future internet

MDPI

Article

Deep Reinforcement Learning for UAV-Based SDWSN Data Collection

Pejman A. Karegar, Duaa Zuhair Al-Hamid and Peter Han Joo Chong *

Department of Electrical and Electronic Engineering, Auckland University of Technology (AUT), Auckland 1010, New Zealand; pejman.karegar@aut.ac.nz (P.A.K.); duaa.alhamid@aut.ac.nz (D.Z.A.-H.)
* Correspondence: peter.chong@aut.ac.nz

Abstract: Recent advancements in Unmanned Aerial Vehicle (UAV) technology have made them effective platforms for data capture in applications like environmental monitoring. UAVs, acting as mobile data ferries, can significantly improve ground network performance by involving ground network representatives in data collection. These representatives communicate opportunistically with accessible UAVs. Emerging technologies such as Software Defined Wireless Sensor Networks (SDWSN), wherein the role/function of sensor nodes is defined via software, can offer a flexible operation for UAV data-gathering approaches. In this paper, we introduce the "UAV Fuzzy Travel Path", a novel approach that utilizes Deep Reinforcement Learning (DRL) algorithms, which is a subfield of machine learning, for optimal UAV trajectory planning. The approach also involves the integration between UAV and SDWSN wherein nodes acting as gateways (GWs) receive data from the flexibly formulated group members via software definition. A UAV is then dispatched to capture data from GWs along a planned trajectory within a fuzzy span. Our dual objectives are to minimize the total energy consumption of the UAV system during each data collection round and to enhance the communication bit rate on the UAV-Ground connectivity. We formulate this problem as a constrained combinatorial optimization problem, jointly planning the UAV path with improved communication performance. To tackle the NP-hard nature of this problem, we propose a novel DRL technique based on Deep Q-Learning. By learning from UAV path policy experiences, our approach efficiently reduces energy consumption while maximizing packet delivery.

Keywords: unmanned aerial vehicle (UAV); software-defined wireless sensor networks (SDWSN); fuzzy UAV route; deep reinforcement learning (DRL)

Citation: Karegar, P.A.; Al-Hamid, D.Z.; Chong, P.H.J. Deep Reinforcement Learning for UAV-Based SDWSN Data Collection. *Future Internet* **2024**, *16*, 398. https://doi.org/10.3390/fi16110398

Academic Editors: Panagiotis Papageorgas, Dimitrios Piromalis and Dionisis Kandris

Received: 24 August 2024
Revised: 17 October 2024
Accepted: 29 October 2024
Published: 30 October 2024

1. Introduction

The growing use of sensors for data collection and activity monitoring has demonstrated the applicability of wireless sensor networks (WSNs) in environmental, urban, and health-related applications. These networks utilize a variety of techniques for data dissemination and processing. However, ground-based WSNs, especially those deployed in environments like forests or farms, often experience substantial energy demands during the data-gathering process [1]. To mitigate this, the integration of Unmanned Aerial Vehicles (UAVs) offers a flexible and agile approach to enhance data acquisition efficiency. In this paper, the terms "Drone" and "UAV" are used interchangeably, referring to any aerial vehicle that facilitates vertical communication with ground-based WSNs [2]. The interaction patterns for data exchange between UAVs and ground-based WSNs differ significantly from conventional WSN operations. Contemporary UAVs possess unique capabilities, allowing them to tailor their flight paths to meet network demands and to strategically hover or circle over designated areas to optimize data retrieval and enhance overall efficiency. The advancement of the Internet of Things (IoT) has paved the way for new requirements for data collection and processing, particularly in applications related to activity monitoring using UAVs [3]. As part of an IoT architecture, UAVs serve two critical

roles: efficient data monitoring and collection from a wireless network, i.e., WSN, and data computation and decision-making. To enable flexible UAV-WSN operations, the WSN structure plays a crucial role in facilitating efficient data dissemination. Various grouping and clustering methods can organize wireless sensor nodes into star or tree structures, with representative nodes responsible for data gathering. Additionally, WSN topology can benefit from the latest technologies such as network virtualization and software-defined networking (SDN) [4,5]. In this context, wireless sensor nodes can be configured with a given function through software definition encouraging the conceptual development of software-defined wireless sensor networks (SDWSN). A function can be represented by the terms "Leaf Sensor Node", "Router Node", or/and "IoT Gateway Node", which can be modeled and tested on a virtual platform before physical deployment [6]. Herein, simulators like the Contiki-Cooja offer a virtual environment for testing and have the capability to simulate those functions and test their performance prior to physical implementation using hardware boards like Texas Instruments CC2538. Therefore, SDWSN and virtualization are significant components of the IoT architecture as they contribute to the flexible operation of the sensor network and hence offer a degree of freedom to the UAV flight path. For integrating the communication between configured WSN and UAV, simulators like CupCarbon can offer smooth communication testing between UAV and WSN. From the UAV path planning point of view, the algorithms of path planning design are crucial for creating adaptable flight paths that enable efficient data gathering by gateway nodes while minimizing energy use. Researchers have examined UAV path design considering factors like speed, communication duration, energy expenditure, and communication protocols [7]. However, factors related to the UAV path, which can be efficiently adjusted based on the reconfigurability of the nodes in the WSN, need to be considered to ensure smooth UAV-WSN communication. Contemporary advancements have introduced methodologies that augment decision-making capabilities in UAV flight path determination and network communication, particularly through the application of deep learning (DL) algorithms [8]. DL, a neural network-centric subset of machine learning, is characterized by its multilayered structure, enabling the processing of extensive datasets and the deciphering of complex interdependencies. Complementarily, reinforcement learning (RL) represents another machine learning strategy, distinguished by its dynamic adaptation through action modification in response to ongoing feedback, with the aim of optimizing a reward function. This technique is optimally suited for scenarios where pre-emptive rule formulation is impractical, as it facilitates experiential learning from errors and subsequent behavioral adjustments. Consequently, RL is adept at navigating multifaceted decision-making processes and addressing pragmatic challenges. Recent scholarly advancements have introduced Deep Reinforcement Learning (DRL), an integrative approach that combines DL and RL. This empowers systems to handle more complex tasks by autonomously extracting features from raw data. DRL has become a prominent method for strategically addressing UAV flight path planning challenges [9–12]. However, optimizing a UAV path using DRL in conjunction with SDWSN-UAV communication has not been considered in the literature. It is worth mentioning that the UAV path can be formulated based on data gathered from the WSN, where the multiple roles/functions of some nodes can enhance system performance by reducing energy consumption and delay, especially in scenarios requiring network restructuring.

This paper intends to introduce the "UAV Fuzzy Travel Path", a novel approach that leverages DRL algorithms for optimal UAV flight path planning to meet the requirements of various environmental applications. Herein, we also shed light on the integration between SDWSN and UAV communication, emphasizing an optimized flight path. This integration provides flexibility in network structure, allowing sensor nodes to be configured with one or multiple functions through software definitions to meet various network scenarios. For instance, the functional role of a node can be changed at the virtual level by implementing any of the three functions, such as reconfiguring a leaf node to act as a router node and vice versa. Consequently, the network's operational behavior can be adjusted to meet service

requirements. In our previous work [13], we detailed the configuration of SDWSN functions on a virtual platform using the Contiki-Cooja network simulator, where each function was coded in C language to perform tasks such as data processing and computation. The proposed UAV-assisted SDWSN model features nodes configured with a gateway function (GW) that receives data from dynamically formed group members via software definition. The node configuration using the Contiki Operating System (OS) mirrors the hardware configuration, utilizing hardware libraries. Concurrently, a UAV is deployed to collect data from GWs along a pre-planned trajectory within a fuzzy span. This approach will be utilized in this paper to investigate the design of an optimized UAV path using DRL that aligns with the data-gathering approach from SDWSN. Our primary aim is to highlight the benefits of integrating SDWSN for applications such as environmental monitoring and UAV path design using DRL.

The main contributions of this work are summarized as follows:

- Utilizing SDWSN Functions: We leverage the flexibility of software-defined wireless sensor networks (SDWSN) to configure network functions. Specifically, we formulate SDWSN to facilitate UAV communication with the wireless sensor network (WSN) and efficient data collection. Notably, nodes within the network can serve as gateway functions, enabling seamless data gathering by the UAV.
- Optimized UAV Path: We propose an optimized UAV flight path using Deep Reinforcement Learning (DRL) techniques, specifically based on Deep Q-Learning. This approach addresses the NP-hard nature of the formulated approximation problem. Moreover, integrating this learned UAV path with SDWSN enhances the overall system's efficiency and flexibility.
- Dual Objective: Our primary objectives are twofold: minimizing the total energy consumption of the UAV system during each data collection round and improving the packet delivery rate to the UAV receiver. Achieving this balance ensures an effective and sustainable system.

The remainder of this paper is structured as follows: Section 2 discusses the related work. Sections 3 and 4 present the system model and RL proposed algorithm. Section 5 evaluates the model. Finally, in Section 6, the conclusion of the work is discussed.

2. Related Work

Intelligent unmanned aerial vehicle (UAV) path planning is a critical prerequisite for a successful and real-time operation. As a result, optimizing UAV flight missions can be a solution for offering the fastest and most efficient path. With a focus on mission planning, researchers investigated several optimization techniques such as the multi-objective particle swarm optimization (PSO) algorithm. However, with the advancement of technology and recent advances in intelligent computation, virtualisation, softwarisation, artificial intelligence, and learning methods, prediction strategies and decision-making for the optimal UAV path are required. In our related work section, we explore various aspects to achieve an optimal UAV path for collecting and analysing data from Wireless Sensor Networks (WSN) situated in both sparse and dense fields. Specifically, we delve into areas related to clustering in WSN, Software-Defined Wireless Sensor Networks (SDWSN), virtualization, UAV path trajectories, and Deep Reinforcement Learning (DRL) algorithms for UAV path optimization.

Considering the significant flight altitude of UAVs, establishing direct connections between the UAV and all nodes in a wireless sensor network during the data-gathering process is inefficient in terms of energy. To address this challenge, grouping the nodes into groups/clusters is looked at based on different communication parameters allowing only the cluster heads (CH) to communicate with the drone [14]. These include factors like node location, the distance with respect to other nodes that are within the transmission range, communication protocol such as ZigBee, node energy consumption, and the size of the cluster. The clustering process involves two main steps: CH selection and cluster construction. In existing literature, two critical elements stand out for research in this area:

topology and data transmission routes wherein developing state-of-the-art topology and efficient data transmission routes within these clusters is essential. This ensures effective communication while minimizing energy consumption [15]. An efficient CH selection algorithm that optimally selects cluster heads is crucial. The CHs play a pivotal role in coordinating the data exchange between the UAV and the WSN nodes. Herein, in the context of ground-based WSN, data gathering often leads to faster battery depletion in sensors located near the sink due to concentrated data traffic. To mitigate this issue, the authors of [16,17] have explored the use of mobile sinks within the network. However, employing a mobile data ferry for data collection introduces additional challenges. Frequent link changes or failures can render the network topology unstable. To address this, Karunanithy Kalaivanan et al. [16] propose a mobile clustering data collection protocol that establishes a highly stable and reliable routing path for data transmission. In their approach, fuzzy logic-based CHs are selected, considering factors such as node speed, the number of neighboring nodes, and average connection time. Notably, this algorithm significantly enhances key performance metrics, including packet delivery ratio, throughput, energy consumption, and end-to-end delay.

In the context of employing UAV as a data ferry to assist in ground network data collection, the authors of [18] focus on clustering methods for the ground Sensor Nodes (SNs) in a single UAV scenario for environmental monitoring and data gathering. The energy model for the SNs considers the energy required for packet reporting and packet forwarding in the context of mesh networks, where each node can serve as a router for neighboring nodes toward the sink. UAV path planning involves a heuristic path, which is a local search algorithm used to solve the Traveling Salesman Problem (TSP) for visiting the WSN clusters. The optimization problem revolves around balancing UAV and sensor energy consumption versus the maximum number of network hops. Tarighi et al. [19] and Dan Popescu [20] explored the use of UAV as a mobile agent for collecting data from SNs. Their research considers the impact of clustering parameters on WSN, making their methods relevant for environmental monitoring applications. The findings suggest that the collaborative approach of UAV-assisted WSNs improves performance in both precision agriculture and ecological agriculture use cases. The challenge in WSN data gathering using UAVs as mobile sinks lies in achieving a balanced trade-off between the length of the UAV path and the energy consumption of SNs. In their work, Ebrahimi et al. [21] propose a novel projection-based method for CH selection and outline a forwarding tree structure. This method efficiently collects and transfers data from SNs and to the CHs, which are responsible for aggregating data from all other nodes within the same cluster. The aggregated data are then transmitted to the UAV acting as a mobile sink, ultimately delivering it to the base station. The UAV follows a trajectory that minimizes the total distance by passing through all the selected CHs once and arriving at the destination. The clustering problem used in the research is based on an updated version of *K-means* clustering [22]. In their proposed algorithm, after applying the K-means clustering method, the clusters are adjusted to achieve nearly equal-sized cells. Simulation results demonstrate that utilizing UAV for data collection at CHs instead of relying solely on the sink significantly reduces the number of relaying transmissions and substantially lowers energy consumption. However, it is worth noting that the approach described in this paper does not align with a realistic ground network formation scenario.

Addressing the challenge of data collection from multiple points involves employing dynamic clustering methods, where only the updated CHs communicate with the UAV. Therefore, developing a novel dynamic network clustering architecture that allows for network re-orchestration based on the UAV's path, and proposing an energy-efficient data communication strategy among ground SNs, are key research areas in UAV-assisted WSN data collection.

To evaluate the communication link before real-world implementation, various methods can be used to virtualize network behavior within the cloud. Efficient utilization of WSN deployments allows multiple applications to coexist on the same virtualized WSN [23].

Virtualization has been recognized as an effective method for conducting software-based simulation testing and adapting these simulations for physical networks. Software-driven virtualization provides a platform for testing and analyzing network scenarios, including dynamic behaviors. This parallel co-simulation, executed in the cloud backend, can streamline the network configuration process by eliminating the need for hardware during testing. For instance, the network simulator Contiki-Cooja is used as a virtualization platform for specific target hardware, such as the TI CC2538 Evaluation Module [24]. In our previous work [6], we explored the virtualization approach for vehicular networks, introducing the innovative concept of controlling WSN functions via the remote cloud. This approach was used to test potential re-orchestration scenarios within the vehicular network. By decoupling the functions of vehicular nodes and loading them as software components in the remote cloud, we enabled the re-organization and integration of the network. This allows for dynamic re-orchestration, where a node's functional role can be altered through software reconfigurations at the virtual level. Consequently, the operational behavior of the vehicular network can be adjusted to meet service requirements. Also, in our previous work [25], we proposed a point-by-point air-to-ground communication system that leverages a clustering structure to partition the ground network into small clusters of sensor nodes distributed over large areas. This structure supports efficient communication between sensor nodes and the UAV. The UAV's flight path is optimized by utilizing the potential dynamics in WSN orchestrations as suggested in our approach. We used the Contiki-Cooja simulator to establish communication between the UAV and the ground WSN, with the communication planning method based on the distance between the sensor nodes and the UAV, as indicated by RSSI measurements. Other researchers, Samir et al. [26], have introduced an innovative virtual design for UAV-assisted vehicular networks, which gathers and processes new information before sending it to the UAV. Current methods for UAV data capture during flights have not accounted for the dynamic topology and the adaptability of a virtual network structure along the UAV's flight path. Enhancing the UAV data collection method can be achieved by analyzing the UAV's travel route and integrating it with software-defined networking and network virtualization. One way to reduce UAV energy consumption during travel is by designing a UAV path framework using topological sorting and optimizing the UAV path shape to enhance energy efficiency and communication fairness. In our previous work [25], the concept of a fuzzy route is explored, where the proposed method considers geographical grouping with one or more nodes acting as ground data collection representatives. This results in more efficient and smoother UAV travel planning. The scheme's performance is evaluated based on UAV energy consumption while visiting points, showing that the proposed network communication analysis and smooth UAV flight path lead to a higher percentage of served SNs and improved ground SN power usage.

From the UAV path planning point of view, numerous studies have focused on data collection applications using drones within a distributed wireless network, where the drone visits each data point individually. To maximize the number of served SNs within the operational connectivity during a limited flight period, identifying the shortest route is crucial. The domain of UAV trajectory design encompasses a variety of evolving methods, including geometric-based path planning and heuristic trajectory planning. Therefore, choosing the appropriate UAV movement pattern can result in a more scalable and energy-efficient flight route design. One approach involves applying geometric constraints to the UAV path, such as circular, spiral, strip-based, and zig-zag patterns, to heuristically enhance UAV energy efficiency [27]. Additionally, various protocols and algorithms, such as iterative genetic algorithms (GA) and ant colony optimization (ACO), have been proposed to address the UAV path problem heuristically. In [28], solutions to the Traveling Salesman Problem (TSP) and the Pickup-and-Delivery Problem (PDP) are utilized for preliminary UAV trajectory design. The authors of [20] also address the challenge of optimizing wireless sensor network coverage using UAV platforms. They formulate this as an optimization problem, employing a heuristic approach based on the TSP to determine the optimal UAV

routing for data collection while minimizing the energy required for data transmission from the SNs. It has been demonstrated that the proposed TSP algorithm outperforms particle swarm optimization (PSO) across various flying heights, speeds, and network sizes. When using the TSP algorithm, the results for the tested configuration in this paper show less than half the total energy consumption at the gateway level. Conventional algorithms struggle to meet the growing coverage demands due to the high energy consumption of UAV movements and the negative impact of slow navigation on coverage performance. Recently, leveraging machine learning in wireless communication has emerged as a promising alternative for optimizing UAV navigation. In [29], a deep learning-based method for UAV navigation is proposed, which does not require sensing data to provide mapping information. However, this method faces challenges with slow convergence, making it unsuitable for real-time navigation scenarios. In [30], the authors aim to design an energy-efficient route for UAVs undertaking long-distance sensing tasks. They propose a DRL-based framework that employs convolutional neural networks (CNNs) for feature extraction and a deep Q-network (DQN) for decision-making. Research on the cooperation of computation tasks between UAV and WSN [31] suggests that tasks can be processed by nodes, UAV, and sink nodes. To manage randomly generated tasks, the UAV needs to dynamically adjust its trajectory to provide computing services for the nodes. Herein, DRL is used for UAV path optimization to enhance energy efficiency during the computation process. However, this research does not address the potential for node failure during this process or the need to restructure the network to re-establish the computation process. Other researchers [32,33] have also highlighted the minimized energy usage through an optimized path for UAV-assisted WSNs. Nonetheless, an approach that offers flexible restructuring of WSNs when needed has not been considered.

In summary, the existing literature on integrating UAV with WSN has not focused on creating a flexible network that can efficiently communicate with UAV, especially when node reconfigurability for various processing and computation tasks is required. Additionally, current algorithms for UAV flight path design have not adequately considered the adaptability of the path to meet the sensor networks' needs. While significant efforts have been made to develop energy-efficient UAV paths, there has been little attention given to optimizing paths that not only save UAV energy but also enable data collection from WSN with minimal downtime, even when structural changes necessitate reconfiguration. Therefore, we propose SDWSN where nodes can be configured with multiple functions to handle various tasks and facilitate efficient data collection by UAV. Furthermore, we propose an optimized path using DRL to ensure energy-efficient UAV flight rounds during data collection from flexible WSN.

3. System Model

In a Wireless Sensor Network (WSN), some sensor nodes can serve dual roles as both leaf nodes and cluster heads. These nodes play a critical role in optimizing network performance by efficiently managing communication and energy consumption. To enhance the efficiency of cluster heads, we propose designing a bounded fuzzy range based on Line-of-Sight (LoS) coverage. By considering the fuzzy region around each cluster head coverage area, we can optimize communication paths and minimize energy consumption. We aim to find the optimal path within the fuzzy region of the WSN. This involves selecting routes that minimize the propulsion of energy usage, reduce the distance between ground gateway nodes and the UAV, and stabilize ground network energy levels. The proposed path should balance these factors effectively. Hence, based on the proposed optimal path, adjustments can be made to select the optimal cluster heads out of potential cluster heads. This decision considers factors such as UAV versus ground network energy, communication quality, and overall network stability.

When evaluating different scenarios within the UAV's operational range, the proposed method's performance metrics include the percentage of sensor nodes served, energy consumption of the ground network, and average UAV energy consumption. There is a

notable trade-off between the UAV's propulsion energy and the ground network's energy costs. A heuristic smooth path design for the UAV can create an energy-efficient route from the UAV's standpoint but leads to an energy-inefficient data collection model for the ground sensor nodes [25]. On the other hand, the Bezier curve path design [25] for the UAV reduces its energy consumption per mission and decreases the percentage of served sensor nodes, while improving the ground network's energy efficiency. Therefore, it is necessary to formulate an optimization problem to find the optimal solution that simultaneously minimizes UAV propulsion energy usage and ground sensor nodes' energy consumption, while maximizing the UAV's communication bit rate.

To enhance the bit rate in air-to-ground communication for the UAV path (q(t)), it is necessary to develop a statistical model that accounts for the communication throughput between the UAV and SNs under LoS conditions. This paper does not consider the air-to-ground connectivity model for occasional link blockages due to NLoS links. Therefore, the total amount of information bits transferred to the UAV over the duration T is a function of the UAV's trajectory, as expressed in (1) [34]:

$$\overline{R}(q(t))_{air-to-Ground} = \int_0^T B \log_2\left(1 + \frac{\gamma_0}{H^2 + \|q(t)\|^2}\right) \quad (1)$$

where B stands for the channel bandwidth, and γ_0 is the reference received signal-to-noise ratio (SNR) at $d_0 = 1$ m. H is the altitude of the UAV while flying over the ground SNs. Also, the UAV energy consumption for a fixed-wing UAV considering variable velocity $v(t)$ and acceleration vectors $a(t)$ is expressed in (2) [34]:

$$\overline{E}(q(t)) = \int_0^T [c_1\|v(t)\|^3 + \frac{c_2}{\|v(t)\|}(1 + \frac{\|a(t)\|^2 - \frac{(a^T(t)v(t))^2}{\|v(t)\|^2}}{g^2})]dt + \frac{1}{2}m\left(\|v(t)\|^2 - \|v(0)\|^2\right) \quad (2)$$

in which c_1 and c_2 are two parameters related to the aircraft's weight, wing, air density, etc., g is the gravitational acceleration with nominal value 9.8 m/s^2, m is the mass of the UAV including all its payload.

The main parameters for designing an optimal UAV path within the fuzzy region considering the RL algorithm are presented in Figure 1. As highlighted in this figure, the parameters are organized into two main categories: parameters that define the UAV energy consumption model and parameters that define the UAV–Gateways communication throughput.

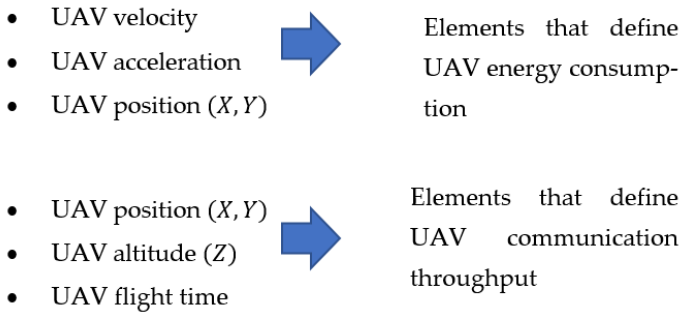

Figure 1. Relevant parameters for RL algorithm definition.

Based on the given UAV energy consumption Equations (2), the UAV energy propulsion usage is highly dependent on the instantaneous acceleration, velocity, and position factors. Since UAV acceleration and velocity are proportional to the amount of second and first derivatives of the instant position, the UAV energy consumption is dependent on the instant location q(t) ultimately.

On the other hand, based on the emerged equation of the UAV communication throughput in (1), it would be the same case with communication throughput. This means

that the UAV communication throughput is highly dependent on the instant location of the UAV. Hence both UAV energy consumption and communication throughput are dependent on the UAV position at each time slot by two various equations.

To define the approximation optimization problem, a cost function of multiple parameters is required to be specified as:

$$\begin{array}{c} \text{Minimise} \\ \text{s.t. List of Contraints} \end{array} \quad \begin{array}{c} q(t) \\ \end{array} \quad \frac{\overline{E}(q(t))_{UAV-propulsion}}{\overline{R}(q(t))_{air-to-Ground}} \tag{3}$$

where $q(t)$ is the UAV path design. $\overline{E}(q(t))_{UAV-propulsion}$ is the UAV propulsion power consumption obtained from (2), $\overline{R}(q(t))_{air-to-Ground}$ is the communication throughput obtained from (1). According to our assumptions, ground network energy consumption is fixed and not dependent on the location of the UAV, while air-to-ground energy consumption is highly dependent on the distance between the UAV and SNs and as a result, the locations of the UAV. Hence, the impact of the ground SNs' energy consumption while communicating with each other on the ground is disregarded in our optimization problem formulation by only taking the UAV–Ground communication energy usage ($Cost_{air-to-Ground}$) into account.

4. Proposed RL Algorithm Model

The model has been proposed based on using Reinforcement Learning (RL) to improve the cost functions (3) by designing an optimal policy. RL is a branch of the machine learning paradigm that deals with the multi-state decision process of a software agent (UAV in our case) while interacting with an environment. The Markov Decision Process (MDP) model is the foundation of Reinforcement Learning (RL) algorithms, particularly in solving path-planning problems. Generally, an MDP for UAV path planning consists of four essential components:

S: The set of all possible states

A: The set of all possible actions

P: The transition probability, which determines the likelihood of moving from one state to another

R: The reward function, which evaluates the performance of an action in each state.

These elements are presented in a typical MDP model, as shown in Figure 2. At a given moment, the agent is in state "s". The decision-making mechanism can select any available action "a" while in state "s". Upon executing action "a", the agent transitions to a new state "s" and receives a reward "R(s,a)" for its action. The reward "R" is a function of the current state "s" and the chosen action "a". The reward is defined as:

$$R(s,a) = \sum_{a \epsilon A} R(s,a) p(s,a) \tag{4}$$

where $Q(s,\ a)$ is defined as a reward value when the agent takes action a at state s. The values of $Q(s,\ a)$ can be generated using the Bellman equations.

$$Q(s,a) = R(s,a) + \ \gamma \sum_{a \epsilon A} p(s,a) \sum_{a\ \prime \epsilon A} Q(\ s\ \prime,\ a\ \prime) \tag{5}$$

Based on the proposed model, we assume that the system has multiple states S (UAV's location, velocity), where at each state, the agent has a finite number of actions a (i.e., moving to neighboring locations within the fuzzy path or adjusting a different velocity for the UAV within the bounded UAV velocity) to choose from. After choosing an action, the agent (UAV) receives a reward $r(s_t, a_t)$, and moves to the next state s_{t+1}. The goal of RL is to learn the best route and find an optimal policy π^* that maximizes the cumulative sum of all rewards. The proposed policy $\pi^* = \{a_1, a_2, \ldots, a_T\}$ defines multiple optimal actions in the whole path that should offer maximum rewards. There are different methods to obtain the RL-optimized path (best policy). One of the most practical ones is utilizing

a table called the Q-Learning method, which is based on updating a table in each epoch (time slot) called Q-table.

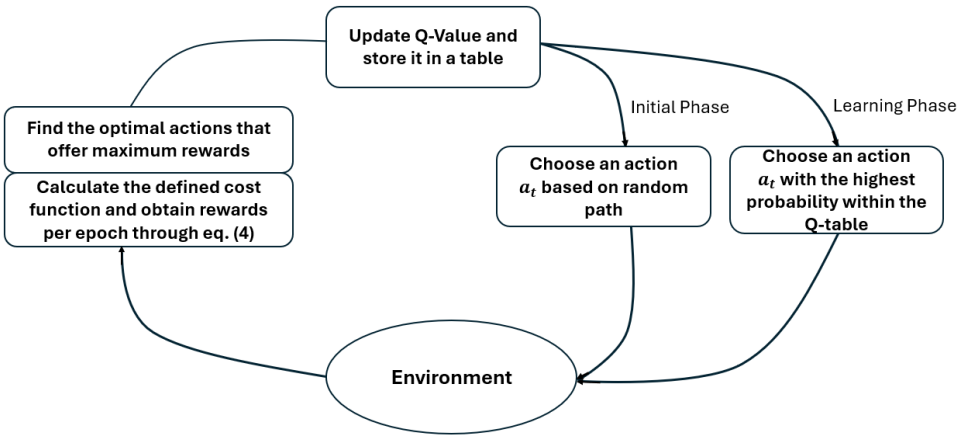

Figure 2. A defined RL algorithm in solving the optimal problem.

The proposed cost function given in Equation (3) requires a subset of policies to minimize the UAV energy consumption while maximizing the communication throughput. Agent (UAV) states S at each epoch considering the bounded fuzzy regions are UAV position (X, Y) and UAV velocity.

To define the environment based on the proposed RL model, the fuzzy boundary is required to be divided into small cells called RL cells in which each cell should be covered by the communication range of the ground gateway. As shown in Figure 3 left, the position of each center can be adjusted based on the predefined gateway position. In the learning policy, the initial path will be designed considering the red points, which are assumed to be within the communication range of gateways (but essentially, they are not receiving the optimal throughput) to train the UAV and get the information from the environment.

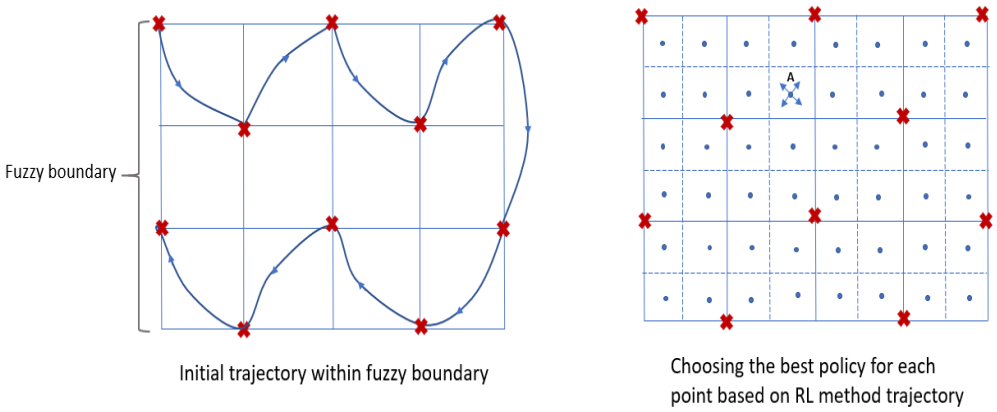

Figure 3. Defined RL model for UAV path design.

In the second phase (training phase), as seen in the right figure, each cell is divided into 4 cells (can be divided into more cells, which may impose more computation and complexity on the algorithm). Once the agent (UAV) reaches each location with different states, the UAV should decide the best action (either moving to left, right, north, or south and declining or increasing the velocity within the range) to gain the best rewards. Note that the chosen actions through the whole path should maximize the amount of cumulative reward for the whole journey, which will provide the optimal policy for the UAV path. Choosing the best action in each epoch (t) is dependent on the Q-values in the Q-table,

which is updated in each epoch (t), by considering Equation (6) which is highly dependent on the amount of all future rewards. γ is a discount factor and α is learning rate.

$$Q^{new}(s_t, a_t) = (1 - \alpha)Q(s_t, a_t) + \alpha.(R_{t+1} + \gamma.\max Q(s_{t+1}, a)) \quad (6)$$

5. Simulation Outcomes

Our simulation is carried out via MATLAB presuming an initial path within the UAV fuzzy range. The initial path is defined based on Figure 4 and after multiple episodes, the UAV path will converge into an optimal path considering the proposed cost function (3).

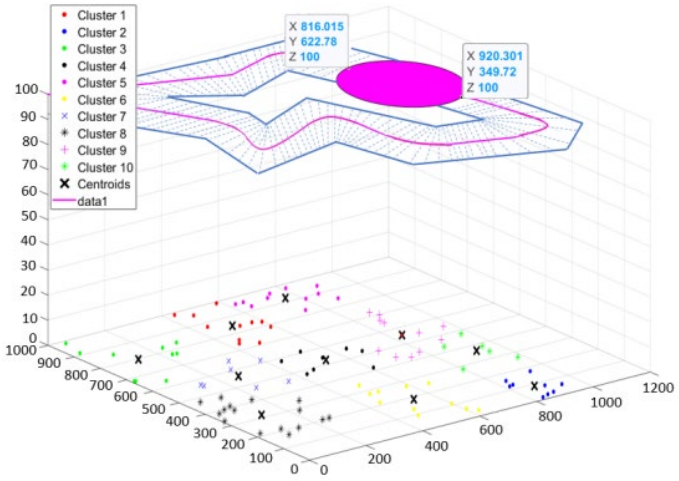

Figure 4. The proposed fuzzy vs. initial path for UAV path design.

The intention here is that the agent (UAV) visits the optimal data points as a part of its designed path to optimize the communication throughput and energy consumption by introducing the optimal location and velocity of the UAV within each state. The agent begins its journey from coordinate (0,0). The rewards are considered +10 for a constructive action while, for a destructive action, the reward is -1. The learning policy using Q-learning is shown in Figure 3.

To begin with, as outlined in Figure 4, a UAV fuzzy range is designed for data collection over 100 ground sensor nodes via gateways. The connectivity window is also mapped over the UAV's adjusted path within the fuzzy route shown in the pink circle. The designed fuzzy route is represented by the hatched area. The initial UAV path is applied via the defined RL model in the UAV fuzzy range shown in the pick route.

Once the Q-table is sufficiently trained, the optimal path can be extracted by choosing the action with the highest Q-value for the current state. The initial and destination positions of UAVs are the same. The UAV is set to fly at an altitude of 100 m, which aligns with real-world conditions. At this height, the UAV can navigate smoothly above the canopies for data-gathering purposes, avoiding any obstructions caused by their height. The UAV velocity range is assumed to be between 5–40 m/s and the size of the grid considered is 1000×1000 m^2. The simulation parameters are outlined in Table 1.

The proposed RL algorithm is highly dependent on the learning parameters choices. Learning rate α determines the extent to which the model adjusts its knowledge with each new data point. We have presumed the learning rate $\alpha = 0.1$ based on our simulations. The other factor is the discount factor γ which determines the significance of future rewards. We have presumed discount factor $\gamma = 0.9$ based on our simulations which ensure that the agent considers long-term rewards while making decisions. The exploration rate is another factor in defining the RL algorithm. The optimal choice for this parameter varies during the experiment, started with a high exploration rate ($\epsilon = 1$) and gradually decreased over time. This strategy, known as epsilon-greedy, facilitates the agent to explore the environment initially and then exploit the learned knowledge as it gains more experience.

Table 1. Simulation parameters for the proposed UAV path design model.

Learning rate α	0.1
Discount factor γ	0.9
Number of episodes	300
Exploration rate ϵ	1
Rewards	+10, −1
Bandwidth of channel (MHz)	20
Packet size (bytes)	512
Transmission range (m)	100
UAV velocity (m/s)	(5–40)
Number of nodes	100
Time step size δ	0.5
Altitude of the UAV (m)	100
C1 Constant	9.26×10^{-4}
C2 Constant	2250

Following the training of the initial path via the proposed Q-Learning RL algorithm, the emerged path has the optimal energy consumption and communication bit rate on UAV ground connectivity. Following 300 episodes of experimentation, the resulting path achieves acceptable performance compared to the initial path design by reaching the episodes to the end. This episode's number choice can provide enough time for the agent to explore the environment and learn from a variety of states and actions. Additionally, it helps the learning algorithm converge to a stable policy relatively quickly, while extending the training episodes might only result in minor improvements at a much higher computational cost.

Figure 5 illustrates the performance of our proposed optimized model, which enhances our initial path performance over the proposed Q-Learning RL. In this figure, the energy consumption and communication bit rate for UAV-Gateway connectivity are computed using Equations (1) and (2) for each episode, allowing us to assess the system's performance across multiple episodes.

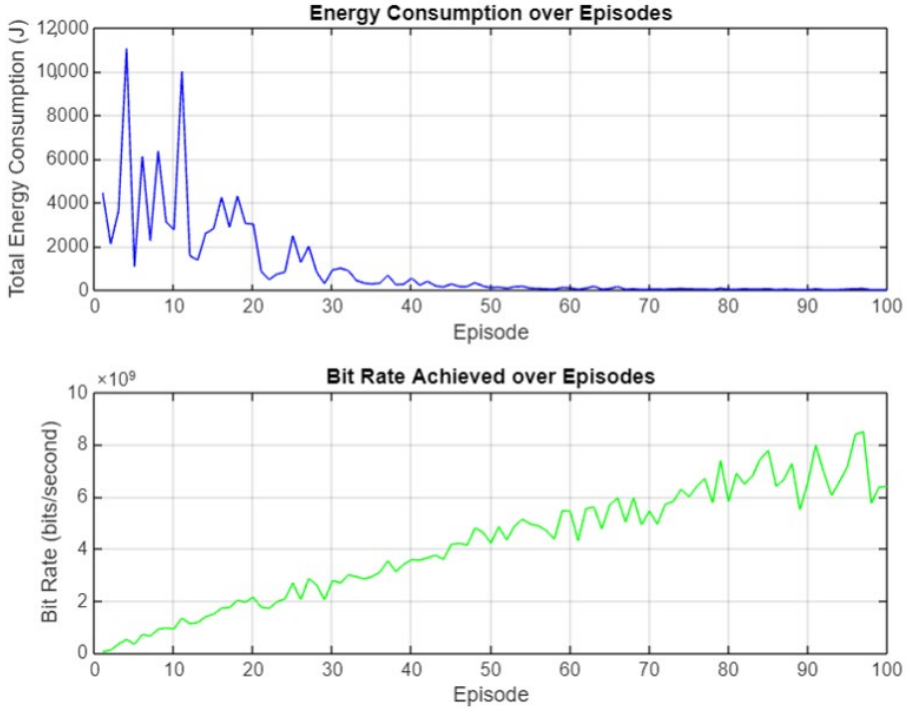

Figure 5. UAV energy consumption and communication bit rate performances over multiple episodes.

The reward function is determined for each epoch as follows: A reward is assigned based on the UAV's connectivity throughput to the gateway, which correlates with its distance to the gateways. The closer the UAV is to the gateway, the better the throughput and the higher the reward. Conversely, moving away from the gateway results in a penalty or negative reward. The same reward function is utilized to assess the UAV's energy consumption, thereby determining the proposed cost function for each epoch accordingly.

As shown in the upper part of Figure 5, energy consumption decreases as the number of episodes increases, while the bit rate declines with the rise in episodes. This indicates that as the number of episodes increases, the likelihood of finding the optimal path improves, although it may also lead to greater complexity and longer runtime for the simulation. Also, with more episodes, the agent has more chances to explore the environment and learn from its actions, which enhances its decision-making over time. Furthermore, the agent can more effectively balance exploring new actions and exploiting known ones, leading to better overall performance. As a result, the proposed cost function in Equation (3) improves with each episode, leading to lower energy consumption and higher bit rates.

The proposed method is compared with the methods in [14,26], including the heuristic fuzzy algorithm without RL and the SCA UAV path approximation model. The results show better performance in the success rate of served sensor nodes with the RL algorithm. This comparison uses various network scalability metrics, keeping other parameters consistent. According to Figure 6, as the number of dispersed nodes increases, the proposed RL-based method shows a significantly larger gap between the SCA and heuristic algorithms. Additionally, the complexity of the proposed communication network is much lower than that of the SCA algorithm [26], which uses complex optimization algorithms for UAV trajectory planning.

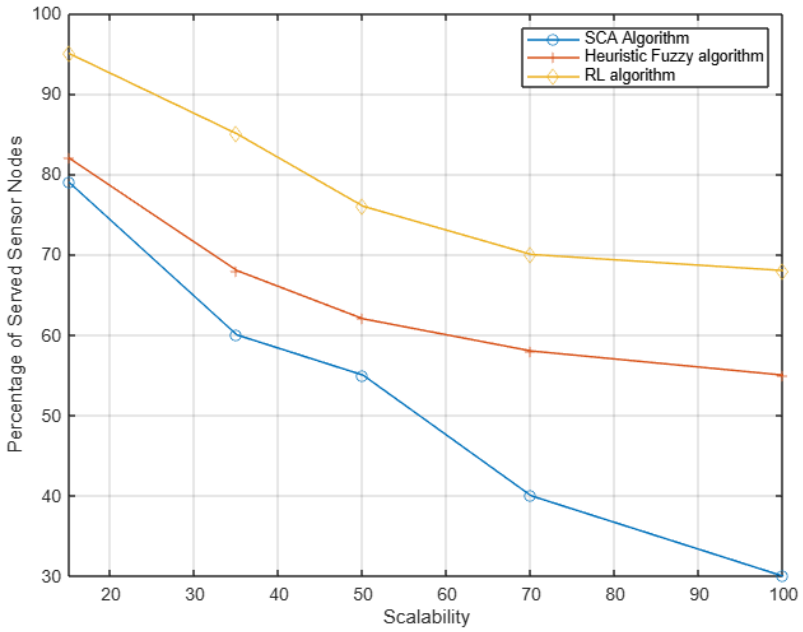

Figure 6. The comparison between the proposed RL data gathering method and heuristic fuzzy algorithm [14] and SCA algorithm [26] on the percentage of served sensor nodes versus scalability.

6. Conclusions

Using UAVs in Software Defined Wireless Sensor Networks (SDWSN) and the Internet of Things (IoT) offers several key benefits including enhanced data collection, energy efficiency, improved network performance, and real-time monitoring. The integration of UAVs with SDWSN and IoT frameworks, facilitated by advanced DRL algorithms, presents a transformative approach to data collection and network performance enhancement. The proposed UAV Fuzzy Travel Path model not only optimizes UAV trajectories to reduce

energy consumption but also significantly improves communication throughput. This innovative solution addresses the complexities of UAV path planning and communication performance, paving the way for more efficient and reliable UAV-assisted data-gathering systems. In the future, we will consider the design of the UAV path with obstacles for scenarios like disaster management. This approach can also include dynamic scenarios for UAV-assisted WSNs, such as mobile nodes and the potential for network restructuring. Consequently, this will necessitate an optimized UAV path that adapts to the network's dynamics.

Author Contributions: Conceptualization, P.A.K.; methodology, P.A.K. and D.Z.A.-H.; software, P.A.K.; validation, P.A.K. and D.Z.A.-H.; investigation, P.A.K. and D.Z.A.-H.; writing—original draft preparation, P.A.K. and D.Z.A.-H.; writing—review and editing, P.A.K., D.Z.A.-H. and P.H.J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

1. Fascista, A. Toward integrated large-scale environmental monitoring using WSN/UAV/Crowdsensing: A review of applications, signal processing, and future perspectives. *Sensors* **2022**, *22*, 1824. [CrossRef]
2. Mohsan, S.A.H.; Khan, M.A.; Noor, F.; Ullah, I.; Alsharif, M.H. Towards the unmanned aerial vehicles (UAVs): A comprehensive review. *Drones* **2022**, *6*, 147. [CrossRef]
3. Mowla, M.N.; Mowla, N.; Shah, A.S.; Rabie, K.; Shongwe, T. Internet of things and wireless sensor networks for smart agriculture applications-a survey. *IEEE Access* **2023**, *11*, 145813–145852. [CrossRef]
4. Jurado-Lasso, F.F.; Marchegiani, L.; Jurado, J.F.; Abu-Mahfouz, A.M.; Fafoutis, X. A survey on machine learning software-defined wireless sensor networks (ml-SDWSNS): Current status and major challenges. *IEEE Access* **2022**, *10*, 23560–23592. [CrossRef]
5. Moubayed, A.; Shami, A.J.A.P.A. Softwarization, virtualization, machine learning for intelligent effective v2x communications. *IEEE Access* **2020**, *14*, 156–173.
6. Al-Hamid, D.Z.; Al-Anbuky, A. Vehicular Networks Dynamic Grouping and Re-Orchestration Scenarios. *Information* **2023**, *14*, 32. [CrossRef]
7. Alsuhli, G.; Fahim, A.; Gadallah, Y. A survey on the role of UAVs in the communication process: A technological perspective. *Comput. Commun.* **2022**, *194*, 86–123.
8. Rezaee, M.R.; Hamid, N.A.W.A.; Hussin, M.; Zukarnain, Z.A. Comprehensive Review of Drones Collision Avoidance Schemes: Challenges and Open Issues. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 6397–6426. [CrossRef]
9. Azar, A.T.; Koubaa, A.; Ali Mohamed, N.; Ibrahim, H.A.; Ibrahim, Z.F.; Kazim, M.; Casalino, G. Drone deep reinforcement learning: A review. *Electronics* **2021**, *10*, 999. [CrossRef]
10. Mannan, A.; Obaidat, M.S.; Mahmood, K.; Ahmad, A.; Ahmad, R. Classical versus reinforcement learning algorithms for unmanned aerial vehicle network communication and coverage path planning: A systematic literature review. *Int. J. Commun. Syst.* **2023**, *36*, e5423. [CrossRef]
11. Xi, M.; Dai, H.; He, J.; Li, W.; Wen, J.; Xiao, S.; Yang, J. A lightweight reinforcement learning-based real-time path planning method for unmanned aerial vehicles. *IEEE Internet Things J.* **2024**, *11*, 21061–21071. [CrossRef]
12. Luo, X.; Chen, C.; Zeng, C.; Li, C.; Xu, J.; Gong, S. Deep reinforcement learning for joint trajectory planning, transmission scheduling, and access control in UAV-assisted wireless sensor networks. *Sensors* **2023**, *23*, 4691. [CrossRef]
13. Al-Hamid, D.Z.; Karegar, P.A.; Chong, P.H.J. Modelling and Implementation Tools for SDWSN Smart Applications. In *2023 28th Asia Pacific Conference on Communications (APCC)*; IEEE: Piscataway, NJ, USA, 2023; pp. 81–86.
14. Karegar, P.A.; Al-Hamid, D.Z.; Chong, P.H.J. UAV-enabled software defined data collection from an adaptive WSN. In *Wireless Networks*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 1–22.
15. Wajgi, D.W.; Tembhurne, J.V. Localization in wireless sensor networks and wireless multimedia sensor networks using clustering techniques. *Multimed. Tools Appl.* **2024**, *83*, 6829–6879. [CrossRef]
16. Kalaivanan, K.; Bhanumathi, V. Reliable location aware and Cluster-Tap Root based data collection protocol for large scale wireless sensor networks. *J. Netw. Comput. Appl.* **2018**, *118*, 83–101. [CrossRef]
17. Velmani, R.; Kaarthick, B. An Efficient Cluster-Tree Based Data Collection Scheme for Large Mobile Wireless Sensor Networks. *IEEE Sens.* **2015**, *15*, 2377–2390. [CrossRef]
18. Tazibt, C.Y.; Bekhti, M.; Djamah, T.; Achir, N.; Boussetta, K. Wireless sensor network clustering for UAV-based data gathering. In Proceedings of the Wireless Days, Porto, Portugal, 29–31 March 2017; pp. 245–247.

19. Tarighi, R.; Farajzadeh, K.; Hematkhah, H. Prolong network lifetime and improve efficiency in WSN-UAV systems using new clustering parameters and CSMA modification. *Int. J. Commun. Syst.* **2020**, *33*, 4324. [CrossRef]
20. Popescu, D.; Stoican, F.; Stamatescu, G.; Ichim, L.; Dragana, C. Advanced UAV–WSN System for Intelligent Monitoring in Precision Agriculture. *Sensors* **2020**, *20*, 817. [CrossRef]
21. Ebrahimi, D.; Sharafeddine, S.; Ho, P.-H.; Assi, C. UAV-Aided ProjectionBased Compressive Data Gathering in Wireless Sensor Networks. *IEEE Internet Things* **2019**, *6*, 1893–1905. [CrossRef]
22. Sinaga, K.P.; Yang, M.S. Unsupervised K-means clustering algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [CrossRef]
23. Khan, I.; Belqasmi, F.; Glitho, R.; Crespi, N.; Morrow, M.; Polakos, P.A. Wireless sensor network virtualization: A survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 553–576. [CrossRef]
24. Al-Hamid, D.Z.; Karegar, P.A.; Chong, P.H.J. A novel SDWSN-based testbed for IoT smart applications. *Future Internet* **2023**, *15*, 291. [CrossRef]
25. Karegar, P.A.; Al-Anbuky, A. UAV-assisted data gathering from a sparse wireless sensor adaptive networks. *Wirel. Netw.* **2022**, *29*, 1367–1384. [CrossRef]
26. Samir, M.; Assi, C.; Sharafeddine, S.; Ebrahimi, D.; Ghrayeb, A. Age of Information Aware Trajectory Planning of UAVs in Intelligent Transportation Systems: A Deep Learning Approach. *IEEE Trans. Veh. Technol.* **2020**, *69*, 12382–12395. [CrossRef]
27. Wu, Q.; Zeng, Y.; Zhang, R. Joint Trajectory and Communication Design for Multi-UAV Enabled Wireless Networks. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 2109–2121. [CrossRef]
28. Li, J.; Ren, B. Joint Optimization on Trajectory, Altitude, Velocity, and Link Scheduling for Minimum Mission Time in UAV-Aided Data Collection. *IEEE Internet Things* **2020**, *7*, 1464–1475. [CrossRef]
29. Wang, C.; Wang, J.; Shen, Y.; Zhang, X. Autonomous Navigation of UAVs in Large-Scale Complex Environments: A Deep Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.* **2019**, *68*, 2124–2136. [CrossRef]
30. Zhang, B.; Liu, C.H.; Tang, J.; Xu, Z.; Ma, J.; Wang, W. Learning-based energy-efficient data collection by unmanned vehicles in smart cities. *IEEE Trans. Ind. Informat.* **2018**, *14*, 1666–1676. [CrossRef]
31. Guo, Z.; Chen, H.; Li, S. Deep Reinforcement Learning-Based UAV Path Planning for Energy-Efficient Multitier Cooperative Computing in Wireless Sensor Networks. *J. Sens.* **2023**, *2023*, 2804943. [CrossRef]
32. Beishenalieva, A.; Yoo, S.J. UAV Path Planning for Data Gathering in Wireless Sensor Networks: Spatial and Temporal Substate-Based Q-Learning. *IEEE Internet Things J.* **2023**, *11*, 9572–9586. [CrossRef]
33. Sun, A.; Sun, C.; Du, J.; Wei, D. Optimizing Energy-Efficiency in UAV-assisted Wireless Sensor Networks with Reinforcement Learning PPO2 Algorithm. *IEEE Sens. J.* **2023**, *23*, 29705–29721. [CrossRef]
34. Zeng, Y.; Zhang, R. Energy-efficient UAV communication with trajectory optimization. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 3747–3760. [CrossRef]