

Article

Joint Optimization on Trajectory, Data Relay, and Wireless Power Transfer in UAV-Based Environmental Monitoring System [†]

Jaewook Lee ¹  and Haneul Ko ^{2,*} 

¹ Department of Information and Communication Engineering, Pukyong National University, Busan 48513, Republic of Korea; jlee0315@pknu.ac.kr

² Department of Electronic Engineering, Kyung Hee University, Yongin-si 17104, Republic of Korea

* Correspondence: heko@khu.ac.kr

[†] This paper is an extended version of our paper published in A preliminary version of this paper was presented at International Conference on Information Networking (ICOIN) 2023, Bangkok, Thailand, 11–13 January 2023 .

Abstract: In environmental monitoring systems based on the Internet of Things (IoT), sensor nodes (SNs) typically send data to the server via a wireless gateway (GW) at regular intervals. However, when SNs are located far from the GW, substantial energy is expended in transmitting data. This paper introduces a novel unmanned aerial vehicle (UAV)-based environmental monitoring system. In the proposed system, the UAV conducts patrols in the designated area, and SNs periodically transmit the collected data to the GW or the UAV. This transmission decision is made while taking into account the respective distance between both the GW and the UAV. To ensure a high-quality environmental map, characterized by a consistent collection of a satisfactory amount of up-to-date data while preventing energy depletion in the SNs and the UAV, the UAV periodically decides on three types of UAV operations. These decisions involve deciding where to move, deciding whether to relay or aggregate the data from the SNs, and deciding whether to transfer energy to the SNs. For the optimal decisions, we introduce an algorithm, called DeepUAV, using deep reinforcement learning (DRL) to make decisions in UAV operations. In DeepUAV, the controller continually learns online and enhances the UAV's decisions through trial and error. The evaluation results indicate that DeepUAV successfully gathers a substantial amount of the current data consistently while mitigating the risk of energy depletion in SNs and the UAV.

Keywords: Internet of Things (IoT); unmanned aerial vehicle (UAV); deep reinforcement learning (DRL)



Citation: Lee, J.; Ko, H. Joint Optimization on Trajectory, Data Relay, and Wireless Power Transfer in UAV-Based Environmental Monitoring System. *Electronics* **2024**, *13*, 828. <https://doi.org/10.3390/electronics13050828>

Academic Editor: Hosam El-Ocla

Received: 12 January 2024

Revised: 7 February 2024

Accepted: 16 February 2024

Published: 21 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last few years, the Internet of Things (IoT) has been rapidly growing across the world, which enables users to experience various environmental monitoring services [1]. In these services, the monitoring server periodically updates its service (e.g., a traffic and air quality monitoring map) by analyzing the received fresh data (e.g., the air quality and traffic intensity on each location) from sensor nodes (SNs) during the service update cycle [1–3]. To provide high-quality environmental monitoring services, fresh data need to be collected steadily from as many SNs as possible [1,4]. Nonetheless, gathering consistent data from numerous SNs is challenging because of their limited battery capacity. To address this issue, various systems have incorporated intermediate nodes capable of relaying data and/or transmitting RF energy [5,6]. However, the efficiency of incorporating intermediate nodes can be not enough due to numerous obstacles and significant propagation loss. Deploying intermediate nodes in an ultra-dense manner can yield substantial benefits, such as energy savings. However, this approach comes with a notable drawback as it significantly escalates the deployment costs. To mitigate this problem, a number of works (e.g., directional antenna transmission [7], scheduling [8,9], and energy waveform optimization [10,11]) have been investigated in the literature. One of the most promising solutions is introducing

unmanned aerial vehicles (UAVs) [12–21] as intermediate nodes to collect data and/or transfer energy, because UAVs can achieve flexibly shorter distance and line-of-sight (LoS) links thanks to their mobility. These solutions fall into two main categories: (1) UAV-based data collection [12–16]; and (2) UAV-based wireless power transfer [17–21]. However, as these studies did not jointly optimize policies for both data collection and wireless power transfer via a UAV, achieving high service quality becomes challenging.

To address this issue and achieve high service quality, we introduce a UAV-based environmental monitoring system designed to optimize UAV operations for data collection, wireless energy transfer, and trajectory. To accomplish this objective, UAV navigates the specified region, and SNs intermittently send their collected data to the gateway (GW) or UAV by taking into account the respective distances to both the GW and UAV. To optimize the service quality of the environmental map (to ensure sufficient fresh data without depleting the energy of the SNs), the UAV periodically makes three types of decisions: (1) determining its movement, (2) deciding whether to relay or aggregate data from SNs, and (3) determining whether to transfer energy to SNs. To achieve these optimal decisions, we introduce a deep reinforcement learning (DRL)-based UAV operation decision algorithm named DeepUAV. Due to the huge scale of the decision space, it is difficult to apply the conventional optimization techniques to our problem. Thus, we introduce the reinforcement learning-based approach for deciding the operation of the UAV. In DeepUAV, the controller commands the optimal decisions of the UAV (i.e., trajectory, data relay, and wireless power transfer). These decisions are made by taking into account the transmission distance between the UAV and GW, the freshness of the aggregated data in the UAV, and the energy levels of the UAV and SNs. It is important to highlight that the controller undergoes continuous online learning, refining the three kinds of decisions through a process of trial and error. Specifically, the information derived from state observation and learning experiences are integrated into a nonlinear function approximator (i.e., a neural network). This neural network undergoes iterative training to provide the controller with near-optimal decisions corresponding to each state. To evaluate and compare the service quality of DeepUAV with that of the related works, we perform an event-driven simulation in an environment where the general crowdsensing map application is requested. From the evaluation results, we demonstrate that DeepUAV can increase the service quality by up to 44% compared to the trajectory and wireless power transfer optimal scheme because DeepUAV simultaneously optimizes three types of UAV operations based on the given environment.

The main contribution of this paper is three-fold: (1) The joint optimization of the trajectory, data relay, and wireless power transfer of the UAV is conducted and thus a sufficient number of fresh data can be collected, leading to the high service quality of the environmental monitoring systems. The previous works focused on optimizing individual UAV operations, such as the trajectory, data relay, or wireless power transfer operations. (2) Because DRL can be exploited in the high-dimensional decision-making problems, DeepUAV can be implemented in practical systems. (3) Extensive evaluation results are presented and scrutinized across diverse environments, offering insights for the development of systems involving UAV-based data collection and wireless power transfer.

The remainder of this paper is structured as follows: Section 2 provides a summary of the related works, and Section 3 presents the system model. Section 4 details the development of DeepUAV. Then, the evaluation results are given in Section 5, followed by the concluding remarks in Section 6.

2. Related Works

To improve the service quality in an environmental monitoring system, a number of works to exploit a UAV have been reported in the literature [12–21], which can be categorized into (1) UAV-based data collection [12–16] and (2) UAV-based wireless power transfer [17–21].

Dai et al. [12] presented a centralized solution based on DRL, focusing on regulating the UAV trajectory and scheduling SNs for data collection. The primary goal is to reduce UAV energy usage while maintaining data freshness. Li et al. [14] proposed a scheme known as the DQN-based Flight Resource Allocation Scheme (DQN-FRAS) to minimize network data loss by determining the UAV data trajectory and data collection schedule. Sun et al. [15] presented an Age of Information (AoI)- and energy-aware data collection scheme tailored for UAV-assisted IoT networks, focusing on minimizing the weighted sum of the expected average AoI and UAV propulsion energy. Liu et al. [13] introduced a distributed control framework for maximizing energy efficiency, the data collection ratio, geographic fairness, and minimizing energy consumption through UAV trajectory planning with charging stations. Dai et al. [16] introduced a UAV crowdsensing framework that prioritizes both minimizing energy consumption across all UAVs and ensuring data freshness while also aiming to maximize the collected data amount and geographical fairness simultaneously.

Yang et al. [17] proposed a novel UAV-enabled hybrid communication system jointly optimizing the UAV's transmission power and trajectory and the SNs' transmission power and duration to maximize the total energy efficiency (EE) of the SNs. Suman and De [18] analyzed UAV-aided RF energy transfer performance and then they provided a closed-form expression on the received power from a UAV. Jiang [19] analyzed a degradation of the received power at SNs due to the inappropriate movement of the UAV in UAV-assisted wireless information and energy transfer systems. Liu et al. [20] developed closed-form expressions for the energy outage probability and rate outage probability, formulating an optimization problem aimed at minimizing the overall outage probability of SNs. Nguyen et al. [21] explored a simultaneous wireless power transfer and information transmission scheme with reconfigurable intelligent surface (RIS)-assisted UAV communication, optimizing the UAV trajectory, power allocation, and phase-shift matrix of the RIS.

Although these related works have achieved high service quality by formulating policies for either data collection from SNs or wireless power transfer, their potential for enhanced service quality increases significantly if these works jointly determine policies for both data collection and energy wireless transfer. For instance, consider a situation where no joint optimization on the trajectory and data relay is conducted. In this situation, even when the UAV is instructed to move closer to the GW (i.e., even though there is a chance that the UAV can relay the data with smaller energy consumption), the UAV can relay the data to the GW at the current location, which can cause the degradation of the service quality due to the energy depletion of the UAV. To obtain high service quality more than the related works, we introduced a UAV system that jointly optimizes the trajectory, data relay, and wireless power transfer in [22] and in this paper, respectively. In [22], the proposed system follows the static data relay policy of transmitting data to the GW after a certain number of data are aggregated. Thus, we cannot achieve high service quality. Thus, we have extended the system in [22] to dynamically determine the policy according to the environment.

3. UAV-Based Environmental Monitoring System

The system model in this paper is illustrated in Figure 1. It comprises two planes: (1) the ground plane; and (2) the UAV plane. Each plane is subdivided into multiple locations, and each location is identified by l . In the ground plane, I SNs are strategically deployed and equipped with distinct battery capacities. Let L_i denote the location of SN i and $E_{SN,i}^{MAX}$ represent the battery capacity of SN i . Additionally, a fixed charger for the unmanned aerial vehicle (UAV) is positioned at the charging location identified as L_C on the ground plane. It is assumed that the monitoring server and the UAV controller are co-located at the gateway (GW).

As depicted in Figure 2, we focus on the environmental monitoring services (e.g., traffic and air quality monitoring services). In this service, the monitoring server aggregates the data (e.g., traffic and air quality intensity) from SNs at the location of interest (LOI) during

the service update cycle T_U . The environmental monitoring map is then updated at the end of each cycle [1–3,23]. We assume that the data sensed by the SNs are dynamically changed.

Specifically, as shown in Figure 2, when it is assumed that the monitoring server periodically updates the environmental map every 5 time slots (i.e., service update cycle T_U is 5), the monitoring server collects the data from SNs during 5 time slots. Then, at the end of the service update cycle, the server updates the environmental map by analyzing the collected data (e.g., predicting the air quality for the air quality map).

Intuitively, if the monitoring server acquires more data during the service update cycle, it can lead to higher service quality, as discussed in [1]. To achieve this, the controller directs the operations of the UAV, including the movement, data relay, and wireless power transfer. Subsequent subsections provide detailed descriptions of the operations of both the sensor nodes (SNs) and the UAV.

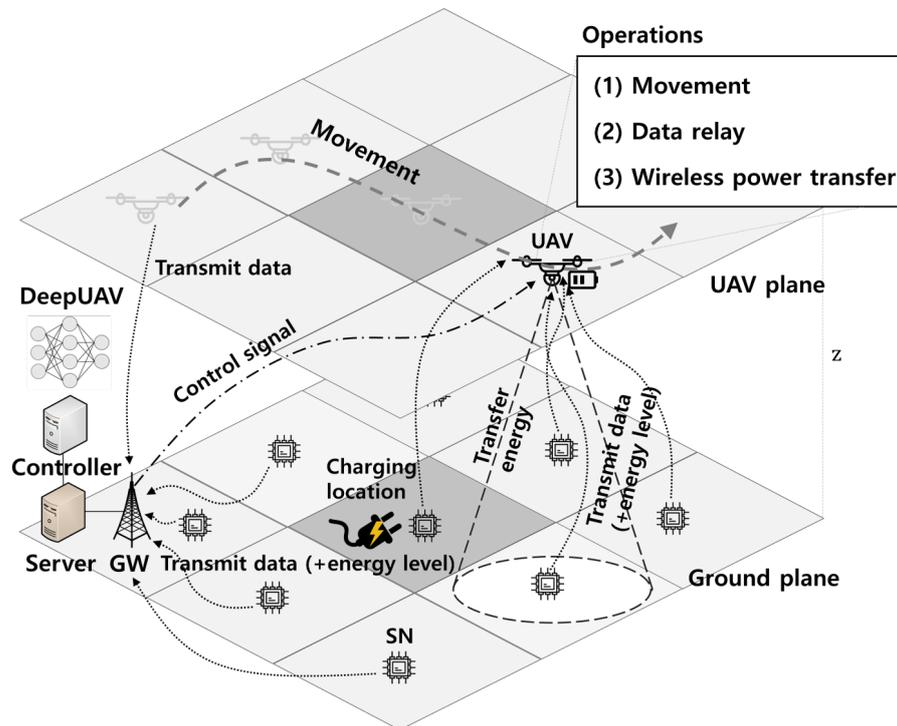


Figure 1. System model.

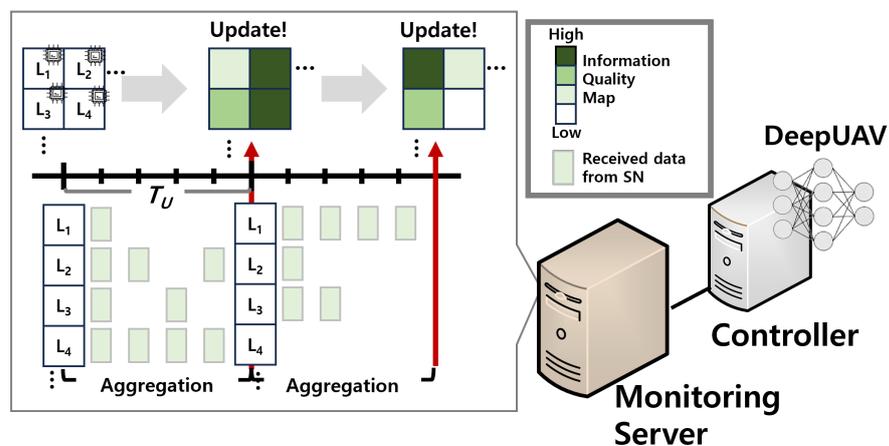


Figure 2. Operating example of the environmental monitoring service.

3.1. Operations of SN

On the ground plane, each SN collects and/or senses its designated target (e.g., air condition and temperature) by expending its energy e_s . Consequently, when the SN runs out of energy, it becomes incapable of collecting and/or sensing the target. After sensing the target, the SNs transmit the data to the GW or UAV, considering the distances to both the GW and the UAV. If the UAV emits a short signal pulse, the SNs can estimate the distance to the UAV by analyzing the received signal power. Additionally, because the GW and SNs are stationary, determining the distance between them is straightforward. In particular, when the distance to the GW is shorter than that to the UAV, the SN sends the sensed data directly to the GW. Conversely, if the distance to the GW is longer than that to the UAV, the SN transmits the data to the UAV. This strategy can reduce the energy consumption for data transmission.

To transmit the sensed data with the minimum energy consumption, it is assumed that SNs use the minimum transmission power $P_T^{\min,d}$ to achieve the desired bit error rate. The minimum transmission power can be calculated as $P_T^{\min,d} = \frac{P_R^{\min}}{|h|^2}$ [24] where h and d are the channel coefficient and the distance between the source node and the destination node, respectively, and P_R^{\min} is the receiving minimum power at the destination node. On the other hand, h is defined as $\sqrt{\beta\hat{h}}$ where \hat{h} and β are denoted by the large-scale fading effect and the small-scale fading, respectively. Especially, β is $\frac{1}{\eta} \left(\frac{4\phi f_c d}{c}\right)^{-\alpha}$, where f_c and c represent the carrier frequency and the speed of light, respectively. α and η denote the path loss exponent and the additional path loss. Therefore, the energy consumption of the SNs for data transmission increases in accordance with the distance.

Meanwhile, upon receiving data from the SN, the application server dispatches an acknowledgment message to the SN. Subsequently, the SN reevaluates its target by monitoring and/or sensing it, before transmitting the data to either the UAV or GW. The energy status of the SN is included in the packet for the sensed data.

3.2. Operations of UAV

On the UAV plane, the UAV maintains a constant altitude denoted as z . In practical systems, the altitude is typically set to the minimum level that avoids obstacles, such as buildings and trees, without the need for frequent descent and ascent [25]. The operations of the UAV fall into three categories: (1) movement; (2) data relay; and (3) wireless power transfer. In other words, the controller makes determinations and issues commands for these operations to the UAV at each time epoch.

For the movement operation, the UAV possesses the capability to either hover in its current position or traverse to an adjacent location, utilizing energies e_H and e_M , respectively. Meanwhile, a fixed charging station for the UAV is stationed at a specific location identified as L_C . Staying in proximity to the charging location enables the UAV to undergo frequent charging and thus it can sustain operations without energy depletion. However, this scenario, where the UAV remains close to the charging location, poses a challenge for SNs deployed far from the charging point. In such instances, these SNs cannot receive assistance from the UAV, i.e., they are unable to transmit sensed data to the UAV or be charged by it. Consequently, the lifetime of the SNs is expected to be shortened, resulting in inadequate data collection. To prevent this situation, the decision regarding the movement operation of the UAV must account for the energy levels of both the UAV and SNs.

As a relay node between the SNs and GW, the UAV has the capability to aggregate data. If the UAV were to relay data to the GW immediately upon receiving it from the SNs, its energy could be depleted rapidly. To avoid this, the UAV can aggregate data and thus it does not have to transmit it to the GW right away upon reception. Instead, it can wait until a certain amount of data are accumulated and/or until the UAV is in close proximity to the GW. At that point, the UAV can deliver the aggregated data simultaneously using a single packet. However, if the data are excessively aggregated, there is a risk of the current service update cycle concluding before the data are transmitted to the GW, rendering the

aggregated data obsolete. Therefore, the decision regarding the transmission operation of the UAV should consider the energy levels of both the UAV and the service update cycle.

To facilitate efficient wireless power transfer, we assume that the UAV utilizes the directional antenna with the fixed angle [26]. Specifically, the UAV at location i in the UAV plane can transfer the energy to only the SN at location i in the ground plane, and this power transfer consumes the energy e_p . It is unnecessary to conduct wireless power transfer to SNs that already possess sufficient energy. Additionally, if the UAV is positioned far from its charging location and excessively engages in wireless power transfer, it may face challenges returning to the charging location due to energy depletion. In summary, the decision regarding the wireless power transfer should consider the current location of the UAV and the energy levels of the UAV and SNs, jointly.

To simultaneously optimize the three types of operations (i.e., movement, data relay, and wireless power transfer), a DeepUAV algorithm is introduced in the subsequent section. In DeepUAV, the controller undergoes continuous online learning, refining the UAV's policies (i.e., sequences of operations) through trial and error. More precisely, state observation information and learning experiences are integrated within a nonlinear function approximator (i.e., a neural network). This neural network is iteratively trained, providing the controller with near-optimal decisions for each state.

4. A Deep Reinforcement Learning (DRL)-Based UAV Operation Decision Algorithm (DeepUAV)

In this section, we present DeepUAV for the optimal policy on the movement, data relay, and wireless power transfer of the UAV. For this, we first define the state space, action space, and reward of the DRL agent (i.e., controller). Important notations in this paper are summarized in Table A1 in Appendix A.

4.1. State Space

We define the state space \mathbf{S} as

$$\mathbf{S} = \mathbf{L} \times \mathbf{E}_U \times \mathbf{T} \times \prod_i (\mathbf{E}_{SN,i} \times \mathbf{D}_{SN,i} \times \mathbf{G}_{SN,i}) \tag{1}$$

where \mathbf{L} and \mathbf{E}_U denote the state spaces for the location and the energy level of the UAV, respectively. \mathbf{T} represents the state space for the remaining time to the next service update cycle. In addition, $\mathbf{E}_{SN,i}$ are the state spaces for the energy level of SN i . $\mathbf{D}_{SN,i}$ and $\mathbf{G}_{SN,i}$ represent the state spaces for the numbers of aggregated data from SN i at the server and the UAV, respectively.

\mathbf{L} is described by

$$\mathbf{L} = \{1, 2, \dots, N_L\} \tag{2}$$

where N_L is the number of locations in the target area.

\mathbf{E}_U is denoted as

$$\mathbf{E}_U = \{0, 1, 2, \dots, E_U^{MAX}\} \tag{3}$$

where E_U^{MAX} is the battery capacity of the UAV.

$\mathbf{E}_{SN,i}$ is represented by

$$\mathbf{E}_{SN,i} = \{0, 1, 2, \dots, E_{SN,i}^{MAX}\} \tag{4}$$

where $E_{SN,i}^{MAX}$ is the battery capacity of the SN.

Because T_U denotes the service update cycle, \mathbf{T} can be defined as

$$\mathbf{T} = \{0, 1, 2, \dots, T_U\}. \tag{5}$$

$D_{SN,i}$ and $G_{SN,i}$ are described by

$$D_{SN,i} = \{0, 1, 2, \dots, D_{max}\} \quad (6)$$

and

$$G_{SN,i} = \{0, 1, 2, \dots, G_{max}\}. \quad (7)$$

where D_{max} and G_{max} are represented by the maximum number of the aggregated data at the UAV and the application server, respectively.

4.2. Action Space

Because the UAV should make three types of decisions, the action space \mathbf{A} can be defined as

$$\mathbf{A} = \mathbf{A}_M \times \mathbf{A}_T \times \mathbf{A}_P \quad (8)$$

where \mathbf{A}_M represents the action space for the movement direction. \mathbf{A}_T and \mathbf{A}_P describe the action space for the data relay and the wireless energy transfer, respectively.

\mathbf{A}_M can be described as

$$\mathbf{A}_M = \{0, 1, 2, 3, 4\} \quad (9)$$

where $A_M = 0$ represents that the UAV hovers at the current location. In addition, $A_M \neq 0$ denotes the movement direction. Specifically, when $A_M = 1$, $A_M = 2$, $A_M = 3$, and $A_M = 4$, the UAV moves to the east, west, south, and north, respectively. The movement action space \mathbf{A}_M can be defined with more direction. However, it is more difficult to find an optimal policy as the action space is bigger. Thus, we defined \mathbf{A}_M by the basic movement operation.

\mathbf{A}_T can be represented as

$$\mathbf{A}_T = \{0, 1\} \quad (10)$$

where A_T denotes whether to transmit or aggregate the data. In other words, if $A_T = 1$, the UAV transmits the aggregated data to the GW. Otherwise (i.e., $A_T = 0$), the UAV aggregates data to the GW.

\mathbf{A}_P can be represented by

$$\mathbf{A}_P = \{0, 1\} \quad (11)$$

where A_P denotes whether the UAV transfers energy for the SN at its current location. That is, $A_P = 1$ represents that the UAV transfers energy for the SN at its current location. On the other hand, if $A_P = 0$, the UAV does not transfer any energy.

4.3. Reward Function

To define the reward function r , we consider the service quality of the environmental monitoring system. In general, the service quality increases as the number of aggregated data increases. However, the additionally increasing service quality diminishes due to a law of diminishing marginal utility [4]. Specifically, when the monitoring server already receives a sufficient number of data from SN i for the current service cycle, the additional data from SN i do not improve the service quality. Meanwhile, because the monitoring server updates the service at the end of the service cycle (i.e., $T = 0$), the monitoring server can obtain a reward (i.e., service quality) at only the end of the service cycle. Therefore, we can define the reward function r (the reward function in this work represents the service

quality of the generalized and simplified application from environmental monitoring applications [1–3]) as

$$r = \delta[T = 0] \frac{\sum_i \min(\frac{D_{SN_i}}{N_D}, 1)}{I} \tag{12}$$

where $\delta[\cdot]$ is a delta function that returns 1 only when the given condition (e.g., $T = 0$) is true. In addition, N_D is the sufficient number of data from SN i for the current service cycle.

4.4. DeepUAV

To obtain the optimal operations (i.e., movement, data relay, and wireless power transfer) of the UAV in a huge scale of the decision space, we propose DeepUAV based on the model-free deep RL algorithm (i.e., deep Q-network (DQN)) [27]. Even though the traditional Q-learning algorithm can obtain the optimal operations of a UAV, it is not practical in real environmental monitoring systems. In the Q-learning algorithm, Q-values for every state–action pair should be calculated and stored in a table. However, in practical environmental monitoring systems, the number of possible states can be tremendous due to the huge scale of the systems (e.g., a huge number of SNs) and then huge memory is needed. Moreover, it is difficult to get enough samples for each state–action pair, which means that the Q-learning algorithm cannot be converged. For this, the optimal action-value function $Q^*(S, A)$ can be defined by [27]

$$Q^*(S, A) = \max_{\pi} E \left[r_t + \sum_k \gamma^k r_{t+k} | S_t, A_t, \pi \right] \tag{13}$$

where π represents a policy that maps states to actions, and γ denotes the discount factor, determining the degree of importance given to future predictions. Subsequently, a deep neural network, specifically a Q-network as introduced in [27], is employed to approximate an action-value function denoted as $Q(S, A, \theta)$, where θ represents the weights of the network. This neural network takes state information as the input and produces Q-values corresponding to each possible action. During the training, the weights θ are adjusted iteratively to minimize the loss function. The loss function at the i th iteration is defined as

$$L_i(\theta_i) = E \left[(y - Q(S_t, A_t, \theta_i))^2 \right] \tag{14}$$

where y is the target value, which can be defined as (15) at the top of the next page.

$$y = \begin{cases} r_t, & \text{if episode terminates at the next state} \\ r_t + \gamma \max_{A'} Q(S_{t+1}, A', \theta_{i-1}), & \text{otherwise} \end{cases} \tag{15}$$

To update weights θ appropriately, we utilize the experience replay and the target network separation as similar to that in [27]. Note that the experience replay breaks the correlation between successive experiences and allows for the model to learn the underlying distribution of labels. Meanwhile, the target network separation is for a stable deep learning procedure.

Figure 3 illustrates the diagram of DeepUAV, while Algorithm 1 delineates the detailed procedure of DeepUAV. Initially, the algorithm initializes essential components, including the replay memory D , the main deep Q-network, and the target deep Q-network (lines 1–3 in Algorithm 1). Subsequently, the DeepUAV algorithm conducts the training phase. First, the controller obtains the initial state S_t (line 5 in Algorithm 1) and determines the action A_t with the ϵ -greedy method (lines 7–12 in Algorithm 1). Following this, the algorithm instructs the UAV with the chosen action A_t and observes the resulting reward r_t and the

next state S_{t+1} . To facilitate the experience replay, the algorithm records the experience (S_t, A_t, r_t, S_{t+1}) in the replay memory D and randomly chooses a set of experiences from this memory (line 15 in Algorithm 1). Using these selected experiences, the algorithm sets the target value y_i by updating the target deep Q-network (line 16 in Algorithm 1) and proceeds to train the main deep Q-network using the loss function and the gradient descent method (line 17 in Algorithm 1). Subsequently, the target deep Q-network is updated periodically by copying the weights from those of the main deep Q-network (line 18 in Algorithm 1).

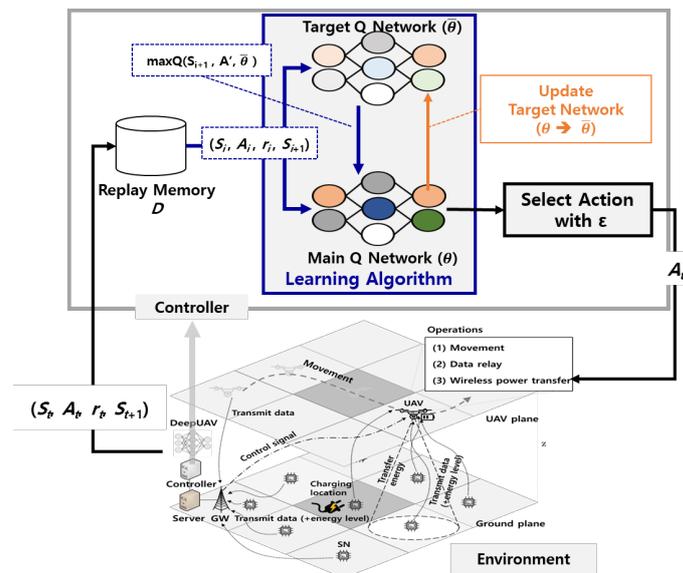


Figure 3. DeepUAV diagram.

5. Evaluation Results

For the performance evaluation, we compare the proposed algorithm, DeepUAV, with the following three schemes: (1) No-UAV where there is no UAV to relay data from SNs and to transfer energy to SNs; (2) DeepCharge, which follows the optimal policy on the wireless energy transfer and the movement of the UAV while adhering to the fixed policy of the always data relay (i.e., always $A_T = 1$); and (3) DeepPush, which follows the optimal policy on the data relay and the movement of the UAV while adhering to the fixed policy of no wireless energy transfer (i.e., always $A_P = 0$). Note that DeepCharge and DeepPush are designed by generalizing the UAV-based data collection schemes and UAV-based wireless power transfer schemes in Section 2.

For the performance evaluation, we consider a 10×10 plane (i.e., 100 locations) where 50 SNs are randomly deployed. In addition, the UAV charger and GW are located at the center of the locations. Also, we consider the general crowdsensing map application. This application generates the environment map (e.g., temperature map) in the urban area based on the aggregated SNs' data (e.g., temperature information) [1].

The energy consumption parameters are set as follows. First, we normalize the energy consumption for sensing data e_S as 1 J. Then, comparing the complexity of the operations, $e_H, e_M, e_P, e_T,$ and e_C are set to 1, 1, 10, 0.2, and 100 J. Also, the maximum battery capacities of the UAV and SN, E_U^{MAX} and $E_{SN}^{MAX,i}$, are set to 500 and 70 J, respectively. The service update cycle and a sufficient number of data from each SN for the current service update cycle, T_U and N_D , are set as 5 and 2, respectively. We consider the average of the service quality based on Equation (12) during the service operation time as the performance parameters.

Algorithm 1 Deep reinforcement learning (DRL)-based UAV operation decision algorithm.

```

1: Make and Rest replay memory  $D$ 
2: Make and Rest the deep Q-network having weights  $\theta$ 
3: Make and Rest target deep Q-network having weights  $\bar{\theta} = \theta$ 
4: for each episode  $j = 1, 2, 3, \dots, J$  do
5:   Obtain the initial state  $S_t$ 
6:   for  $t = 1, 2, 3, \dots, T$  do
7:     Randomly Determines probability value  $p$ 
8:     if  $p \leq \epsilon$  then
9:       select a random action  $A_t$ 
10:    else
11:       $A_t = \arg \max_A Q(S_t, A, \theta)$ 
12:    end if
13:    Command the action  $A_t$  to UAV, and obtain and calculate the next state  $S_{t+1}$  and
    the reward  $r_t$  and  $t$ 
14:    Put experience  $(S_t, A_t, r_t, S_{t+1})$  into replay memory  $D$ 
15:    Get a set of experiences  $(S_i, A_i, r_i, S_{i+1})$  into replay memory  $D$ 
16:    Set
      
$$y_i = \begin{cases} r_i, & \text{if episode terminates at the next state} \\ r_i + \gamma \max_{A'} Q(S_{i+1}, A', \bar{\theta}), & \text{otherwise} \end{cases}$$

17:    Performs a gradient descent step on  $(y_i - Q(S_i, A_i, \theta))^2$  with respect to  $\theta$ 
18:    For every  $K$  steps, update the target deep Q-network with  $\bar{\theta} = \theta$ 
19:  end for
20: end for

```

5.1. Convergence Process Analysis

Figures 4a,b represent the service quality and the training loss during the training phase, respectively. In each epoch, 50 iterations are conducted. From Figures 4a,b, it can be shown that the service quality and training loss increase and decrease as the training epoch goes on, except for the initial epochs, respectively. This indicates that DeepUAV effectively learns how the selected UAV operations affect the achievable service quality. Meanwhile, at the initial training phase (i.e., epochs < 100), high training loss and low service quality are obtained because DeepUAV has less experience in learning how the selected UAV operations affect the service quality.

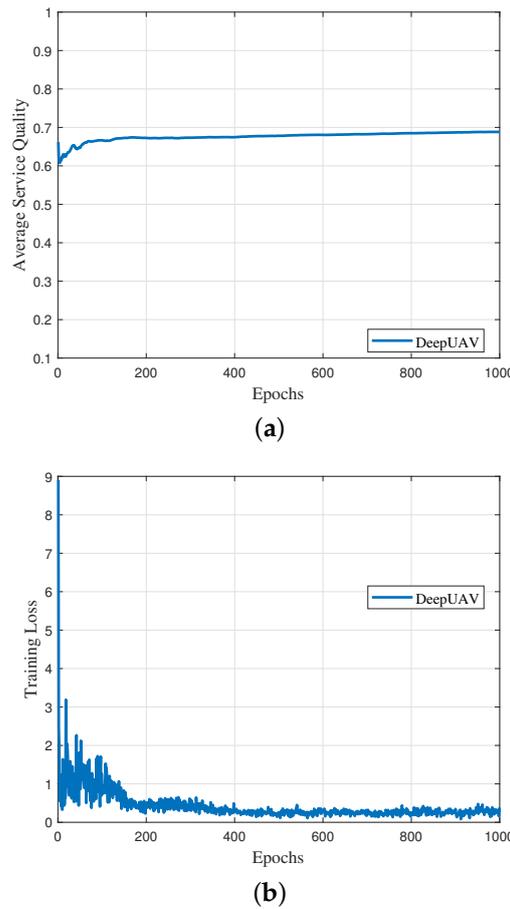


Figure 4. Learning results. (a) Expected total rewards. (b) Training loss.

5.2. Effect of I

Figure 5 represents the effect of the number of SNs I on the service quality. In this simulation result, DeepUAV shows the enhanced service quality by 28%~48% when it is compared to DeepPush and DeepCharge. Especially, in the default simulation setting, DeepUAV demonstrates increased service quality by up to 44%. From Figure 5, it can be found that DeepUAV has the highest service quality regardless of I . This is because DeepUAV simultaneously optimizes three types of decisions (i.e., the decision on where to move, decision on whether to relay or aggregate data, and decision on whether to transfer energy).

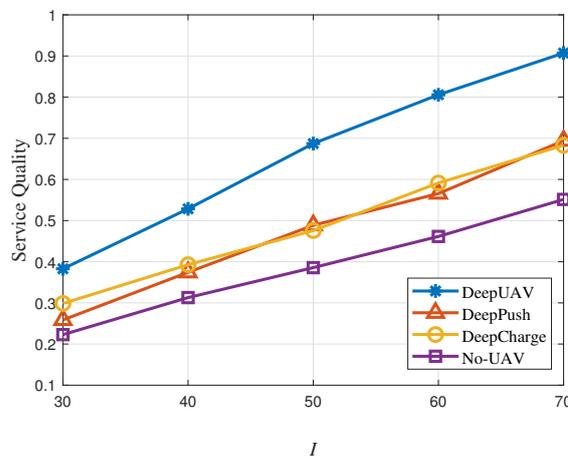


Figure 5. Effect of the number of SNs on the service quality.

Meanwhile, as shown in Figure 5, the service qualities of all the schemes increase as I increases. This is because more SNs can transmit their sensed data for the target application and more data indicate a higher service quality.

5.3. Effect of L_G

Figure 6 represents the effect of the number of locations in the ground plane, L_G , on the average service quality. DeepUAV increases its service quality by up to 31%~60% compared to DeepPush and DeepCharge. From Figure 6, it can be observed that the service qualities of all the schemes are degraded as the number of locations L_G increases. This is because, when L_G increases, the SNs are more sparsely deployed in the ground plane and thus the SNs need more energy consumption to transmit their data to the GW. In addition, the UAV needs to travel long distances to help the SNs, which causes more energy consumption of the UAV.

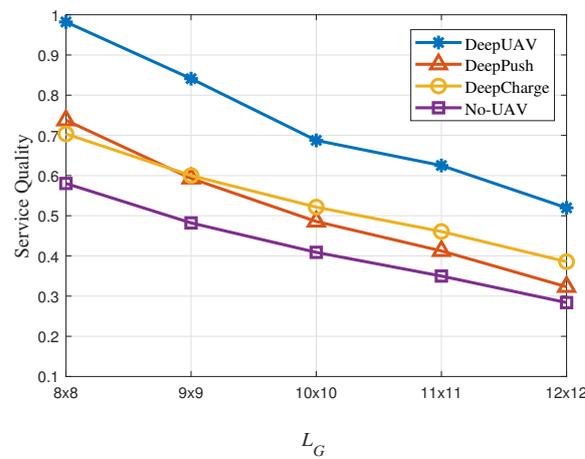


Figure 6. Effect of the number of locations on the service quality.

5.4. Effect of T_U

Figure 7 represents the effect of the service update cycle T_U on the service quality. In this simulation result, DeepUAV demonstrates that it can enhance the service quality by up to 13%~58% compared to DeepPush and DeepCharge. When T_U is set to a short value, SNs need to frequently sense the target data and report the sensed data to the monitoring server. In this situation, SNs consume more energy within a short duration, which can cause the energy depletion of SNs. Thus, as shown in Figure 7, the service qualities of all the schemes decrease as T_U decreases.

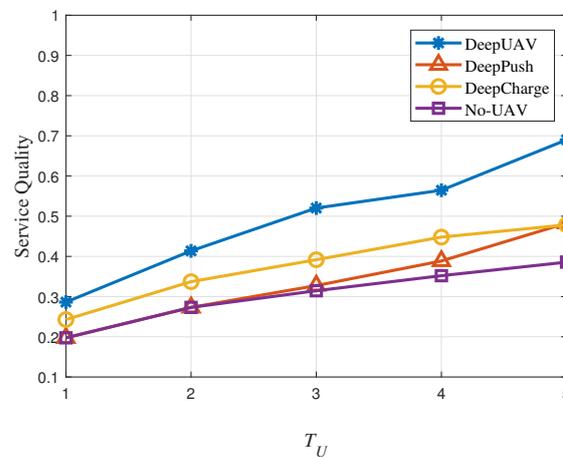


Figure 7. Effect of the service update cycle on the service quality.

Meanwhile, from Figure 7, it can be found that DeepPush has almost the same service quality with No-UAV when a service update cycle is short (i.e., T_U is 1, 2, or 3). This is because, when a service update cycle is short, most SNs directly transmit their data to the GW before the UAV approaches. That is, the effectiveness of introducing the UAV in DeepPush is not significant.

5.5. Effect of $E_{SN,i}^{MAX}$

Figure 8 shows the effect of the average battery capacity of SNs, $E[E_{SN,i}^{MAX}]$. When DeepUAV is compared to DeepPush and DeepCharge, DeepUAV can enhance the service quality by up to 24%~100% compared to DeepPush and DeepCharge. From Figure 8, the service qualities of all the schemes increase as the battery capacity of the SNs increases. This is because SNs having more battery capacity can sense and transmit more data to the GW for a longer time.

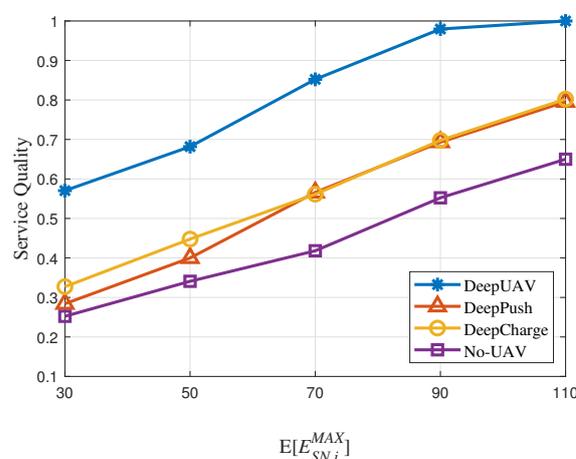


Figure 8. Effect of E_{SN}^{MAX} .

Meanwhile, when SNs have a lower battery capacity, the charging energy of SNs to prevent the energy depletion of SNs is more important than reducing the transmission energy consumption via the UAV relay operation. Therefore, when SNs have a low battery capacity (i.e., $E[E_{SN,i}^{MAX}] < 70$), the service qualities of DeepCharge and DeepUAV are better than that of DeepPush.

6. Conclusions

To enhance the service quality with a sufficient number of fresh data, we proposed a deep reinforcement learning (DRL)-based UAV operation decision algorithm (DeepUAV). DeepUAV interacts with the monitoring environment to learn and improve the UAV operation decisions (i.e., movement, data relay, and wireless power transfer) through trial and error. The evaluation results demonstrate that DeepUAV improves the service quality by up to 44% compared to the scheme with non-fully optimal operations. In this work, because we focused on a single UAV, the UAV in DeepUAV infrequently transfers energy to SNs located far from the GW and frequently returns to the GW for recharging. Consequently, we cannot achieve better service quality. To address our limitation, we plan to enhance our algorithm to consider multiple UAVs by using multi-agent deep reinforcement learning as one of our future works.

Author Contributions: Conceptualization, H.K.; methodology, J.L.; software, J.L.; validation, J.L. and H.K.; formal analysis, J.L. and H.K.; investigation, H.K.; resources, H.K.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L. and H.K.; visualization, J.L. and H.K.; supervision, H.K.; project administration, H.K.; funding acquisition, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation (NRF) of Korea Grant funded by the Korean Government (MSIP) (No. 2022R1F1A1063183).

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Important notations in this paper are summarized in Table A1.

Table A1. Summary of notations.

Notation	Description
\mathbf{S}	State space
\mathbf{L}	State space for the location of the UAV
\mathbf{E}_U	State space for the energy level of the UAV
\mathbf{T}	State space for the remaining time to the next service update cycle
$\mathbf{E}_{SN,i}$	State space for the energy level of the SN i
$\mathbf{D}_{SN,i}$	State space for the number of aggregated data from SN i at the UAV
$\mathbf{G}_{SN,i}$	State space for the number of aggregated data from SN i at the GW
E_U^{MAX}	Battery capacity of the UAV
$E_{SN,i}^{MAX}$	Battery capacity of SN i
T_U	Service update cycle
\mathbf{A}	Action space
\mathbf{A}_M	Action space for the movement direction
\mathbf{A}_T	Action space for the data relay
\mathbf{A}_P	Action space for the wireless power transfer
e_H	Energy consumption to hover at the current location
e_M	Energy consumption to move to an adjacent location
e_S	Energy consumption to sense data via the SN
e_T	Energy consumption per distance to transmit data energy from the SN to the UAV or GW
e_P	Energy consumption to transfer energy to the SN from the UAV
e_C	Energy consumption to transfer energy to the UAV from the UAV charger
r	Reward
$Q(S, A)$	Action-value function
π	Policy mapping states to actions
γ	Discount factor to control the amount of weight to future prediction
θ	Weights of the deep neural network
$L_i(\theta)$	Loss function at the i th iteration with weights θ
D	Replay memory

References

1. Liu, W.; Yang, Y.; Wang, E.; Wang, L.; Zeglache, D.; Zhang, D. Multi-Dimensional Urban Sensing in Sparse Mobile Crowdsensing. *IEEE Access* **2019**, *7*, 82066–82079. [[CrossRef](#)]
2. Sun, H.; Ma, Y.; Quek, T.Q.S.; Wang, X.; Guo, K.; Zhang, H. MDC Enhanced IoT Networks: Network Modeling and Performance Analysis. *IEEE Internet Things J.* **2023**, *7*, 839–854. [[CrossRef](#)]
3. Li, Y.; Chai, Z.; Tan, F.; Liu, X. Temporal Data Scheduling in Internet of Vehicles Using an Improved Decomposition-Based Multi-Objective Evolutionary Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 5282–5295. [[CrossRef](#)]

4. Ko, H.; Kim, T.; Jung, D.; S, P. Software-Defined Electric Vehicle (EV)-to-EV Charging Framework With Mobile Aggregator. *IEEE Syst. J.* **2023**, *17*, 2815–2823. [[CrossRef](#)]
5. Wang, X.; Li, J.; Ning, Z.; Song, Q.; Guo, L.; Guo, S.; Obaidat, M.S. Wireless powered mobile edge computing networks: A survey. *ACM Comput. Surv.* **2023**, *55*, 263. [[CrossRef](#)]
6. Zhang, F.; Yu, C.; Tang, S. Unified Performance Analysis of Stochastic Clustered Cooperative Systems With Distance-Based Relay Selection. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 6180–6194.
7. Zhu, Z.; Wang, N.; Hao, W.; Wang, Z.; Lee, I. Robust beamforming designs in secure MIMO SWIPT IoT networks with a nonlinear channel model. *IEEE Internet Things J.* **2021**, *8*, 1702–1715. [[CrossRef](#)]
8. Sha, Q.; Liu, X.; Ansari, N. Efficient Multiple Green Energy Base Stations Far-Field Wireless Charging for Mobile IoT Devices. *IEEE Internet Things J.* **2023**, *10*, 8734–8743. [[CrossRef](#)]
9. Liu, X.; Ansari, N.; Sha, Q.; Jia, Y. Efficient Green Energy Far-Field Wireless Charging for Internet of Things. *IEEE Internet Things J.* **2022**, *9*, 23047–23057. [[CrossRef](#)]
10. Zhang, Y.; Clerckx, B. Waveform Design for Wireless Power Transfer with Power Amplifier and Energy Harvester Non-Linearities. *IEEE Trans. Signal Process.* **2023**, *71*, 2638–2653. [[CrossRef](#)]
11. Feng, Z.; Clerckx, B.; Zhao, Y. Waveform and Beamforming Design for Intelligent Reflecting Surface Aided Wireless Power Transfer: Single-User and Multi-User Solutions. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 5346–5361. [[CrossRef](#)]
12. Dai, Z.; Wang, H.; Liu, C.; Han, R.; Tang, J.; Wang, G. Mobile crowdsensing for data freshness: A deep reinforcement learning approach. In Proceedings of the IEEE Conference on Computer Communications, INFOCOM, Vancouver, BC, Canada, 10–13 May 2021; pp. 1–10.
13. Liu, C.; Dai, Z.; Zhao, Y.; Crowcroft, J.; Wu, D.; Leung, K.K. Distributed and energy-efficient mobile crowdsensing with charging stations by deep reinforcement learning. *IEEE Trans. Mob. Comput.* **2021**, *20*, 130–146. [[CrossRef](#)]
14. Li, K.; Ni, W.; Tovar, E.; Guizani, M. Joint flight cruise control and data collection in uav-aided internet of things: An onboard deep reinforcement learning approach. *IEEE Internet Things J.* **2021**, *8*, 9787–9799. [[CrossRef](#)]
15. Sun, M.; Xu, X.; Qin, X.P.Z. Aol-Energy-Aware UAV-Assisted Data Collection for IoT Networks: A Deep Reinforcement Learning Method. *IEEE Internet Things J.* **2021**, *8*, 17275–17289. [[CrossRef](#)]
16. Dai, Z.; Liu, C.; Han, R.; Wang, G.; Leung, K.; Tang, J. Delay-Sensitive Energy-Efficient UAV Crowdsensing by Deep Reinforcement Learning. *IEEE Trans. Mob. Comput.* **2023**, *22*, 2038–2052. [[CrossRef](#)]
17. Yang, H.; Ye, Y.; Chu, X.; Sun, S. Energy Efficiency Maximization for UAV-Enabled Hybrid Backscatter-Harvest-then-Transmit Communications. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 2876–2891. [[CrossRef](#)]
18. Suman, S.; De, S. Performance Analysis of UAV-aided RF Energy Transfer. In Proceedings of the IEEE International Conference on Communication Systems & Networks, COMSNETS, Bengaluru, India, 7–11 January 2020; pp. 575–578.
19. Jiang, R.; Xiong, K.; Liu, T.; Wang, D.; Zhong, Z. Coverage Probability-Constrained Maximum Throughput in UAV-Aided SWIPT Networks. In Proceedings of the IEEE International Conference on Communications Workshops, ICC Workshops, Dublin, Ireland, 7–11 June 2020; pp. 1–6.
20. Liu, Y.; Xiong, K.; Lu, Y.; Ni, Q.; Fan, P.; Letaief, K. UAV-Aided Wireless Power Transfer and Data Collection in Rician Fading. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 3097–3113. [[CrossRef](#)]
21. Nguyen, K.; Masaracchia, A.; Sharma, V.; Poor, H.; Duong, T. RIS-assisted UAV communications for IoT with wireless power transfer using deep reinforcement learning. *IEEE J. Sel. Top. Signal Process.* **2022**, *16*, 1086–1096. [[CrossRef](#)]
22. Lee, J.; Seo, S.; Ko, H. UAV-Based Data Collection and Wireless Power Transfer System with Deep Reinforcement Learning. In Proceedings of the IEEE International Conference on Information Networking, ICOIN, Bangkok, Thailand, 11–14 January 2023; pp. 400–403.
23. Peng, Z.; Gao, S.; Xiao, B.; Guo, S.; Yang, Y. CrowdGIS: Updating Digital Maps via Mobile Crowdsensing. *IEEE Trans. Autom. Sci. Eng.* **2018**, *15*, 369–380. [[CrossRef](#)]
24. Lee, J.; Ko, H. Energy and Distribution-Aware Cooperative Clustering Algorithm in Internet of Things (IoT)-Based Federated Learning. *IEEE Trans. Veh. Technol.* **2023**, *72*, 13799–13804. [[CrossRef](#)]
25. Zhou, F.; Wu, Y.; Hu, R.; Qian, Y. Computation Rate Maximization in UAV-Enabled Wireless Powered Mobile Edge Computing Systems. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 1927–1941. [[CrossRef](#)]
26. Ko, H.; Pack, S. Phase-Aware Directional Energy Transmission Algorithm in Multiple Directional RF Energy Source Environments. *IEEE Trans. Veh. Technol.* **2019**, *68*, 359–367. [[CrossRef](#)]
27. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.; Veness, J.; Bellemare, M.; Graves, A.; Riedmiller, M.; Fidjeland, A.; Ostrovski, G.; et al. Human-Level Control Through Deep Reinforcement Learning. *Nature* **2015**, *518*, 529–533. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.