . The UAV's data collection mission aims to maximize the overall throughput while minimizing the flight duration, which is constrained by the maximum flight time, NFZs, obstacle avoidance and landing points.

State during mission time slot  $t$ ,  $s_t = (\mathbf{D}_t, \mathbf{p}_t, b_t, \mathbf{M}, \mathbf{U})$ , representing state-space with device data, UAV position, remaining flying time, environment map (NFZ and landing/takeoff positions) and IoT device positions, respectively. UAV has  $\mathcal{A} = \{\text{north, east, south, west, hover, land}\}$  six actions, which later extended to ten actions encompassing diagonal moves to increase directional capacity.

The optimization problem was formulated as a reward function within the framework of a Markov decision process (MDP) and was realized using DRL model. Readers are referred to Bayerlein et al. (2020) for further details as we follow a similar methodology to solve this optimization problem.

## 4. Solution approaches

### 4.1. Q-Learning

Q-learning is a model-independent methodology in the field of reinforcement learning (RL) that enables an agent to refine and optimize its strategy by interacting with its environment. At any given point in time  $t$ , the agent observes a state  $s_t$ , and performs an action  $a_t$  based on this observation. The environment then issues a reward  $r(s_t, a_t)$  depending on the effectiveness of the action that sets the agent to a new state  $s_{t+1}$ . The overall goal for the agent is to determine a policy  $\pi$ , i.e. a strategy that maximizes the accumulated rewards. This strategy  $\pi$  describes an optimal distribution of actions resulting from the prevailing state.

Q-learning attempts to iteratively refine the state-action-value function, commonly referred to as the Q-function quantifying the expected total return resulting from the execution of a particular action in a particular state and serves as a central component in the policy ( $\pi$ ) learning process and given by,

$$Q^\pi(s, a) = \mathbb{E}_\pi[R_t | s_t = s, a_t = a], \quad (5)$$

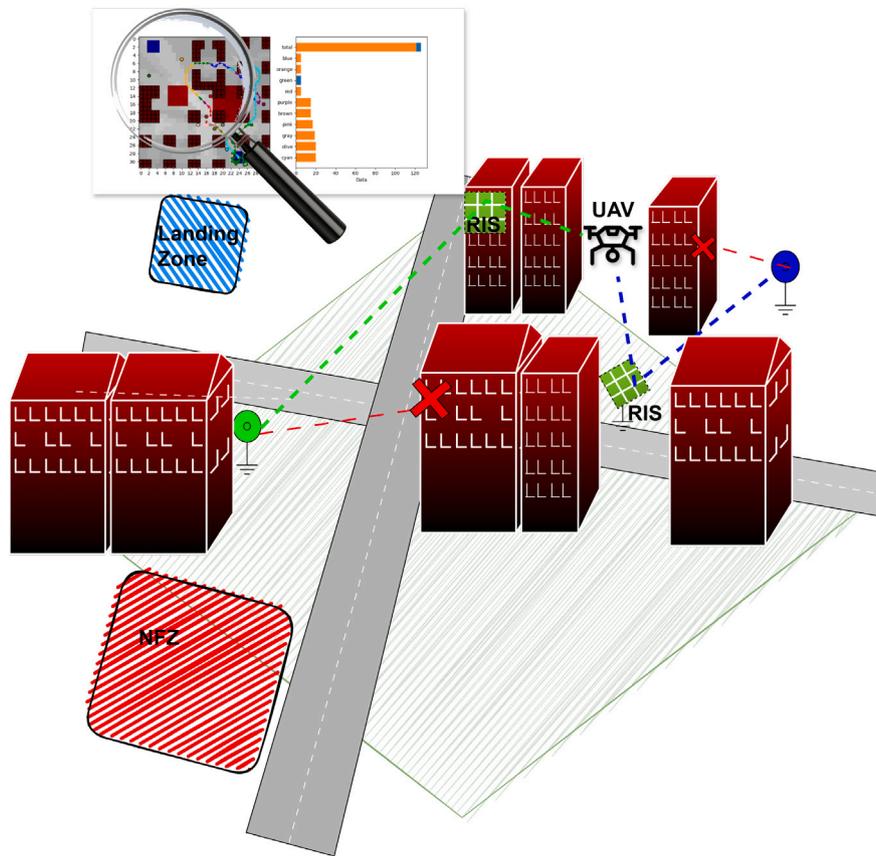


Fig. 2. RIS system model.

where represents the expected discounted cumulative return from the current state  $s_t$  until reaching a terminal state at time  $T$  given by,

$$R_t = \sum_{j=t}^T \gamma^{(j-t)} r(s_j, a_j), \quad (6)$$

where  $\gamma$  represents the discount factor balancing the weight of immediate and future rewards.

#### 4.2. DDQN and experience replay

The Double Deep Q-Learning (DDQN) algorithm is a further development of Deep Q-Networks (DQN) and uses deep neural networks to approximate the Q-function with improved convergence properties. DDQN shows superior performance due to its refined update mechanism, especially in environments characterized by extensive state and action spaces. In this method, a secondary target network is used for the Q-function updates. This mitigates the overestimation error that occurred in its predecessors and promotes more stable convergence throughout the training phase (Hasselt, 2010).

Experience replay is a technique that is often used to reduce temporal correlations within the training data set. In this approach, the agent's experiences are encapsulated as tuples of the form  $(s_t, a_t, R_t, s_{t+1})$  and archived in a replay buffer. Throughout the training process, the agent improves its learning process by extracting random minibatches from this memory. Such a method makes it easier for the agent to draw on previous experience and utilize it evenly, promoting a more even distribution of learning opportunities (Sutton and Barto, 2018).

#### 4.3. Centred global map

In the study, we used the centred global map technique advocated by Bayerlein et al. (2020) to process the map layers with a focus on

the location of the UAV. This method improves the ability to generalize across different scenario parameters and represents a pioneering example of the centring approach to global maps.

The map centring procedure is portrayed in Fig. 3. During this process, the maps were enlarged to a dimension of  $(2M - 1) \times (2M - 1)$ . The centre of the original map was then repositioned to match the UAV's location so that the agent can capture the entire map regardless of its position.

Using a centred map has advantages as the relational assignment of neurons within the "flatten" layer changes. Without centring, the neurons in this layer correspond to the features at fixed positions. With the map centred, however, these neurons correlate with features in relative positions, which increases the efficiency of the learning processes, especially when the agent has to initiate actions based on its relative position.

#### 4.4. DDQN architecture

We adopted the neural network model of Bayerlein et al. (2020) encompassing the preprocessing of the centred map and its subsequent use in the neural network, which is portrayed in Fig. 4. The centred map is passed through convolutional layers and then fed into fully connected layers. In these layers, the ReLU activation function is used and the outputs represent the Q-values corresponding to each action. During training, the softmax policy is used to ensure the balance between exploration and exploitation of the agent. This policy determines the agent's probability for each action based on the current state. During evaluation, the action with the highest Q-value is determined and this action is implemented (Bayerlein et al., 2020).

The agent is capable of reconstructing a defined number of past episodes essential for any DRL algorithm's training. Each episode consists of a tuple with state, action, reward, and the following state

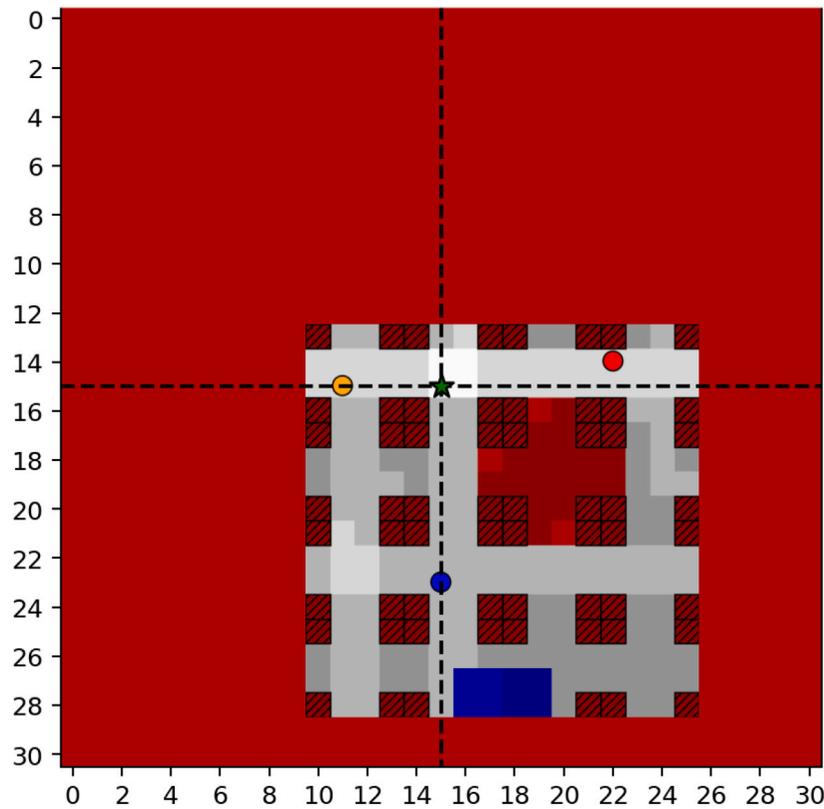


Fig. 3. Centred map with the UAV's location represented by a green star and the intersection of the dashed lines (Bayerlein et al., 2020).

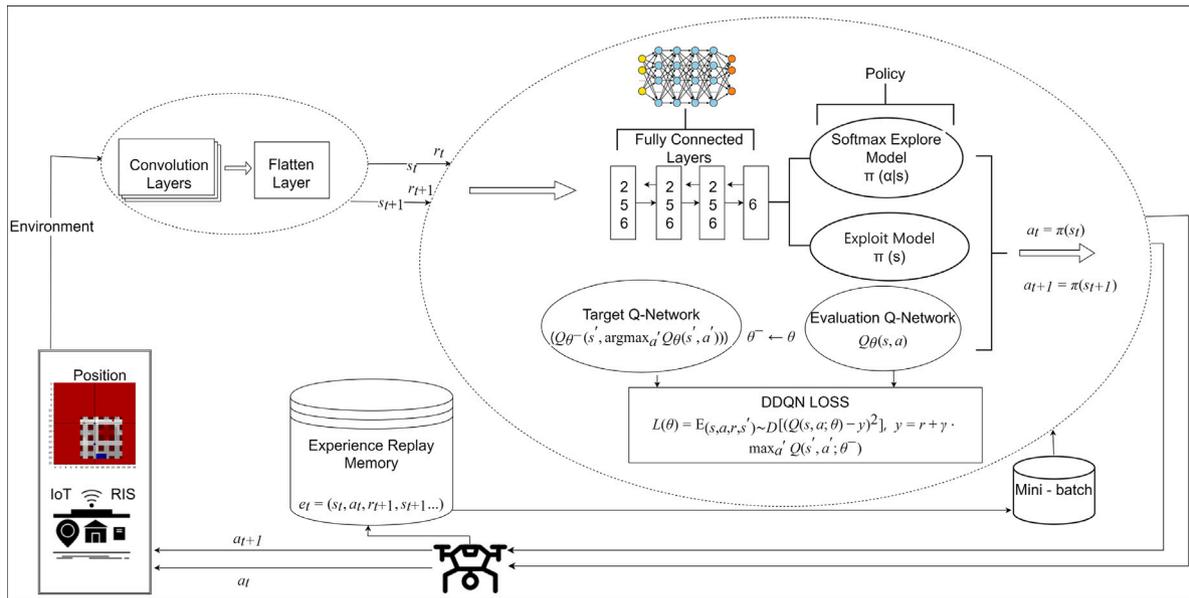


Fig. 4. Double DQN architecture with combined experience replay.

$(s_t, s_{t+1}, r_t, a_t)$ . These tuples are used to train the policy neural network using batch learning. As shown Fig. 4, the training process includes calculating the loss  $\mathcal{L}$  and updating the policy  $Q_\theta$  by minimizing  $\mathcal{L}$ , in accordance with the methodology described in Bayerlein et al. (2020).

#### 4.5. Model assumptions and constraints

The proposed Double DQN model operates under several assumptions and constraints to effectively address the problem of UAV trajectory optimization in RIS-assisted IoT networks. The environment is

modelled as a finite Markov decision process (MDP), assuming that the current state–action pair fully determines the next state and reward. The state representation incorporates UAV location, energy levels, communication quality, and RIS phase configurations, while the action space includes discrete movements and RIS reconfigurations, executed within fixed time slots. The reward function is designed to prioritize throughput maximization and energy efficiency, penalizing actions that result in excessive energy consumption or violations of operational constraints, such as no-fly zones. To enhance training stability, the DQN employs experience replay with a finite buffer, assuming that

random sampling breaks temporal correlations and improves convergence. However, the model is constrained by the UAV's limited energy budget, which restricts the number of actions and overall mission duration. Additionally, computational resources onboard the UAV impose limitations on the complexity of the neural network architecture and training processes. The  $\epsilon$ -greedy policy for exploration assumes a balance between exploration and exploitation to converge towards the optimal policy, while the long-term reward optimization is influenced by the discount factor, constraining the trade-off between immediate and future rewards. These assumptions and constraints form the foundation for the model's design and implementation, providing a framework that balances computational feasibility with performance optimization.

## 5. Navigation & Reconnaissance and RIS integration

In Phase-1 of this section, we explain the methodology on how the directional capacity is enhanced and the flexible reconnaissance is maintained, while the details of the integration methodology of RIS are provided in Phase-2.

### 5.1. Phase-1: Enhanced directional capacity and flexible reconnaissance (DCFR)

The system model described in Section 3.1 is used for Phase-1. Since the movement of the UAV in four directions, i.e., *east, west, north, south* does not provide enough directional capacity, the UAV is arranged in eight directions encompassing the movement action towards *east, west, north, south, south-east, south-west, north-east, north-west*. Noting that hovering and landing movements are default actions of the agent. In addition to improving the directional capacity, we have also developed Algorithm 1 to further improve the UAV's reconnaissance flexibility.

Our investigations in Bayerlein et al. (2020) reveal that in some scenarios (e.g. when the amount of data was low and the IoT device was far away) the UAV did not become active to collect data and did not capture this data. This situation was attempted to be solved with Algorithm 1 that was added to the simulation. If the UAV has enough energy while there is data to collect in the environment, the UAV will be activated to make further discoveries. After these changes, the simulation was retrained with 2 million iterations. The results were compared to the pre-trained model and the impact on the data collection rate was analysed. The effects of these changes were analysed with manually created scenarios, and the effects on the average data collection rate and energy efficiency in 1000 random scenarios were investigated.

Therefore, the solution approach provided in Phase-1 reflects both the enhanced directional capacity and the algorithm developed in Algorithm 1, which is referred to as DCFR model in rest of the paper.

**Algorithm 1** Enhancing UAV's reconnaissance flexibility.

---

```

1. while getRemainingData()>0:
2.   ...
3.   if getRemainingData() > 0 → data available at IoT
       else proceed to Step 6 for landing.
4.   if isLanding≠true, proceed to Step 7.
5.   if getUAVEnergy() >10 unit, proceed to Step 7,
       else isLanding=true and proceed to Step 9.
6.   → chose a random direction but landing position
       and proceed to Step 3.
7.   ...
8.   ...
9. end while

```

---

### 5.2. Phase-2: Integration of RIS

Without loss of generality, the integration of RIS is considered along with the energy efficiency it brings to the overall system model. MDP based reward problem considered a parameter called remaining flight time  $b_t$ .

For comparison purposes, regular data collection is assumed to be relayed over a DF system. For such system, we assumed a rate of 8.94 bit/s/Hz, as highlighted in Table 2, which is reproduced from Bjornson et al. (2020) for adopting the impact of RIS into our system model from energy efficiency perspective. For a given rate, as highlighted in Table 2, we can consider that energy efficiency of RIS is around 30% better compared to DF for each communication attempt of time slot  $n$ , which also performs better compared to single input single output system (SISO). Therefore, we can assume that 30% performance improvement against DF is a lower bound because beyond certain rate requirements, DF also performs better than SISO case as highlighted in Bjornson et al. (2020). In RIS comparison scenario, every communication at time slot  $n$  is assumed to be assisted by an RIS. Each sensor is served by a different RIS, where sensors and the UAV have a NLOS channel, while sensor-RIS-UAV channels maintain LOS. RIS model considers  $E_{opt}$  elements and its EE reflects the relevant optimized phase shifts in rate calculations and thus the  $SNR_{r_k}$  as formulated in Eq. (4). Given above assumptions, for a given rate  $R$  and bandwidth  $B$ , we follow the energy efficiency (EE) in Bjornson et al. (2020) and formulate both RIS and DF scenarios, respectively, as follows.

$$EE_{r_k} = B \cdot \frac{R}{P_{total}^{r_k}(E_{opt})}, \quad EE_{DF} = B \cdot \frac{R}{P_{total}^{DF}}, \quad (7)$$

where  $P_{total}^{r_k}(E_{opt})$  and  $P_{total}^{DF}$  represent the power consumption of RIS and DF, respectively, and readers are referred to Bjornson et al. (2020) for further details, as we focus our attention on the energy efficiency brought by the RIS as a saving coefficient (SC), which is derived as follows,

$$SC = \frac{EE_{r_k} - EE_{DF}}{EE_{r_k}}. \quad (8)$$

Therefore, using

$$\widehat{E_{UAV}}(n) = E_{UAV}(n) - (E_{UAV}(n) \cdot SC), \quad (9)$$

we can update the UAV energy dissipation per mission time slot  $t$  assuming that RIS assists all the communication in each time slot  $n$ , resulting in a slower consumption of  $b_t$  flight duration. This translates  $b_t - E_{UAV}(t)$  into  $b_t - \widehat{E_{UAV}}(t)$  leveraging RIS, where  $t$  is composed of  $\delta$  times of  $n$  time slots. Considering EE of each time slot  $n$ , overall EE  $\widehat{E_{UAV}}(t)$  is averaged over  $\delta$  attempts, we can then formulate the average EE brought by RIS  $r_k$  compared to DF for each mission time slot  $t$ , as follows.

$$\widehat{E_{UAV}}(t) = \frac{1}{\delta} \sum_{n=\delta t}^{\delta(t+1)-1} \widehat{E_{UAV}}(n), \forall t, \quad (10)$$

Note that  $\widehat{E_{UAV}}(t)$  designed in a way so that it reflects the average EE brought by RIS compared to DF as a *lower bound* and this value is fed into the MDP-based reward maximization problem, provoking less consumption of remaining flight time  $b_t$  compared to the case without RIS.

## 6. Experimental evaluation

In this section, we first provide the evaluations for the proposed ‘‘DCFR Model’’ contrasted against the model developed in Bayerlein et al. (2020). Then, we investigate the evaluation results after RIS integration as opposed to the model in Bayerlein et al. (2020) and to the proposed DCFR model. To clarify, ‘‘DCFR Model’’ is an extension to the model proposed in Bayerlein et al. (2020), which we call as ‘‘Bayerlein

**Table 2**

Comparing the energy efficiency of RIS to that of decode and forward (DF) for a specified rate, which is reproduced from Bjornson et al. (2020).

Achievable rate [bit/s/Hz]	RIS energy efficiency	Decode-and-forward (DF) energy efficiency	RIS energy efficiency compared to DF [%]
8.8	113.26	89.01	21.42
8.9	112.11	81.02	27.73
<b>8.94</b>	<b>111.65</b>	<b>77.95</b>	<b>≈30</b>
9	110.95	73.49	33.76
9.1	109.78	66.44	39.48
9.2	108.61	59.89	44.86
9.3	107.44	53.84	49.89
9.4	106.26	48.28	54.57

**Table 3**

Symbols and descriptions used for the grid map (Bayerlein et al., 2020).

Symbols	Description
	Take off / landing zone
	No-flying zone (NFZ)
	Physical obstructions
	IoT devices (sensors)
	Cumulative shadowing
	Taking off (left) and landing (right) positions
	Directed UAV movement at the time of data collection, e.g. from green IoT device
	Hovering and communicating with green device
	UAV to move without collecting data

Model” in rest of the paper, and the RIS-integrated model, namely “RIS Model”, is also a cumulative extension to the DCFR Model.

It is important to understand the meaning of the symbols used at the grid map considered in our analyses that is outlined in Table 3. These symbols mainly show the constraints of the map, such as NFZ and physical obstructions, as well as represent the actions of the agent (UAV).

The dataset and simulation environment used in this study are designed to evaluate the proposed methodology comprehensively. The UAV initiates each new mission within a map grid composed of  $16 \times 16$  cells, each grid cell having a size of  $10 \times 10$  m. IoT devices are randomly positioned within this grid, with data quantities assigned randomly between 5 and 20 units per device. UAV energy levels are initialized randomly between 50 and 150 units for each trial to simulate diverse operating conditions. The UAV starts with  $b_t$  remaining flight time, which is decremented by one after every action the UAV takes, given the action space in the regular scenario. This is generalized with the RIS saving coefficient, which is set to 30% by default compared to the DF scenario, as highlighted in Table 2, and can be adjusted to any value of interest. The predetermined rate  $R$  is leveraged to optimize the number of reflecting elements  $E$ , further enhancing energy efficiency.

The UAV flies at a constant altitude sufficient to establish LOS and NLOS channels between sensors and the UAV while considering obstacles such as buildings, NFZs, and RIS positions. The communication environment includes urban scenarios with shadowing effects modelled using ray tracing, ensuring a realistic propagation environment. The UAV agent has no prior knowledge of the channel and learns it dynamically following the methodology and propagation parameters outlined in Bayerlein et al. (2020) and Esrafilian et al. (2019). Monte Carlo simulations were conducted over 1000 trials to ensure robust statistical evaluation.

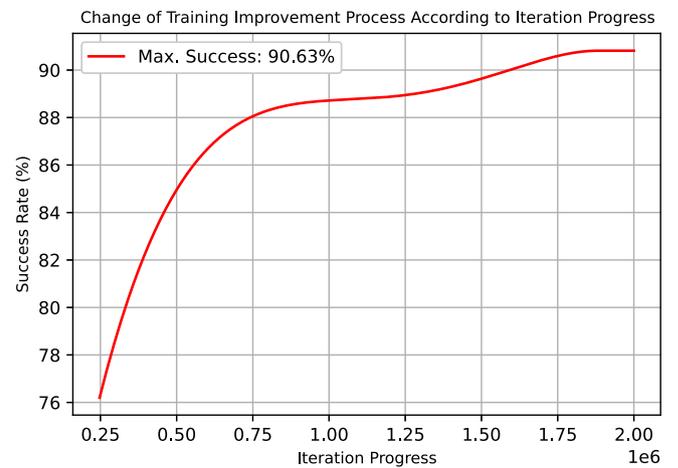


Fig. 5. Presenting the success rate of the UAV after 2 million training iterations (generated on the basis of the peak values achieved during training).

We also use three main metrics to evaluate the performance of the agent in different scenarios comparing aforementioned models.

- *Collection ratio* reveals the data collection performance of the model.
- *Successful landing* helps understand if a UAV has safely landed given a mission is completed.
- We introduced a new metric, namely *collected data per unit energy*, revealing if the data is collected in an energy efficient manner. This metric is very helpful along with the collection ratio. If the collection ratio is high, then one can readily understand whether this data collection process was conducted by consuming less energy.

### 6.1. Phase 1: Evaluation of DCFR model

Bayerlein Model proposed in Bayerlein et al. (2020) underwent retraining incorporating the modifications outlined in Section 5.1, and the training outcomes were captured as depicted in Fig. 5. It is evident from the figure that the UAV achieved its peak success rate at 90.63%. Beyond this point, it exhibited no further enhancements and maintained a plateau in performance.

#### 6.1.1. Evaluating the directional capacity of UAV

In this scenario, 15 units of data, each corresponding to one of the 10 IoT devices in the system, were loaded and their positions were manually adjusted. The UAV’s energy capacity is set at 150 units. The UAV’s directional capacity and data collection performance while moving in four directions (East, West, North, South) are illustrated in Fig. 6(a). Furthermore, using the same IoT device positions, units of data, and energy capacity, the UAV’s directional capacity and data collection performance in eight directions (East, West, North, South, Southeast, Southwest, Northeast, Northwest) are depicted in Fig. 6(b).

As the results in Fig. 6 show, the UAV, which could move in eight directions, completed the data collection task in less time. Despite collecting all the available data, it did not exhaust its energy reserves (black arrows indicate that the drone moves when no further data collection is required). In addition, as shown in Fig. 6(a), after collecting all data except the data from the red IoT device, the UAV attempted one last data collection before landing, but was unsuccessful. In Fig. 6(b), however, data was successfully collected from all devices owing to the enhanced directional capacity. Even though Fig. 6 portrays the impact of the UAV’s directional capacity, the improved data collection performance is an outcome of the proposed DCFR model encompassing the enhanced directional capacity and the proposed Algorithm 1.

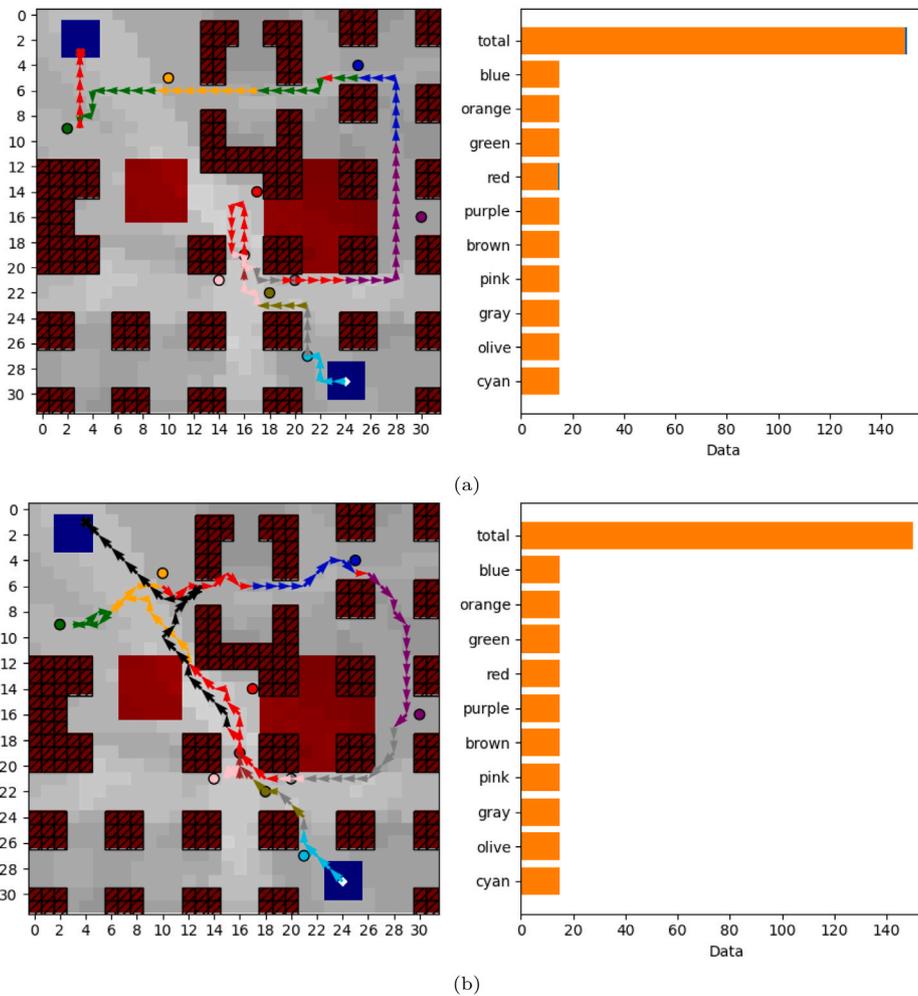


Fig. 6. (a) The trajectory of the UAV, capable of movement in four directions, alongside the data collection performance, (b) The trajectory of the UAV, maneuverable in eight directions, accompanied by the corresponding data collection performance.

More explicitly, the eight-directional movement allowed the UAV to minimize redundant paths and adjust its trajectory to cover more IoT devices within the same energy constraints. As shown in Fig. 6, the UAV operating in eight directions not only completed the data collection task more efficiently but also retained sufficient energy reserves, demonstrating superior trajectory optimization. These results highlight the critical role of increased manoeuvrability in improving UAV operations, particularly in dynamic IoT environments.

### 6.1.2. Manually configured scenarios

In scenarios where sufficient energy is available but certain sensors provide less data or are too far away, the UAV has the option of not collecting data from these sensors and flying to its landing position. In some cases, the UAV was not able to detect data from these devices and therefore could not initiate take-off (Bayerlein et al., 2020). In such a scenario, there is a risk of losing crucial data from sensors with relatively small but important information. To solve this problem, we have proposed an algorithm described in Algorithm 1 as part of the DCFR approach. By refining the algorithm to recognize data availability and enabling further exploration when energy allows, our system achieves higher efficiency in data collection compared to the approach described in Bayerlein et al. (2020).

**Scenario-1: Reconnaissance capability.** In Fig. 7, 5 data units are loaded into the blue device and 15 data units into the orange device. The amount of energy of the UAV is 150 units. The position of the IoT devices in Fig. 7(b) was manually placed according to the positions

used in Fig. 7(a) to allow a fair comparison of the data collection performance between Bayerlein model proposed in Bayerlein et al. (2020) and our DCFR model. In Fig. 7, the blue colour indicates the situation in which no data is collected and the orange colour indicates the situation in which it is collected.

In Fig. 7(a), the UAV remained still because it could not detect the data on the IoT devices. Contrarily, in Fig. 7(b), the UAV recognizes the remaining data on the red IoT device and lands safely after collecting all the data. The data collection success rate in this case becomes 100%. It is evident that the DCFR model using Algorithm 1 enables the UAV to make more discoveries if there is data on the IoT devices.

The inability of the Bayerlein model to detect data on IoT devices, as seen in Fig. 7(a), is due to its static exploration strategy, which fails to adapt to sparse or low-data scenarios. In contrast, the DCFR model, equipped with Algorithm 1, dynamically adjusts its exploration and prioritizes underutilized data sources, as shown in Fig. 7(b). This ensures no data is left uncollected, achieving a 100% success rate. This result is particularly significant for real-world applications where critical but sparse data must be prioritized, such as in disaster response or emergency monitoring.

**Scenario-2: Data collection from devices with low amount of data.** In this scenario, data ranging from 5 to 20 units are manually allotted to 10 IoT devices, while the UAV’s energy level is measured at 150 units. The positioning of the IoT devices remains the same in both cases, as portrayed in Figs. 8(a) and 8(b). The distribution of data on these devices is as follows: blue (20 units), orange (10 units), green (18

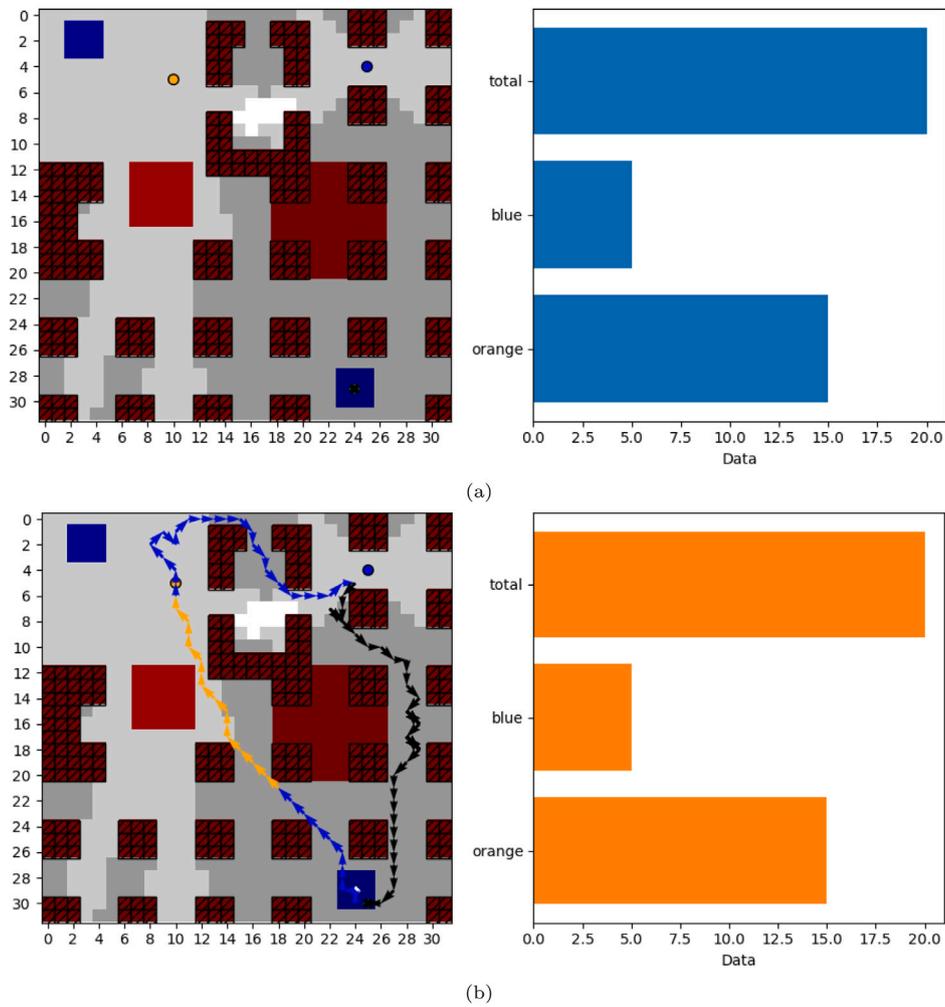


Fig. 7. (a) The case in which the UAV is unable to detect the data on devices, (b) The trajectory of the UAV with the DCFR model encompassing Algorithm 1.

units), red (5 units), purple (19 units), brown (12 units), pink (15 units), grey (14 units), olive green (13 units) and cyan (12 units).

In Fig. 8(a), despite five attempts using Bayerlein model to establish a communication with the red device (having the least amount of data), the UAV was unable to receive any data and then landed at the opposite landing point due to insufficient energy. In this case, the overall success rate of data collection stands at 96.38%. However, with the DCFR model retrained using the algorithm described in Algorithm 1, the data collection success rate rose to 100%, as shown in Fig. 8(b). Remarkably, the UAV still had enough energy to return safely to its starting point after completing the data collection from IoT devices. This outcome underscores the importance of integrating adaptive learning mechanisms for efficient energy utilization in data collection missions.

**Scenario-3: Data collection from the no-fly zone.** In this particular scenario, a sum of 150 units of data, ranging from 5 to 25 units each, was distributed among 10 IoT devices. One of these devices was located within a no-fly zone for UAVs, while another was situated at a considerable distance from the remaining devices. The UAV starts with an initial energy level of 150 units.

In Fig. 9(a), due to the presence of the brown device within a no-fly zone, the UAV is unable to approach closely, resulting in minimal data collection from this device. Furthermore, as the purple device is far away from the remaining devices, the UAV cannot fully detect it, so no data is collected from it. In contrast, in Fig. 9(b), the brown and purple devices contain significantly more data than the blue device, so the UAV prioritizes connecting to these devices after collecting some

data from the blue device. Consequently, it successfully collects all data from the purple device and almost all from the brown device. However, it only receives a small amount of data from the blue device, which contains 5 units of data. Upon examination of the figures, the total data collected amounts to 120 units in Fig. 9(a) and 145 units in Fig. 9(b). In both cases, the UAV lands safely within the intended landing zone. It is noteworthy that the DCFR model, considering the improvements of Algorithm 1 and representing Fig. 9(b), increases the amount of collected data by 16.7% compared to Fig. 9(a) representing Bayerlein model.

More explicitly, in environments with regulatory constraints, such as no-fly zones, the DCFR model demonstrated a 16.7% improvement in data collection (Fig. 9(b)) compared to the Bayerlein model (Fig. 9(a)). This is attributed to the DCFR model's ability to prioritize devices with high data volumes while respecting operational constraints. These results have practical implications for deploying UAVs in urban and regulated environments, emphasizing the need for intelligent trajectory planning.

### 6.1.3. Overall system performance with arbitrary parameter values

Parameter values for the overall system evaluation are provided in Table 4 involving 3 to 10 IoT devices randomly assigned with data quantities ranging from 5 to 20 units. The positioning of these devices on the map is also randomized in each trial. Additionally, the UAV's energy level is randomly set between 50 and 150 units for each trial. The evaluation results are averaged over 1000 trials.

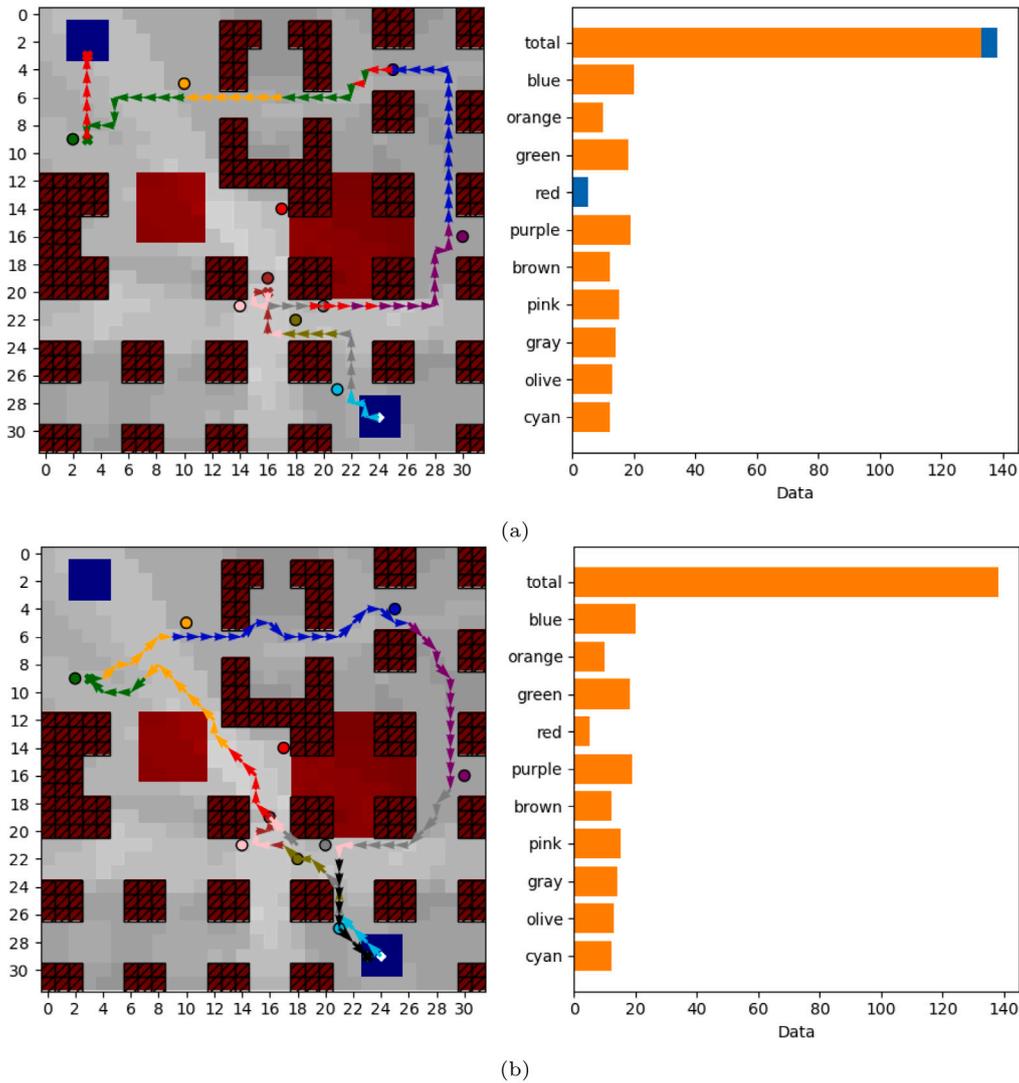


Fig. 8. (a) Performance of the model in Bayerlein et al. (2020), (b) Performance of the proposed model retrained using Algorithm 1.

Table 4

Parameter values for the overall system evaluation.

Parameters	Values
Number of IoT device	Randomized between 3-10
Device positioning	Random in each trial
UAV energy level	Randomized between 50–150
Amount of IoT data	Randomized between 5–20
Number of trials	1000
Map grid dimensions	16 × 16 cells, each 10 m × 10 m
UAV altitude	10 m
Communication time slots	4 per mission time slot
Path Loss Exponent (LoS)/(NLoS)	2.27/3.64 (Bayerlein et al., 2020)
Shadowing Variance (LoS)/(NLoS)	2/5 (Bayerlein et al., 2020)
Transmission power	Normalized (cell-edge SNR of −15 dB at grid corner)
UAV velocity	Constant, remains within the grid
Channel model	LoS/NLoS with log-distance path loss and shadow fading
RIS reflecting elements	Optimized for energy efficiency (Bjornson et al., 2020)
RIS saving coefficient	30% improvement over DF scenarios; adjustable based on rates (Bjornson et al., 2020)
UAV movement directions	4 or 8 directions

Table 5

Overall system performance evaluation of the DCFR model compared to Bayerlein model.

Evaluation metrics	Bayerlein model (Bayerlein et al., 2020)	DCFR model
Successful landing <sup>a</sup>	0.983	0.986
Collection ratio <sup>a</sup>	0.831	0.899
Available data	80 242	81 890
Collected data	66 055	72 534
Dissipated energy	83 072	85 331
Collected data per unit energy	0.795	0.850

<sup>a</sup> Values are normalized within the range of 0 to 1.

As illustrated in Table 5 and Fig. 10, a marginal improvement in the already high “Successful Landing” performance is observed. The main focus of the evaluation is therefore on the improvement of the “Data Collection Rate”, which shows an increase of 8.18%. In addition, the success of data collection per unit of energy has increased by 6.92%. These results clearly demonstrate the effectiveness of the proposed algorithm used to retrain the model to improve data collection performance and increase energy efficiency while maintaining similar “successful landing” performance.

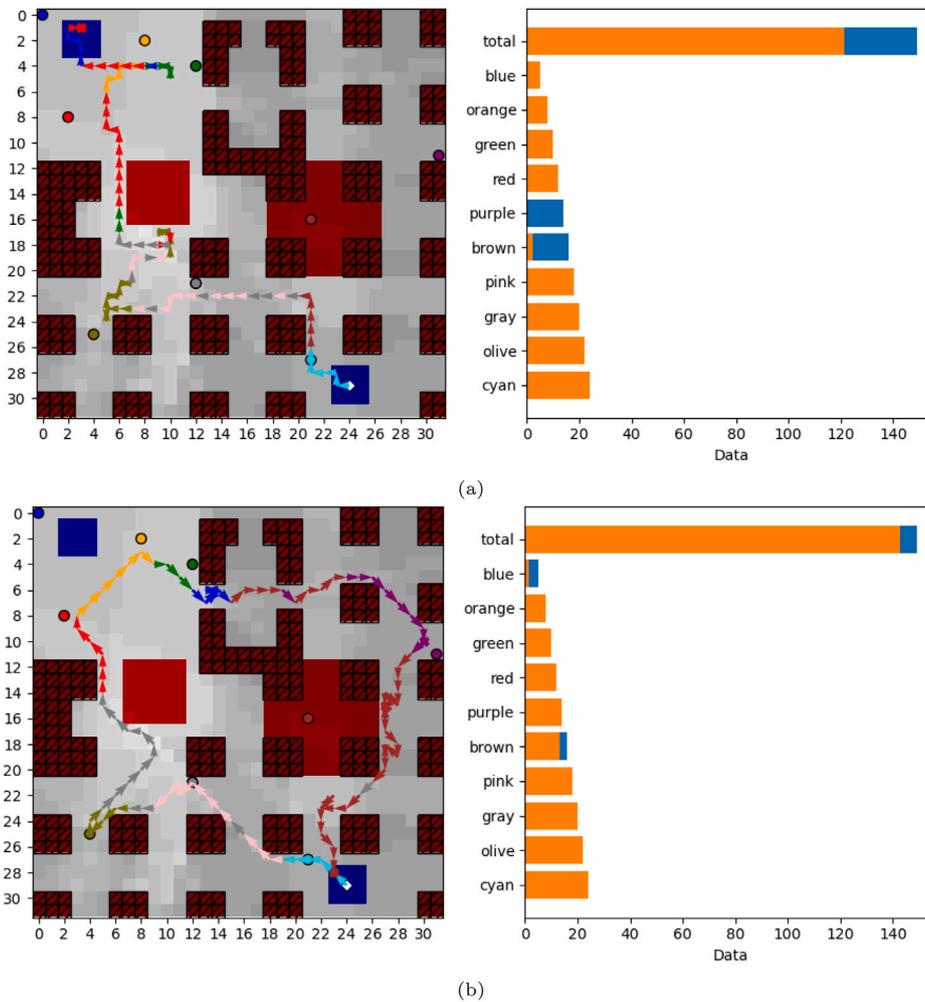


Fig. 9. (a) No-fly zone scenario with the model from Bayerlein et al. (2020), (b) No-fly zone scenario with the proposed model considering Algorithm 1.

The increased “Data Collection Ratio” indicates the effectiveness of the DCFR model in ensuring higher coverage and retrieval of IoT device data. This improvement highlights the role of Algorithm 1 in optimizing UAV trajectories and energy usage. The higher “Collected Data per Unit Energy” reflects the enhanced energy efficiency achieved through the DCFR model, underscoring its potential for prolonged and efficient UAV operations in IoT networks.

These results emphasize the advantages of integrating the DCFR model over existing approaches, demonstrating its ability to achieve superior performance in data collection and energy management while maintaining a consistently high rate of successful landings. The findings validate the robustness of the proposed model and its practical applicability to dynamic and constrained UAV-assisted IoT environments.

### 6.2. Phase-2: Evaluation of RIS model

As a cumulative extension to the DCFR model, RIS is incorporated using the methodology outlined in Section 3.2. This addition enabled the UAV to conserve more energy, consequently leading to an increase in the overall system performance. Within the scope of this study, after integrating RIS into the system to bolster data collection and energy efficiency, the simulation is averaged over 1000 random trials. The results obtained are presented in Table 6 and Fig. 11. A thorough examination of the results reveals that with the integration of RIS, the UAV’s data collection ratio increases by 10.59% and 2.22% compared to the Bayerlein model and DCFR model, respectively. This improvement underscores the effectiveness of RIS in

optimizing UAV trajectories and signal coverage, allowing for higher data retrieval rates. Moreover, the amount of data collected per unit of energy sees a substantial increase of 22.64% and 14.7% compared to the Bayerlein model and DCFR model, respectively. These results highlight the role of RIS in reducing energy dissipation during UAV operations, enabling extended mission durations and enhanced sustainability. Finally, RIS Model achieves a 0.995 successful landing rate, representing an improvement of 1.22% over the Bayerlein model and 0.91% over the DCFR model. This demonstrates the reliability of the RIS model in ensuring safe mission completion, even under variable energy constraints.

The RIS model achieves a 0.995 successful landing rate, representing an improvement of 1.22% over the Bayerlein model and 0.91% over the DCFR model. This demonstrates the reliability of the RIS model in ensuring safe mission completion, even under variable energy constraints.

## 7. Conclusions and open challenges

This study explores the impact of integrating RIS technology to augment data collection process and energy efficiency of UAVs. The investigation unfolds in two phases. Initially, we enhance data collection capacity by refining UAV directional capacity and reconnaissance capability, resulting in an 8.18% increase in data collection rate and a 6.92% rise in data collected per unit dissipated energy. Subsequently, RIS integration is conducted, elevating UAV data collection capacity

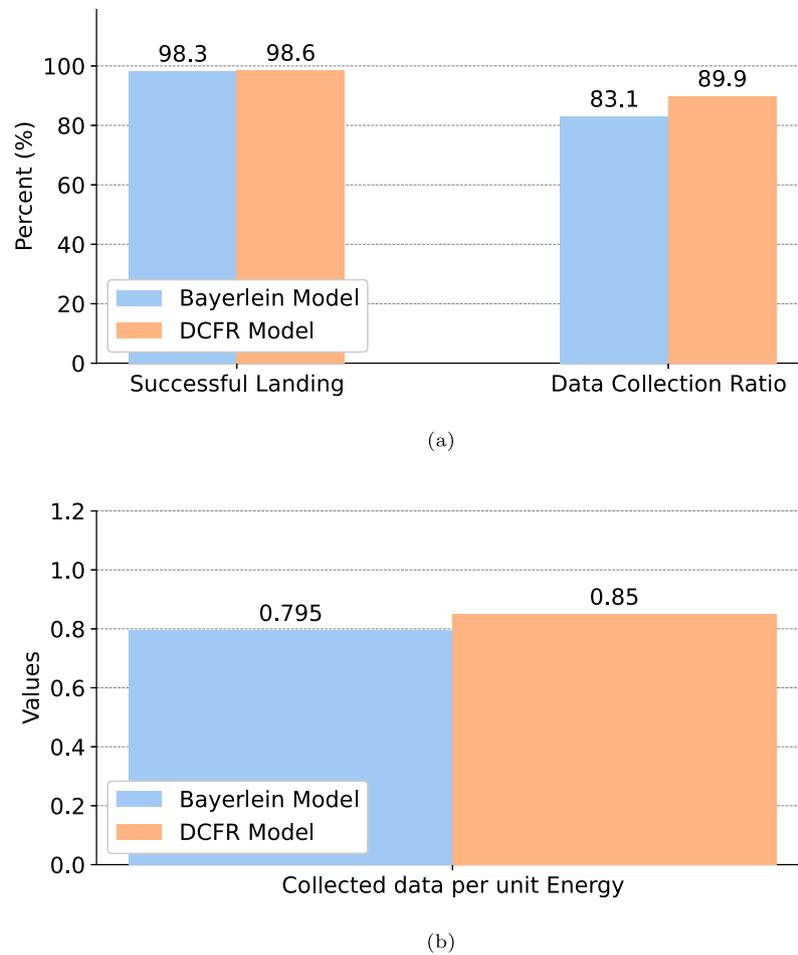


Fig. 10. Evaluating the overall performance of the Bayerlein model and DCFR model for; (a) Successful landing and data collection ratio, (b) Collected data per unit Energy.

Table 6

Evaluation of overall system performance following the integration of RIS.

Evaluation metrics	Bayerlein model (Bayerlein et al., 2020)	DCFR model	RIS model
Successful landing <sup>a</sup>	0.983	0.986	0.995
Collection ratio <sup>a</sup>	0.831	0.899	0.919
Available data	80 242	81 890	83 167
Collected data	66 055	7253 4	75 682
Dissipated energy	83 072	85 331	77 548
Collected data per unit energy	0.795	0.850	0.975

<sup>a</sup> Values are normalized within the range of 0 to 1.

by up to 10.59% and data collected per unit dissipated energy by up to 22.64%.

These advancements underscore RIS's potential in refining the efficiency of UAV-assisted data collection. Additionally, leveraging robust learning algorithms like DDQN in UAV data collection processes could offer deeper insights into the UAV's surroundings and boost data collection capacity, especially in dynamic environments. Such findings pave the way for more sustainable and enduring UAV mission execution.

Future research could delve into assessing RIS performance across diverse scenarios and conducting in-depth analyses of its effects on other heterogeneous networks.

#### CRedit authorship contribution statement

**Idris Ertas:** Writing – original draft, Software. **Halil Yetgin:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

We extend our gratitude to the authors of Bayerlein et al. (2020) and Bjornson et al. (2020) for generously sharing their structured and reproducible simulation codes.

#### Data availability

No data was used for the research described in the article.

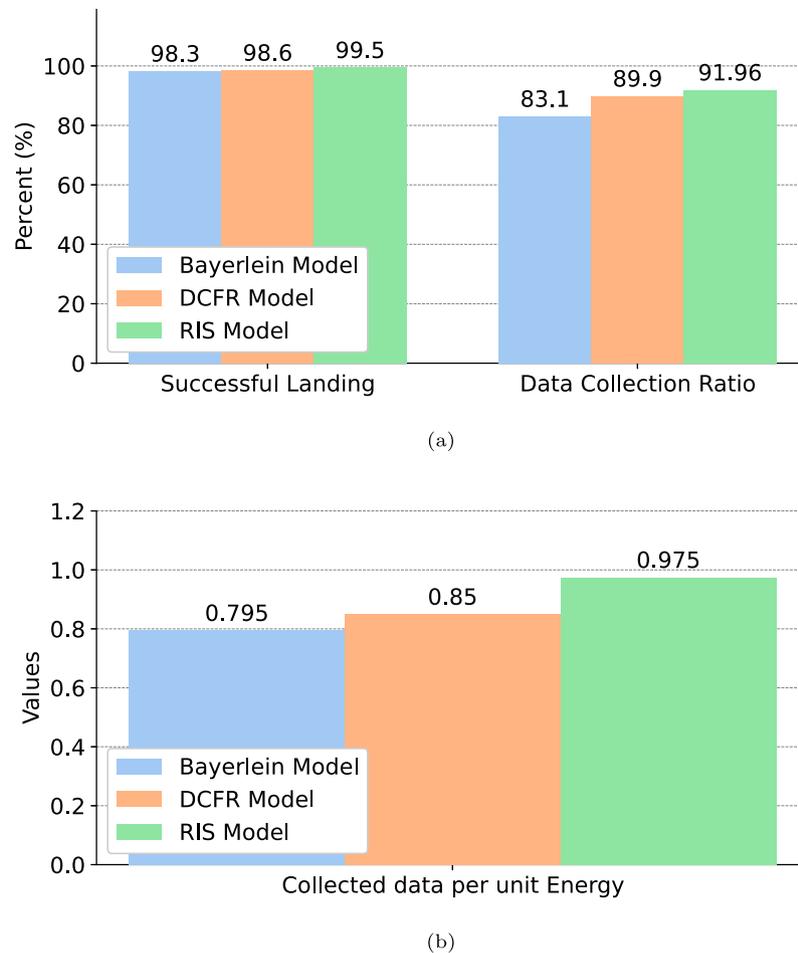


Fig. 11. Evaluating overall performance of the Bayerlein, DCFR and RIS models regarding; (a) Successful landing and data collection ratio, (b) Collected data per unit Energy.

## References

- Al-Hilo, A., Samir, M., Elhattab, M., Assi, C., Sharafeddine, S., 2023. RIS-assisted UAV for timely data collection in IoT networks. *IEEE Syst. J.* 17 (1), 431–442. <http://dx.doi.org/10.1109/JSYST.2022.3215279>.
- Al-Jarrah, M., Alsusa, E., Al-Dweik, A., So, D.K.C., 2021. Capacity analysis of IRS-based UAV communications with imperfect phase compensation. *IEEE Wirel. Commun. Lett.* 10, 1479–1483. <http://dx.doi.org/10.1109/LWC.2021.3071059>.
- Almasoud, A.M., 2023. Robust anti-jamming technique for UAV data collection in IoT using landing platforms and RIS. *IEEE Access* 11, 70635–70651. <http://dx.doi.org/10.1109/ACCESS.2023.3294596>.
- Atzori, L., Iera, A., Morabito, G., 2010. The internet of things: A survey. *Comput. Netw.* 54, 2787–2805. <http://dx.doi.org/10.1016/j.comnet.2010.05.010>.
- Bayerlein, H., Theile, M., Caccamo, M., Gesbert, D., 2020. UAV path planning for wireless data harvesting: A deep reinforcement learning approach. pp. 1–6.
- Benmad, I., Driouch, E., Kardouchi, M., 2022. Trajectory planning for data collection in multi-UAV assisted WSNs. In: 2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring). IEEE, pp. 1–6.
- Bjornson, E., Ozdogan, O., Larsson, E.G., 2020. Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying? *IEEE Wirel. Commun. Lett.* 9 (2), 244–248. <http://dx.doi.org/10.1109/LWC.2019.2950624>.
- Cao, C., Lian, Z., Wang, Y., Su, Y., Jin, B., 2021. A non-stationary geometry-based channel model for IRS-assisted UAV-MIMO channels. *IEEE Commun. Lett.* 25, 3760–3764. <http://dx.doi.org/10.1109/LCOMM.2021.3116192>.
- Chen, G., Wu, Q., He, C., Chen, W., Tang, J., Jin, S., 2023. Active IRS aided multiple access for energy-constrained IoT systems. *IEEE Trans. Wirel. Commun.* 22 (3), 1677–1694. <http://dx.doi.org/10.1109/TWC.2022.3206332>.
- Chen, J., Yan, F., Mao, S., Shen, F., Xia, W., Wu, Y., Shen, L., 2019. Efficient data collection in large-scale UAV-aided wireless sensor networks. In: 2019 11th International Conference on Wireless Communications and Signal Processing. WCSP, IEEE, pp. 1–5.
- Chu, Z., Xiao, P., Mi, D., Hao, W., Chen, Q., Xiao, Y., 2023. IRS-assisted wireless powered IoT network with multiple resource blocks. *IEEE Trans. Commun.* 71 (4), 2335–2350. <http://dx.doi.org/10.1109/TCOMM.2023.3242365>.
- Cicek, C.T., 2021. A reinforcement learning algorithm for data collection in UAV-aided IoT networks with uncertain time windows. In: 2021 IEEE International Conference on Communications Workshops. (ICC Workshops, IEEE, pp. 1–6.
- Esrifiliani, O., Gangula, R., Gesbert, D., 2019. Learning to communicate in UAV-aided wireless networks: Map-based approaches. *IEEE Internet Things J.* 6 (2), 1791–1802. <http://dx.doi.org/10.1109/JIOT.2018.2879682>.
- Fan, X., Liu, M., Chen, Y., Sun, S., Li, Z., Guo, X., 2023. RIS-assisted UAV for fresh data collection in 3D urban environments: A deep reinforcement learning approach. *IEEE Trans. Veh. Technol.* 72 (1), 632–647. <http://dx.doi.org/10.1109/TVT.2022.3203008>.
- Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M., 2013. Internet of things (IoT): A vision, architectural elements, and future directions. *Future Gener. Comput. Syst.* 29, 1645–1660. <http://dx.doi.org/10.1016/j.future.2013.01.010>.
- Haider, S.A., Zikria, Y.B., Garg, S., Ahmad, S., Hassan, M.M., AlQahtani, S.A., 2022. AI-based energy-efficient UAV-assisted IoT data collection with integrated trajectory and resource optimization. *IEEE Wirel. Commun.* 29 (6), 30–36. <http://dx.doi.org/10.1109/MWC.001.2200105>.
- Hasselt, H., 2010. Double Q-learning. *Adv. Neural Inf. Process. Syst.* 23.
- Huang, C., Zappone, A., Alexandropoulos, G.C., Debbah, M., Yuen, C., 2019. Reconfigurable intelligent surfaces for energy efficiency in wireless communication. *IEEE Trans. Wirel. Commun.* 18, 4157–4170. <http://dx.doi.org/10.1109/TWC.2019.2922609>.
- Li, X., Yin, B., Yan, J., Zhang, X., Wei, R., 2022. Joint power control and UAV trajectory design for information freshness via deep reinforcement learning. In: 2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring). IEEE, pp. 1–5.
- Lian, Z., Su, Y., Wang, Y., Ji, P., Jin, B., Zhang, Z., Xie, Z., 2023. A novel geometry-based 3-D wideband channel model and capacity analysis for IRS-assisted UAV communication systems. *IEEE Trans. Wirel. Commun.* 22, 5502–5517. <http://dx.doi.org/10.1109/TWC.2023.3234555>.
- Liu, L., Wang, A., Sun, G., Li, J., Pan, H., Quek, T.Q.S., 2024. Multi-objective optimization for data collection in UAV-assisted agricultural IoT. *IEEE Trans. Veh. Technol.* 1–17. <http://dx.doi.org/10.1109/TVT.2024.3514664>.
- Mei, H., Yang, K., Liu, Q., Wang, K., 2022. 3D-trajectory and phase-shift design for RIS-assisted UAV systems using deep reinforcement learning. *IEEE Trans. Veh. Technol.* 71, 3020–3029. <http://dx.doi.org/10.1109/TVT.2022.3143839>.

- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M., 2013. Playing atari with deep reinforcement learning. URL <http://arxiv.org/abs/1312.5602>.
- Mondal, A., Junaedi, A.M.A., Singh, K., Biswas, S., 2022. Spectrum and energy-efficiency maximization in RIS-aided IoT networks. *IEEE Access* 10, 103538–103551. <http://dx.doi.org/10.1109/ACCESS.2022.3209823>.
- Mondal, A., Mishra, D., Prasad, G., Johansson, H.A., 2024. Joint trajectory, user-association, and power control for green UAV-assisted data collection using deep reinforcement learning. *IEEE Trans. Intell. Veh.* 1–12. <http://dx.doi.org/10.1109/TIV.2024.3416067>.
- Mozaffari, M., Saad, W., Bennis, M., Nam, Y.-H., Debbah, M., 2019. A tutorial on UAVs for wireless networks: Applications, challenges, and open problems. *IEEE Commun. Surv. Tutor.* 21, 2334–2360. <http://dx.doi.org/10.1109/COMST.2019.2902862>.
- Park, Y.M., Tun, Y.K., Han, Z., Hong, C.S., 2022. Trajectory optimization and phase-shift design in IRS-assisted UAV network for smart railway. *IEEE Trans. Veh. Technol.* 71, 11317–11321. <http://dx.doi.org/10.1109/TVT.2022.3189024>.
- Qi, W., Yang, C., Song, Q., Guan, Y., Guo, L., Jamalipour, A., 2024. Minimizing age of information for hybrid UAV-RIS-assisted vehicular networks. *IEEE Internet Things J.* <http://dx.doi.org/10.1109/JIOT.2024.3359666>, 1–1.
- Sutton, R.S., Barto, A.G., 2018. "Reinforcement Learning: an Introduction". MIT Press, URL <https://t.ly/vjRmb>.
- Wang, P., Liu, K., Ma, Y., Gao, Q., 2024. Aol and energy-aware data collection for IRS-assisted UAV-IoT networks under jamming. *IEEE Internet Things J.* <http://dx.doi.org/10.1109/JIOT.2024.3519668>, 1–1.
- Wang, F., Zhang, X., 2023. Active-IRS-Enabled Energy-Efficiency Optimizations for UAV-Based 6G Mobile Wireless Networks. *IEEE*, pp. 1–6. <http://dx.doi.org/10.1109/CISS56502.2023.10089767>.
- Yao, Y., Lv, K., Ma, N., Yue, X., Qin, X., Yun, X., 2022. Energy efficient air-to-ground communication networks with reconfigurable intelligent surface. *J. Commun. Netw.* 24 (5), 555–565. <http://dx.doi.org/10.23919/JCN.2022.000025>.
- Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M., 2014. Internet of things for smart cities. *IEEE Internet Things J.* 1, 22–32. <http://dx.doi.org/10.1109/JIOT.2014.2306328>.
- Zhou, H., Erol-Kantarci, M., Liu, Y., Poor, H.V., 2023. A survey on model-based, heuristic, and machine learning optimization approaches in RIS-aided wireless networks. *IEEE Commun. Surv. Tutor.* early-access. <http://dx.doi.org/10.1109/COMST.2023.3340099>.