


## Article

# Deep-Reinforcement-Learning-Based Joint Energy Replenishment and Data Collection Scheme for WRSN

Jishan Li <sup>1</sup>, Zhichao Deng <sup>1</sup>, Yong Feng <sup>1,\*</sup> and Nianbo Liu <sup>2</sup>

<sup>1</sup> Yunnan Key Laboratory of Computer Technology Applications, Kunming University of Science and Technology, Kunming 650500, China; 20212204267@stu.kust.edu.cn (J.L.); dengzhichao@stu.kust.edu.cn (Z.D.)

<sup>2</sup> School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; liunb@uestc.edu.cn

\* Correspondence: fybraver@163.com

**Abstract:** With the emergence of wireless rechargeable sensor networks (WRSNs), the possibility of wirelessly recharging nodes using mobile charging vehicles (MCVs) has become a reality. However, existing approaches overlook the effective integration of node energy replenishment and mobile data collection processes. In this paper, we propose a joint energy replenishment and data collection scheme (D-JERDG) for WRSNs based on deep reinforcement learning. By capitalizing on the high mobility of unmanned aerial vehicles (UAVs), D-JERDG enables continuous visits to the cluster head nodes in each cluster, facilitating data collection and range-based charging. First, D-JERDG utilizes the K-means algorithm to partition the network into multiple clusters, and a cluster head selection algorithm is proposed based on an improved dynamic routing protocol, which elects cluster head nodes based on the remaining energy and geographical location of the cluster member nodes. Afterward, the simulated annealing (SA) algorithm determines the shortest flight path. Subsequently, the DRL model multiobjective deep deterministic policy gradient (MODDPG) is employed to control and optimize the UAV instantaneous heading and speed, effectively planning UAV hover points. By redesigning the reward function, joint optimization of multiple objectives such as node death rate, UAV throughput, and average flight energy consumption is achieved. Extensive simulation results show that the proposed D-JERDG achieves joint optimization of multiple objectives and exhibits significant advantages over the baseline in terms of throughput, time utilization, and charging cost, among other indicators.

**Keywords:** wireless rechargeable sensor networks; unmanned aerial vehicles; deep reinforcement learning; route protocol



**Citation:** Li, J.; Deng, Z.; Feng, Y.; Liu, N. Deep-Reinforcement-Learning-Based Joint Energy Replenishment and Data Collection Scheme for WRSN. *Sensors* **2024**, *24*, 2386. <https://doi.org/10.3390/s24082386>

Academic Editors: Rajan Shankaran, Wei Ni, Xiaojing Chen and Bochun Wu

Received: 4 March 2024

Revised: 1 April 2024

Accepted: 2 April 2024

Published: 9 April 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

Benefiting from the advancements in wireless power transfer (WPT) technology [1,2], wireless rechargeable sensor networks (WRSNs) equipped with wireless charging systems have emerged as an effective solution for addressing node energy constraints [3,4]. Unlike traditional approaches [5–7], WRSN fundamentally overcomes the predicament of nodes relying solely on battery power. It provides a promising solution for the sustainable energy replenishment of nodes [8–10]. WRSN normally includes one or more mobile charging vehicles (MCVs) and a base station (BS) for the battery replacement of MCVs. MCV can move autonomously and is equipped with a wireless charging device to replenish energy for nodes by wireless charging. However, in remote and harsh environments, e.g., farmlands, forests, or disaster areas, it becomes challenging for ground vehicles to enter designated areas and perform close-range wireless charging [11–13]. Utilizing UAVs as aerial mobile vehicles to provide remote services to ground devices is considered to bring significant benefits to WRSN [14]. UAV possesses excellent maneuverability, coverage capabilities, and relatively low operating costs. It can perform various tasks within the network coverage,

such as routing, communication, data collection, and energy supply [15–18]. In UAV-assisted wireless rechargeable sensor networks (UAV-WRSNs), UAVs can serve both as energy transmitters and data transceivers. By continuously accessing and exchanging data with nodes within their communication range, they can forward the data to the BS. This eliminates the energy consumption caused by multihop transmissions between nodes. Additionally, UAVs provide range-based charging to ensure simultaneous data collection and energy replenishment.

Although UAV offers new possibilities for the future development of WRSN, more research is needed to consider the unified process of node energy replenishment and data collection in UAV-WRSN. There are still unresolved issues in UAV-WRSN. For example, in a one-on-one node service scenario, UAV must hover multiple times, increasing hover time and energy consumption, leading to subsequent node data overflow or energy depletion. Moreover, target node selection is crucial to enable UAV to cover more nodes with each hover. In practical environments, there is a significant difference in energy consumption among nodes, and some nodes may run out of energy earlier due to heavy loads. Therefore, balancing node energy consumption and solving the problem of energy imbalance among nodes are equally important. Based on the above issues, this paper proposes D-JERDG, which integrates node joint energy replenishment and data collection using deep reinforcement learning.

The main contributions of this study can be summarized as follows:

- First, we consider deploying UAV in delay-tolerant WRSN and combining wireless power transfer (WPT) with wireless information transfer (WIT) technologies. To achieve the unified process of sensor node energy replenishment and mobile data collection, we propose a deep-reinforcement-learning-based method called D-JERDG for UAV-WRSN.
- We introduce a cluster head selection algorithm based on an improved dynamic routing protocol to minimize the number of UAVs hovering and balance node energy consumption. Based on the obtained cluster head visiting sequence, we employ the simulated annealing algorithm [19] to approximate the optimal solution for the traveling salesman problem (TSP), thereby reducing the UAV flight distance. We then employ the DRL model MODDPG [20] and design a multiobjective optimized reward function to control the UAV instantaneous speed and heading, aiming to minimize the node death rate and the UAV's average energy consumption.
- Simulation results are conducted to evaluate the feasibility and effectiveness of D-JERDG. The results demonstrate that D-JERDG outperforms existing algorithms in node death rate, time utilization, and charging cost.

The remaining parts of this paper are as follows: Section 2 introduces related work, Section 3 describes the system model and problem formulation of UAV-WRSN, Section 4 presents the cluster head selection algorithm based on an improved dynamic routing protocol and the MODDPG algorithm in D-JERDG, Section 5 validates the effectiveness and feasibility of D-JERDG through comparative experiments, and Section 6 summarizes the paper and discusses future research directions.

## 2. Related Work

In this section, we provide a brief overview of the existing work in three relevant domains: cluster-based networks [21–28], traditional algorithm-based UAV trajectory planning [29–37], and DRL-based UAV trajectory planning [38–43].

In traditional WRSN, the one-to-one node charging mode can often lead to inefficient movement of MCV, resulting in wastage of energy. Moreover, the convergence nodes bear the primary data transmission tasks, leading to faster energy depletion and premature failure of nodes, thus affecting the overall network lifetime. To balance the energy consumption of nodes, the charging method often involves dividing the nodes into several clusters using a hierarchical clustering approach. Within each cluster, a few nodes closest to the base station (BS) are selected as cluster heads, and communication links are established between

the cluster heads and the BS, such as LEACH [21], K-means [22], Hausdorff [23], and HEED [24]. These methods typically use a rotating cluster head approach to reduce energy consumption and extend network lifetime. For example, in [25–28], the authors studied the division of the network into multiple clusters using clustering algorithms. They employed a mobile charger (MC) to periodically replenish the energy of anchor nodes in each cluster according to the generated shortest visiting path. Wu et al. designed a joint solution that integrates both aspects and proposed a heuristic-based MC scheduling scheme to maximize the charging efficiency of the MC while minimizing its energy consumption [25]. Li et al. proposed an energy-efficiency-oriented heterogeneous paradigm (EEHP), which is a routing protocol based on multihop data transmission. It reduces the energy consumption for data transmission by employing multihop data transfer and shortens the charging distance for the MC [26]. Han et al. used the K-means clustering algorithm to divide the network into multiple clusters and proposed a semi-Markov model to update anchor nodes. They deployed two MCs that periodically moved in opposite directions to visit anchor nodes, charging the sensor nodes (SNs) within the charging range and collecting data from the cluster heads [27]. Zhao et al. focused on achieving joint optimization of efficient charging and data collection in randomly deployed WRSN. They periodically selected anchor points and arranged MCV to visit these locations and charge the nodes sequentially [28].

For sustainable monitoring, WRSN can be applied in remote and resource-limited areas, such as rural farmlands, forests, or disaster zones [29]. In such harsh environments, UAVs can be used as auxiliary aerial base stations to efficiently collect data and replenish node energy, which significantly benefits WRSN [30]. For example, in [31–34], researchers studied scheduling and designing UAV trajectories to improve the system charging efficiency. Xu et al. studied a new UAV-enabled wireless power transfer system to maximize the total energy received by considering the optimization of the trajectory of UAV under the maximum velocity constraint [31]. Liu et al. proposed UAV-WRSN and solved the subproblems of UAV scheduling and trajectory optimization separately. They aimed to minimize the number of UAV hover points, SN with repeated coverage, and UAV flight distance [32]. Wu et al. investigated the trajectory optimization problem for UAV in UAV-WRSN. They decomposed the problem of maximizing energy efficiency into an integer programming problem and a nonconvex optimization problem, effectively reducing the UAV flight distance and algorithm complexity and maximizing the energy utilization efficiency of the UAV [33]. Zhao et al. proposed an improved ant colony algorithm to plan UAV flight trajectories, achieving shorter flight paths and network lifetimes [34]. In [35–37], researchers explored using UAVs as data collectors and mobile chargers, simultaneously providing data collection and energy transfer to the nodes. Baek et al. considered node energy consumption and replenishment to maximize the WRSN lifetime. They optimized the UAV hover locations and durations to maximize the remaining energy of the nodes [35]. Lin et al. studied the collaboration between rechargeable sensors and UAVs to accomplish regular coverage tasks. They introduced a new concept of coverage called periodic area coverage, aiming to maximize the energy efficiency of the UAV [36]. Hu et al. formulated a nonconvex optimization problem to minimize the average age of information (AoI). They divided it into a time allocation problem and UAV trajectory optimization problem and solved it optimally using dynamic programming and ant colony algorithms [37].

DRL has proven to be an effective solution for decision-making problems on sequential data. It has been widely applied in various fields and has achieved notable results. In the context of UAV-WRSN, several studies are worth mentioning: Bouhamed et al. employed two reinforcement learning (RL) methods, namely, deep deterministic policy gradient (DDPG) and Q-learning (QL), to train UAV for data collection tasks. DDPG was utilized to optimize UAV flight trajectories in environments with obstacle constraints, while QL was used to determine the order of visiting nodes [38]. Liu et al. proposed a UAV path planning based on reinforcement learning, which enables the UAV to respond to the position change of the cluster nodes, reduces the flight distance and the energy consumption, and increases the time utilization ratio [39]. Li et al. proposed a flight resource allocation framework

based on the DDPG algorithm. They optimized the UAV instantaneous heading, speed, and selection of target nodes. They utilized a state representation layer based on long short-term memory (LSTM) to predict network dynamics and minimize data packet loss [40]. Shan et al. presented a DRL-based trajectory planning scheme for multi-UAVs in WRSN. They established a network model for multi-UAV path planning. They optimized the network model using an improved hybrid energy-efficient distributed (HEED) clustering algorithm to obtain the optimal charging path [41]. Liu et al. considered WRSN assisted by UAVs and vehicles. In their study, UAVs served as mobile chargers to replenish energy for nodes, while mobile vehicles acted as mobile base stations to replace UAV batteries. The authors utilized a multiobjective deep Q-network (DQN) algorithm to minimize sensor downtime and optimize UAV energy consumption [42]. Wang et al. proposed a dynamic spatiotemporal charging scheduling scheme based on deep reinforcement learning, given the discrete charging sequence planning and continuous charging duration adjustment in mobile charging scheduling, to improve the charging performance while avoiding the power death of nodes [43]. Table 1 shows how the related work differs from our scheme.

**Table 1.** Comparison of related works with D-JERDG.

References	Optimization Objective	Optimization Scheme	Learning-Based	Charging Mode
[20]	Minimize UAV energy consumption while maximizing UAV throughput	DRL	Yes	One-to-multiple
[32]	Minimize hovering points and flying distance of UAV	Particle swarm optimization	No	One-to-multiple
[33]	Maximize energy efficiency	Polynomial-time approximation	No	One-to-one
[34]	Minimize trajectory path while maximizing network lifetime	Ant colony algorithm	No	One-to-one
[35]	Maximize energy of nodes	Geometry-based algorithm	No	One-to-one
[40]	Minimize the overall data packet loss	DRL	Yes	One-to-multiple
[41]	Minimize charging path	DRL Dynamic routing	Yes	One-to-one
[42]	Minimize death time of nodes and UAV energy consumption	DRL	Yes	One-to-one
[43]	Minimize node death rate	DRL	Yes	One-to-one
D-JERDG	Minimize node death rate and flight energy consumption while maximizing UAV throughput	DRL Dynamic routing	Yes	One-to-multiple

### 3. System Model and Problem Formulation

In this section, we present the system model and problem formulation of UAV-assisted wireless rechargeable sensor networks (UAV-WRSNs). For the sake of clarity, Table 2 lists the symbols used in this paper.

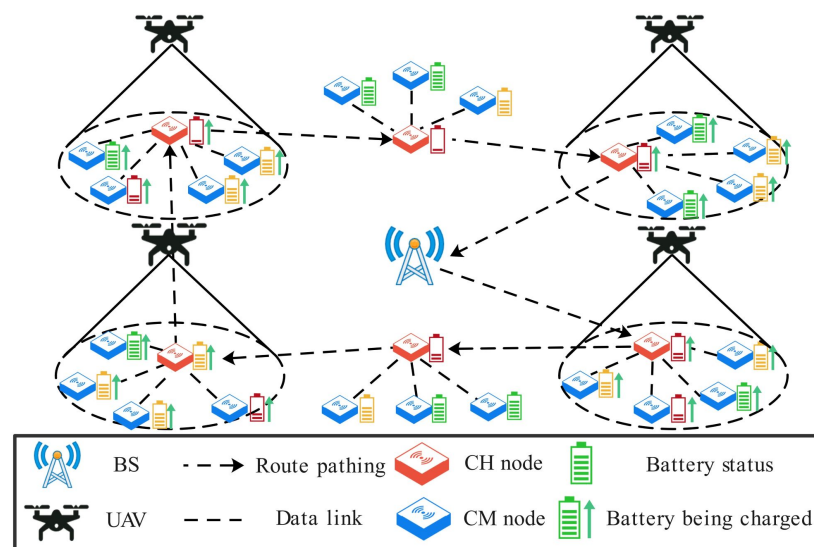
**Table 2.** Notation and definition.

Notations	Definitions	Notations	Definitions
$k$	total number of nodes	$D_i(t)$	buffer length of node $i$
$m$	number of CH nodes	$v_u(t)$	speed of the UAV
$h$	number of CM nodes	$\theta_u(t)$	yaw angle of the UAV
$Q_{max}$	buffer capacity of nodes	$\alpha$	weight coefficient of CM nodes
$E_{max}$	battery capacity of nodes	$R_c$	maximum charging distance
$RE_i(t)$	remaining energy of node $i$	$R_d$	maximum data transmission distance

### 3.1. System Model

#### 3.1.1. Network Model

As shown in Figure 1, the network is deployed in a two-dimensional area and consists of a UAV, a base station (BS), and randomly distributed sensor nodes. The UAV is assumed to have computing and wireless transmission capabilities, be equipped with a sufficiently charged battery, and be able to obtain its position through the Global Positioning System (GPS). The BS has ample energy and communication capabilities, enabling direct wireless communication with the UAV. The sensor nodes can monitor their remaining energy and data fusion capabilities. Cluster member (CM) nodes send data packets to cluster head (CH) nodes, which aggregate the data and store them in a data buffer. The UAV takes off from the BS and follows a “fly-hover” pattern to visit each CH node in a predetermined order continuously. The CH nodes consume energy to upload data packets to the UAV, while the UAV provides charging to the nodes within its coverage range using radio frequency transmission technology.



**Figure 1.** Network model of the UAV-WRSN.

#### 3.1.2. Sensor Node Model

The network deploys  $k$  sensor nodes with fixed and known positions. Consider the set of nodes in the network as  $S_x \triangleq \{s_1, s_2, s_3, \dots, s_k\}$ . The CH nodes are denoted as  $ch_i \triangleq \{ch_1, ch_2, ch_3, \dots, ch_m\}$ , and the CM nodes are denoted as  $cm_j \triangleq \{cm_1, cm_2, cm_3, \dots, cm_h\}$ . Assuming that node batteries can be charged within a very short time compared with the data collection time by the UAV, we can neglect the charging time. Regarding the energy consumption of nodes in sleep mode, since this portion of energy consumption is very small and can be considered negligible, we have not included it as a significant influencing factor, and therefore, it has not been incorporated into the node energy consumption. Nodes only consume energy when receiving and transmitting data. The data buffer length of the CH node  $ch_i$  at  $t$  is denoted as  $D_{ch_i}(t) = [0, Q_{max}]$ . Based on the literature [43], assuming that, at any time  $t$ , the energy consumption of the CM node  $cm_j$  sending a unit data packet (where the unit data packet size  $f$  is 1024 kb) to the CH node  $ch_i$  can be given as follows:

$$E_j^{cm}(t) = e_t \sum_{j=1}^h f_{j,i} \quad (1)$$

where  $e_t$  indicates the energy consumption of transmitting unit data, the energy consumption of the CH node  $ch_i$  can be divided into two parts: reception energy and transmission energy:

$$E_i^{ch}(t) = \sum_{j=1}^h f_{j,i} \sum_{i=1}^m e_r + E_{ch_i}^t(t) \quad (2)$$

where  $e_r$  indicates the energy consumption of receiving unit data, and the remaining energy of the CM node  $cm_j$  and the CH node  $ch_i$  at  $t$  can be respectively defined as follows:

$$RE_j^{cm}(t) = E_j^{cm}(t-1) - E_j^{cm}(t) \quad (3)$$

$$RE_i^{ch}(t) = E_i^{ch}(t-1) - E_i^{ch}(t) \quad (4)$$

### 3.1.3. UAV Model

The UAV maintains a fixed flight altitude of  $H$ . It acquires the positions of all nodes before taking off from the base station and maintains communication with the base station throughout the operation. At  $t$ , the three-dimensional coordinates of the UAV can be represented as  $[x_u(t), y_u(t), H]$ . The distance between the UAV and the node plays a crucial role in determining the link quality and the ability of the node to upload data. In our scenario, we assume that the UAV has limited communication and charging ranges, denoted as  $R_d$  (maximum data transmission radius) and  $R_c$  (maximum charging radius), respectively.  $\Delta d_{u, ch_i^{tar}}$  represents the distance between the UAV and the target CH node if  $\Delta d_{u, ch_i^{tar}} \leq R_d$ ; the UAV enters the hovering state to collect data from the target CH node and charge all nodes within the charging range. The instantaneous speed  $v_u(t)$  and the instantaneous heading angle  $\theta_u(t)$  describe the flight control of the UAV. For safety reasons,  $v_u(t)$  must be within the minimum and maximum speed limits.

$$V_{\min} \leq v_u(t) \leq V_{\max} \quad (5)$$

This paper employs a rotor-wing UAV, and the energy consumption of the UAV can be divided into propulsion energy consumption and communication energy consumption. The propulsion energy consumption is further divided into flight energy consumption and hovering energy consumption [20]. When the UAV has an instantaneous speed  $v_u(t)$ , the propulsion energy consumption model of the UAV can be given as follows:

$$E_u^{pro}(v_u(t)) = P_0 \left(1 + \frac{3v_u(t)^2}{U_{tip}^2}\right) + P_f \left(\sqrt{1 + \frac{v_u(t)^4}{4v_0^4}} - \frac{v_u(t)^2}{2v_0^2}\right)^{1/2} + \frac{1}{2} d_0 \rho_{air} s_{rotor} A_{rotor} v_u(t)^3 \quad (6)$$

where  $P_0$  and  $P_f$  are constants that represent the blade profile power and induced power in hovering status, respectively;  $U_{tip}$  indicates the tip speed of the rotor blade;  $v_0$  denotes the mean rotor-induced speed; and  $d_0$ ,  $\rho_{air}$ ,  $s_{rotor}$ , and  $A_{rotor}$ , respectively, denote the fuselage drag ratio, air density, rotor solidity, and rotor disc area. At  $t$ , UAV speed is  $v_u(t)$ , and the flight energy consumption is expressed as  $E_u^{fly} = E_u^{pro}(v_u(t))$ . It is worth noting that the UAV is in hovering status with flight speed  $v_u(t) = 0$  and hovering energy consumption  $E_{hover} = E_u^{pro}(v_u(t) = 0)$ . To facilitate the expression, we divide the time domain  $T$  into  $n$  time steps, where each time step is denoted as  $t_0, t_1, t_2, \dots, t_n$ ; thus, the total flight energy consumption of the UAV at each time step  $t_n$  is given as follows:

$$E_u^{total}(t_n) = \int_0^{t_n} E_u^{pro}(v_u(t)) dt \quad (7)$$

the UAV's average flight energy consumption is given as follows:

$$E_u^{ave} = \frac{E_u^{total}(T)}{T} \quad (8)$$

### 3.1.4. Transmission Model

The wireless communication link between the UAV and the nodes mainly consists of a line-of-sight (LoS) link and a non-line-of-sight (NLoS) link, similar to many existing works [20,39,40]. Let the coordinate of the target CH node  $ch_i^{tar}$  be  $[x_{ch_i^{tar}}, y_{ch_i^{tar}}, 0]$ ; the free path loss of the LoS link between the UAV and  $ch_i^{tar}$  at  $t$  is modeled as follows:

$$\begin{aligned} h_{u,ch_i^{tar}}(t) &= \gamma_0 \Delta d_{u,ch_i^{tar}}(t)^{-2} \\ &= \frac{\gamma_0}{H^2 + (x_u(t) - x_{ch_i^{tar}})^2 + (y_u(t) - y_{ch_i^{tar}})^2} \end{aligned} \quad (9)$$

where the channel's power gain at a reference distance of  $\Delta d_{u,ch_i^{tar}} = 1$  m is denoted by  $\gamma_0$ , and  $\Delta d_{u,ch_i^{tar}}(t)$  denotes the Euclidean distance between the UAV and  $ch_i^{tar}$  at  $t$ , where the path loss of the NLoS link is given as  $\eta^{NLoS} \gamma_0 \Delta d_{u,ch_i^{tar}}(t)^{-2}$ , and  $\eta^{NLoS}$  is the attenuation coefficient of the NLoS links.

At  $t$ , the LoS link probability between the UAV and  $ch_i^{tar}$  is modeled as follows:

$$P_{u,ch_i^{tar}}^{LoS}(\theta_{ch_i^{tar}}(t)) = \frac{1}{1 + \alpha \exp(-\beta(\theta_{u,ch_i^{tar}}(t) - \alpha))} \quad (10)$$

where  $\alpha$  and  $\beta$  are constant sigmoid parameters that depend on the signal propagation environment and propagation distance.  $\theta_{ch_i^{tar}}(t)$  is the elevation angle of the UAV and  $ch_i^{tar}$  in degree; it is given as  $\theta_{ch_i^{tar}}(t) = \frac{180}{\pi} \sin^{-1}\left(\frac{H}{\Delta d_{u,ch_i^{tar}}}\right)$ . The NLoS link probability is given as follows:

$$P_{u,ch_i^{tar}}^{NLoS}(\theta_{ch_i^{tar}}(t)) = 1 - P_{u,ch_i^{tar}}^{LoS}(\theta_{ch_i^{tar}}(t)) \quad (11)$$

let  $g_{u,ch_i^{tar}}(t)$  be the wireless channel gain between the UAV and  $ch_i^{tar}$ ; it is modeled as follows:

$$g_{u,ch_i^{tar}}(t) = (P_{u,ch_i^{tar}}^{LoS}(\theta_{u,ch_i^{tar}}(t)) + \eta^{NLoS} P_{u,ch_i^{tar}}^{NLoS}(\theta_{u,ch_i^{tar}}(t))) h_{u,ch_i^{tar}}(t) \quad (12)$$

in this paper, we assume that the uplink and downlink channels are approximately equal. As a result, the channel power gain between the UAV and  $ch_i^{tar}$  is given as follows:

$$\begin{aligned} h_{u,ch_i^{tar}}(t) &\approx g_{u,ch_i^{tar}}(t) = (P_{u,ch_i^{tar}}^{LoS}(\theta_{u,ch_i^{tar}}(t)) \\ &\quad + \eta^{NLoS} P_{u,ch_i^{tar}}^{NLoS}(\theta_{u,ch_i^{tar}}(t))) h_{u,ch_i^{tar}}(t) \end{aligned} \quad (13)$$

Assuming that the UAV establishes a communication link with  $ch_i^{tar}$  at  $t$  and  $ch_i^{tar}$  maintains a constant transmit power  $P_t$  to upload data to the UAV, according to Shannon's formula [44], the data transmission rate between the UAV and  $ch_i^{tar}$  can be given as follows:

$$R_{ch_i^{tar}}^t(t) = B \log_2 \left( 1 + \frac{P_t \left| g_{u,ch_i^{tar}}(t) \right|^2}{\sigma^2} \right) \quad (14)$$

where  $B$  and  $\sigma^2$  represent the channel bandwidth and noise power, respectively, and  $P_t$  denotes the transmit power of the node; hovering time (upload data time) is given as follows:

$$t_{hover} = \frac{D_{ch_i^{tar}}(t)}{R_{ch_i^{tar}}^t(t)} \quad (15)$$

the energy consumption for  $ch_i^{tar}$  to upload data at time  $t$  can be given as follows:

$$E_{ch_i^{tar}}^t(t) = P_t t_{hover} \quad (16)$$

### 3.2. Problem Formulation

In the proposed D-JERDG, we consider the UAV capabilities of energy replenishment and data collection. The UAV departs from the BS and moves according to a planned flight path. It is responsible for collecting data generated by all cluster heads in the network. After completing a mission cycle, the UAV returns to the BS to recharge and transmit the collected data back to the BS. The flight path consists of the hovering positions and the CH node access sequence. First, to improve data collection efficiency, the UAV needs to serve as many nodes as possible within the maximum flight time  $T$  and collect data. Second, the charging strategy for the target CH node is crucial to minimize the node death rate and flight costs. To ensure that high-energy-consuming nodes and remote area nodes receive energy replenishment while reducing ineffective flight time, the UAV flight distance and speed need to be controlled.

In summary, the overall objective of D-JERDG is to minimize the UAV flight energy consumption and maximize its throughput while minimizing the node death rate. This problem is a multiobjective optimization problem, and the objective functions can be formulated as follows:

$$P1 : \min(N_d, E_u^{ave}, L_u^{total}) \quad (17)$$

The constraint condition can be formulated as follows:

$$0 \leq \sum_{n=0}^N t_n \leq T \quad (18)$$

$$\Delta d_{u, ch_i^{tar}} < R_d \quad (19)$$

$$\forall \Delta d_{u, cm_j} < R_c \quad (20)$$

$$N_d^i = \begin{cases} 0, RE_i(t) = 0 \\ 1, RE_i(t) > 0 \end{cases} \quad (21)$$

$$v_{\min} \leq v_u(t) \leq v_{\max} \quad (22)$$

$$0 < \theta_u(t) \leq 2\pi \quad (23)$$

$$0 \leq x_u(t) \leq L \quad (24)$$

$$0 \leq y_u(t) \leq L \quad (25)$$

In objective functions,  $N_d^i (i = (1, 2, \dots, k))$  is denoted as the number of dead nodes,  $L_u^{total}$  represents the total flight distance, and  $E_u^{ave}$  represents the UAV's average flight energy consumption. In the constraint condition, Formula (18) represents the UAV with a maximum flight time of no more than  $T$ . Formula (19) represents the maximum data transmission range of the node into the UAV to establish a communication link with the UAV. Formula (20) indicates that nodes can harvest energy from the UAV within the maximum charging range of the UAV. Formula (21) represents the energy state. If  $RE_i(t) = 0$ , node  $i$  is considered a dead node. If  $RE_i(t) < 0$ , node  $i$  is in a normal state. Formulas (22) and (23) represent the size limit of the UAV's speed and heading angle. Formulas (24) and (25) represent that the UAV's flight range does not extend beyond the two-dimensional area.

## 4. A Joint Energy Replenishment and Data Collection Algorithm Based on Deep Reinforcement Learning

In this section, we presented the effective implementation of D-JERDG in UAV-WRSN. First, we introduced a UAV-WRSN with a random topology structure. Then, we described our proposed cluster head selection algorithm. Using the obtained set of CH nodes, we applied the simulated annealing algorithm to solve the traveling salesman problem (TSP). Afterward, we provided the state space, action space, and reward function of the MODDPG algorithm.

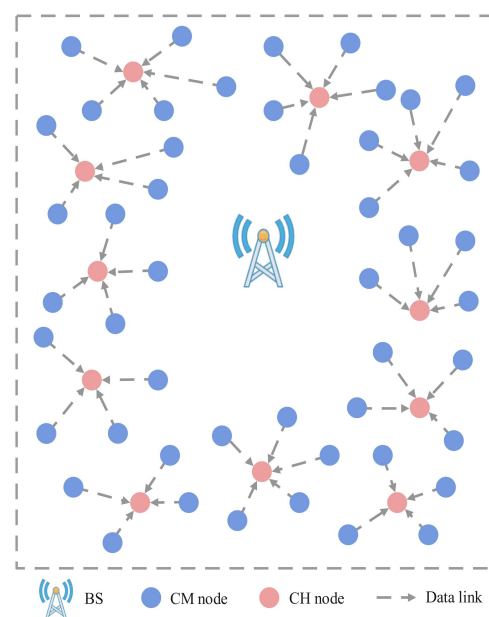


#### 4.1. Cluster Head Selection Algorithm

Referring to [39], this network studied in this paper is a UAV-WRSN that uses a dynamic routing protocol. Figure 2 illustrates the generated clusters and data flow. Compared with static routing protocols, dynamic routing protocols can evenly consume the energy of nodes. However, in a traditional LEACH routing protocol [45], the CM nodes determine whether to become CH nodes based on randomly generated thresholds. For each CM node  $cm_j$ ,  $w_j^{th}$  is given as follows:

$$w_j^{th} = \begin{cases} \frac{p}{1-p(r \bmod \frac{1}{p})}, & \text{if } j \in cm_j \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

where  $p$  is the expected percentage of CH nodes, and  $r$  is the current number of rounds. This method lacks consideration for the remaining energy and location of the nodes. We have redesigned the CH node selection mechanism based on these two factors. Due to insufficient energy information before the network operation, we utilize the K-means algorithm to partition all the data into  $m$  clusters and identify  $m$  centroids within each cluster. The overall optimization objective of the K-means algorithm is to minimize the sum of distances between each point and its respective centroid within the clusters. Based on the distances between nodes and centroids, we have determined  $m$  CH nodes. As a result, when the UAV reaches a CH node, it can cover more nodes, thereby improving the charging efficiency.

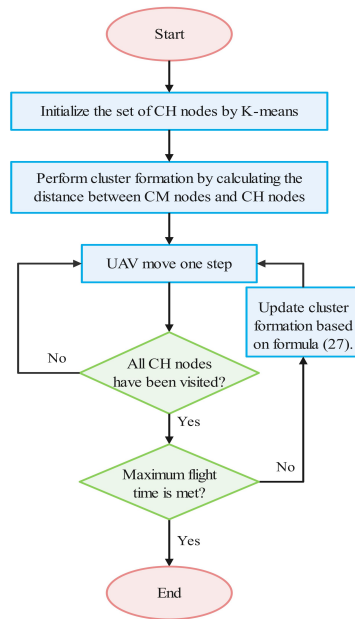


**Figure 2.** Illustration of clusters and data flow.

At each round, the cluster member nodes at each cluster elect CH nodes based on their own energy status and location information to balance the energy consumption of the nodes. Dynamic routing protocols are used within each cluster. We stipulate no communication link between the sensor nodes and BS. CM nodes use CH nodes as the next hop for communication, while CH nodes collect data sent by member nodes and transmit them to the UAV. (If a CH node has nonzero data storage after the UAV completes a flight cycle, it will continue to transmit data as a CM node in the next mission.) The weight of CM nodes can be obtained using the following formula:

$$w_{cm_j} = \kappa \frac{E_{\max} - RE_j^{cm}(t)}{RE_{ave}(t)} + (1 - \kappa) \frac{d_{i,j}}{d_{ave}} \quad (27)$$

where  $RE_{ave}(t)$  represents the average remaining energy of CM nodes,  $d_{ave}$  represents the average distance between CM nodes and CH nodes,  $d_{i,j}$  represents the distance between the CH node  $ch_i$  and the CM node  $cm_j$ , and  $\kappa$  is a weighting coefficient. The weight value  $w_j$  represents the priority or importance of CM; among CM nodes, the one with the highest  $w_j$  will be selected as the CH node in the next mission round. The CH node will then join the cluster as a CM node. Based on the obtained coordinates of the CH nodes, this paper utilizes the SA algorithm to solve the TSP. The pseudocode for the SA algorithm is presented in Algorithm 1. The specific algorithmic process is shown in Figure 3.



**Figure 3.** Illustration of the algorithm process.

#### 4.2. UAV Control Algorithm Based on MODDPG

In contrast with reinforcement learning algorithms based on an action value, our scenario involves a continuous action space for controlling the instantaneous speed and heading angle of the UAV. This necessitates a distinct modeling and solution approach. To tackle this continuous action space, we propose utilizing the MODDPG model for policy learning. The algorithmic specifics of MODDPG are elaborated in Algorithm 2.

---

#### Algorithm 1 Solve TSP by using simulated annealing algorithm

---

**Input:** The coordinates of CH nodes  $l_i$

**Output:** The sequence of CH nodes  $l_{new}$

- 1: Initialize maximum iteration temperature  $K_0 = 50,000$ , the sequence of CH nodes  $l_i$ , and annealing coefficient  $q = 0.98$ .
  - 2:  $K \leftarrow K_0$ .
  - 3: **while**  $K > K_0$  **do**
  - 4:   Swap the order of any two nodes to generate a new visiting sequence  $l_{new}$ .
  - 5:   Calculating  $l_{new}$ , denoted as  $df$ .
  - 6:   **if**  $df > 0$  **then**
  - 7:     **if**  $\exp(\frac{df}{K_0}) > rand(0,1)$  **then**
  - 8:        $l_f \leftarrow l_{new}$
  - 9:     **end if**
  - 10:   **else**
  - 11:      $K = K \times q$
  - 12:   **end if**
  - 13: **end while**
-

**Algorithm 2** MODDPG**Input:** State space  $s_{rl}$ **Output:** The action of UAV  $a_{rl}$ 

- 1: Initialize Actor network parameters  $\theta^\mu$ , and Target Actor network parameters  $\theta^{\mu'}$ ,  $\theta^{\mu'} \leftarrow \theta^\mu$ .
- 2: Initialize Critic network parameters  $\theta^Q$ , and Target Critic network parameters  $\theta^{Q'}$ ,  $\theta^{Q'} \leftarrow \theta^Q$ .
- 3: Initialize experience Replay Buffer.
- 4: **for** episode = 0 to  $M$  **do**
- 5:     Initialize  $s_{rl}$ .
- 6:     **for** time step= $t_1, t_2, \dots, t_n$  **do**
- 7:         **repeat**
- 8:             With probability of  $\epsilon$  choose an action  $a_{rl} = clip(\mu(s_{rl}|Q_\mu) + \epsilon, a_{low}, a_{high})$ .
- 9:             Perform action  $a_{rl}$  and observe reward  $r_{rl}$  and next state  $s'_{rl}$ .
- 10:             Store the transition  $(s_{rl}, a_{rl}, r_{rl}, s'_{rl})$  from  $P$ .
- 11:             **if**  $P \geq$  Batch size **then**
- 12:                 Randomly sample Mini batch transitions  $(s_{rl}, a_{rl}, r_{rl}, s'_{rl})$  from  $P$ .
- 13:                 Compute  $y_{rl}$  (19).
- 14:                 Update Critic network by minimizing the critic loss (20).
- 15:                 Update Actor network by maximizing the actor loss (21).
- 16:                 Soft Update Target Network Parameters.
- 17:                  $\theta^{Q'} \leftarrow \tau\theta^Q + (1 - \tau)\theta^{Q'}$
- 17:                  $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1 - \tau)\theta^{\mu'}$
- 18:             **end if**
- 19:     **until**  $\sum_{n=0}^N t_n \geq T$
- 20:     **end for**
- 21: **end for**

- State Space: The state space set is defined as follows:

$$s_{rl}(t) = \left\{ d_{u, ch_i^{tar}}^x(t), d_{u, ch_i^{tar}}^y(t), x_u(t), y_u(t), N_f(t), N_d(t), \forall i \in [1, m] \right\} \quad (28)$$

Specifically, with a state space size of 6,  $d_{u, ch_i^{tar}}^x$  and  $d_{u, ch_i^{tar}}^y$  represent the relative distances in the horizontal and vertical coordinates between the UAV and the target CH node  $i$ , and  $N_f(t)$  records the cumulative number of times that the UAV has continuously exceeded the restricted area by time  $t$ .

- Action Space: At each  $t$ , the action of the UAV is defined as follows:  
 $\theta_u(t) \in (0, 2\pi]$ : Instantaneous heading angle of the UAV along the horizontal direction at  $t$ .  
 $v_u(t) \in [0, v_{max}]$ : Instantaneous speed of the UAV at  $t$ .  
The action of the UAV is defined as  $a_{rl} = [v_u(t), \theta_u(t)]$
- Reward Function: At any time step  $t_n$ , the UAV receives a reward  $r_{rl}$ . To ensure the throughput of the UAV, the UAV receives an immediate reward  $r_0$  whenever it successfully establishes communication with the target node  $ch_i^{tar}$ . The reward function is defined as follows:

$$r_{rl} = \begin{cases} r_0(t) + r_1(t), & \text{if } \Delta d_{u, ch_i^{tar}}(t) \leq R_d \\ r_1(t), & \text{otherwise} \end{cases} \quad (29)$$

$$r_0(t) = r_{serve} + \eta_1 \lambda_{cover}(t) + \eta_2 R_{ch_i^{tar}}^{tran}(t) \quad (30)$$

$$r_1(t) = -\eta_2 \Delta d_{u, ch_i^{tar}}(t) - \eta_3 (E_u^{pro}(t) + N_d(t)) \quad (31)$$

Specifically, the reward function is divided into two parts,  $r_0$  and  $r_1$ , where  $r_{serve}$  represents the reward for the UAV establishing a connection with CH nodes;  $\lambda_{cover}$  indicates the number of nodes covered by the UAV; and  $\eta_1$ ,  $\eta_2$ , and  $\eta_3$  are reward weight factors for the optimization objectives.  $r_1$  is based on the UAV actions at each time step and includes factors such as the distance between the UAV and CH nodes, the energy consumption of the UAV flight, and the number of dead nodes  $N_d(t)$ .

The MODDPG network framework is shown in Figure 4; it consists of four networks, with the main network and target network each comprising two networks: the actor network, critic network, target actor network, and target critic network. The main network and target network share the same network structure, where the actor network outputs UAV actions, and the critic network evaluates these actions to update the actor network. The MODDPG algorithm optimizes the UAV instantaneous speed and heading angles, ultimately learning the optimal policy.

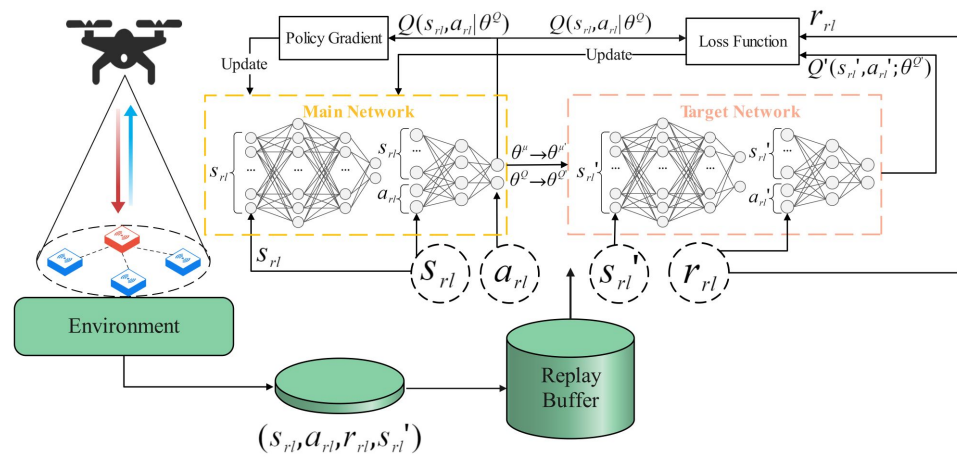


Figure 4. MODDPG network framework.

The calculation of the target value  $y_{rl}$  is as follows:

$$y_{rl} = r_{rl} + \gamma Q'(s_{rl}', \mu'(s_{rl}'; \theta^{\mu'}); \theta^{Q'}) \quad (32)$$

The critic network is trained using gradient descent to minimize the loss between the predicted value  $Q(s_{rl}, a_{rl}; \theta^Q)$  and the target value  $y_{rl}$ .

$$Loss = \frac{1}{M} \sum (y_{rl} - Q(s_{rl}, a_{rl}; \theta^Q))^2 \quad (33)$$

Update the actor network using a gradient-based policy algorithm:

$$\nabla_{\theta^{\mu}} J = \frac{1}{M} \sum \nabla_{a_{rl}} Q(s_{rl}, a_{rl}; \theta^Q) |_{s=s_{rl}, a=\mu(s_{rl})} \nabla_{\theta^{\mu}} \mu(s_{rl}; \theta^{\mu}) |_{s_{rl}} \quad (34)$$

where  $J$  represents the total discounted cumulative reward. The specific algorithm flow is described as follows: Initialize the replay buffer, critic network, and actor network parameters  $\theta^Q$  and  $\theta^{\mu}$ , as well as the target critic network and target actor network parameters  $\theta^{Q'}$  and  $\theta^{\mu'}$  (lines 1–3). Perform the exploration phase for the UAV in lines 4–10, and generate an action  $a_{rl} = clip(\mu(s_{rl}|Q_{\mu}) + \epsilon)$  from the actor network. This process is repeated until the maximum task time is reached, and the experiences are stored in the replay buffer. Updating network parameters, when the experience replay buffer  $P$  is full, start training the networks to update actor network and critic network parameters (lines 11–15). Update the critic network parameters by minimizing the loss function (line 14), and update the actor network parameters using the deterministic policy gradient (line 15). Finally, use the

soft update technique to update the parameters of the target actor network, controlling the update frequency to converge faster toward the optimal policy (lines 17–21).

## 5. Experimental Results

In this section, we conducted simulation results to validate the effectiveness of D-JERDG. Specifically, we provided numerical results and analyzed the convergence of D-JERDG by adjusting the number of CM nodes, CH nodes, and maximum charging radius. We compared D-JERDG with MODDPG and the random method. We analyzed the differences between the algorithms from five aspects: node death rate, data overflow rate, UAV throughput, average flight energy consumption, and time utilization. By comparing these metrics, we were able to evaluate and highlight the differences and advantages of D-JERDG over the other algorithms in terms of efficiency and performance.

- **D-JERDG:** D-JERDG uses the K-means clustering algorithm to divide nodes into multiple clusters, designing the CH node selection mechanism based on an improved dynamic routing protocol. The SA algorithm determines the CH node access sequence, and the UAV flight actions are controlled by inputting the UAV and network states into the DRL model MODDPG.
- **MODDPG (multiobjective deep deterministic policy gradient) [20]:** This method selects the node with the highest data volume as the target node. By observing the relative positions of the UAV and the target node, as well as the node states, a DRL model is established to control the UAV learning of all node positions. Through accessing the nodes, data collection and range-based charging are achieved.
- **Random:** The method is based on the MODDPG [20] definition, with the difference being that the random algorithm randomly selects a node as the target node and uses the DRL model to optimize the UAV's flight trajectory.

### 5.1. Simulation Settings

We consider the UAV taking off from the BS, the maximum flight time  $T$  to be set at 10 min. The sensor nodes are randomly distributed within a  $400 \times 400 \text{ m}^2$  square two-dimensional area. The UAV flies at an altitude of 10 m, with a maximum data transmission radius  $R_d = 10 \text{ m}$  and a maximum charging radius  $R_c = 30 \text{ m}$ . The maximum flight speed of the UAV is set to 20 m/s [20]. The coordinates of the nodes are randomly generated and remain fixed. The nodes' maximum data storage capacity is  $Q_{max} = 100 \text{ Mb}$  [40]. The total energy of the nodes is set to  $E_{max} = 800 \text{ J}$  [40], and their initial energy is randomly generated within the range of  $[0 \text{ J}, 800 \text{ J}]$  [40]. The transmission energy consumption for CM nodes is  $e_t = 0.4 \text{ J}$  [43], and the reception energy consumption for CH nodes is  $e_r = 0.5 \text{ J}$  [43]. Table 3 and Table 4, respectively, show the network parameters of MODDPG and the main environmental parameters, as specified in reference [20]. The operating system environment of the simulation experiment is TensorFlow 2.5.0, TensorLayer 2.2.5, and Python 3.7 on a Linux server with four NVIDIA 3090 GPUs.

**Table 3.** Network parameters.

Parameters	Values
Episodes ( $M$ )	400
Actor network structure	$400 \times 300 \times 300$
Critic network structure	$400 \times 300$
Actor network hidden layers	3
Critic network hidden layers	2
Batch size	64
Learning rate for actor ( $lr_a$ )	$1 \times 10^{-3}$
Learning rate for critic ( $lr_c$ )	$1 \times 10^{-3}$
Discount factor ( $\gamma$ )	0.9

**Table 3.** Cont.

Parameters	Values
Replay buffer ( $P$ )	8000
Target network soft update rate ( $\tau$ )	$1 \times 10^{-3}$
Reward weights ( $\eta_1, \eta_2, \eta_3$ )	50, 100, 5

**Table 4.** Main simulation parameters.

Parameters	Values
Node transmit power ( $P_t$ )	$1 \times 10^{-3}$ W
Buffer capacity of nodes ( $Q_{max}$ )	100 Mb
Battery capacity of nodes ( $E_{max}$ )	800 J
The initial energy of nodes	$rand(0 \text{ J}, 800 \text{ J})$
The initial buffer length of nodes	$rand(0 \text{ Mb}, 5 \text{ Mb})$
Weighting coefficient ( $\kappa$ )	0.7
Channel bandwidth ( $B$ )	1 MHz
Channel power gain ( $\gamma_0$ )	−30 dB
Noise power ( $\sigma^2$ )	−90 dBm
NLoS path loss coefficient ( $\eta^{NLoS}$ )	0.2
LoS probability coefficient ( $\alpha, \beta$ )	10, 0.6
Blade profile power ( $P_0$ )	79.8563 W
Induced power ( $P_f$ )	88.6279 W
Tip speed of rotor blade ( $U_{tip}$ )	120 m/s
Mean rotor induced velocity ( $v_0$ )	4.03 m/s
Fuselage drag ratio ( $d_0$ )	0.6
Air density ( $\rho_{air}$ )	1.225 km/m <sup>3</sup>
Rotor solidity ( $S_{rotor}$ )	0.05
Rotor disc area ( $A_{rotor}$ )	0.503 m <sup>2</sup>

## 5.2. Evaluation Metrics

We introduce five key performance metrics to analyze and compare the optimization effects of D-JERDGD on UAV-WRSN in terms of system charging efficiency, data collection efficiency, and UAV flight energy consumption.

- **Node death rate:** The node death rate is determined by the proportion of dead nodes, where node  $i$  is considered dead when  $RE_i(t) = 0$ . A lower node death rate indicates a higher system charging efficiency. This metric determines which charging strategy can maintain the maximum number of surviving nodes in the experiment.
- **Node data overflow rate:** The node data overflow rate is determined by the proportion of nodes with data storage exceeding the maximum capacity, represented as  $D_i(t) = 100$ . This metric measures the effectiveness of UAV data collection and the clustering algorithm.
- **UAV throughput:** The UAV throughput is the ratio of the number of nodes accessed by the UAV to the total number of nodes within the maximum flight time. It measures the overall efficiency of the UAV-WRSN system.
- **Average flight energy consumption:** The average flight energy consumption is defined as the ratio of the total energy consumption during UAV flight to the maximum flight time. It is a key indicator for evaluating the energy consumption of the UAV-WRSN system.
- **Time utilization rate:** The time utilization rate represents the ratio of the hover time to the flight time of the UAV during a flight mission. A lower time utilization indicates a higher proportion of time occupied by UAV flights, serving as a metric for evaluating the data collection efficiency of the UAV. The time utilization rate  $R_{tu}$  is given as follows:

$$R_{tu} = \frac{t_{hover}}{t_{fly}} (T = t_{hover} + t_{fly}) \quad (35)$$

### 5.3. Convergence and Stability

After D-JERDG is implemented, we first test the convergence and stability of the proposed model. The episode was set as 400, and the model of the number of different clusters, the number of nodes in each cluster, and the maximum charging radius of the UAV were trained. “C20N200R30” means that there are 20 CH nodes and 180 CM nodes, the maximum charging radius of UAV is  $R_c = 30$  m, and the other sections in this chapter follow the same pattern. Then the functional relationship between episode and reward is shown in Figure 5. As can be seen from Figure 5, the reward of most models increased rapidly before 150 episodes. After 200 episodes, the value of rewards becomes stable and oscillates around a certain value until the training ends. The results indicate that the trained D-JERDG algorithm successfully converges and can make high-reward decisions in a dynamically changing network state. Figure 5a shows the reward curves under different numbers of CH nodes and CM nodes, ranging from “C20N200R30” to “C40N400R30” experimental scenarios. Figure 5b displays the reward curves for different numbers of CH nodes, with the experimental scenarios ranging from “C15N200R30” to “C30N200R30”. Figure 5c illustrates the reward curves under different UAV charging radii, with the experimental scenarios ranging from “C20N200R15” to “C20N200R30”.

Additionally, Figure 6 displays the loss curve and temporal difference error (TD error) curve of network training for the three algorithms, with the experimental scenario being “C20N200R30”. In Figure 6a, the curve trends of the three algorithms are mostly consistent before 150 episodes. After that, due to the increased UAV throughput in D-JERDG, some fluctuations occur. However, overall, the loss curves of the three algorithms converge before 400 episodes. In Figure 6b, we show the TD error curve of network training for the three algorithms. Overall, all three algorithms can converge before 400 iterations. However, before 100 episodes, D-JERDG has a slower convergence rate compared with the other two algorithms. Because we expect D-JERDG to learn the positions of more nodes in the early training episodes, we have implemented a strategy where the set of CH nodes initialized using the K-means algorithm is not fixed. The purpose of this approach is to ensure that D-JERDG can learn the positions of as many nodes as possible. By flexibly adjusting the CH node set, D-JERDG can better adapt to the varying node distributions in different environments, enhancing its performance and training effectiveness. After 100 episodes, as the rewards for D-JERDG start to increase and gradually converge, it also exhibits convergence in the TD error curve.

### 5.4. Performance Comparison

Specifically, without loss of generality, we conducted experiments with different numbers of nodes and varying UAV flight speeds. The number of nodes was set from “C10N100R30” to “C40N400R30”, with a maximum data transmission radius of  $R_d = 10$  m, a maximum charging radius of  $R_c = 30$  m, and a maximum flight speed ranging from 15 m/s to 20 m/s. With an increasing number of nodes, both the energy consumption and data volume of CH nodes increase within the same period. A larger flight speed allows the UAV to reach nodes faster, but it also increases energy consumption. On the other hand, a smaller speed can save energy but may sacrifice some throughput and result in a certain number of dead nodes, thereby affecting UAV throughput, the number of dead nodes, and time utilization. Additionally, as the number of nodes increases, in our proposed approach, the UAV flight time and distance are influenced by the increasing number of CH nodes, which also affects flight energy consumption. Based on these hypotheses, we conducted comparative experiments.

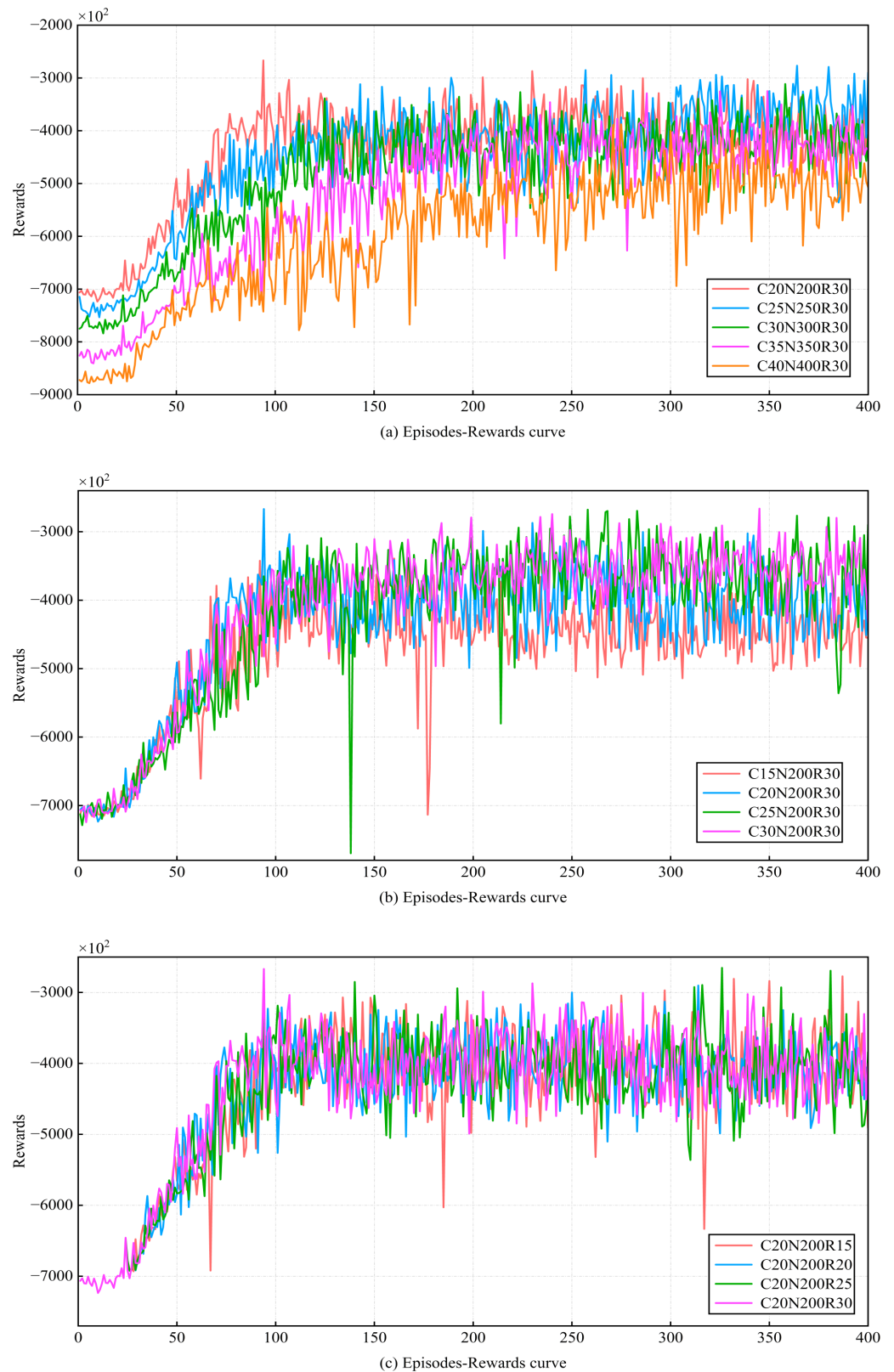
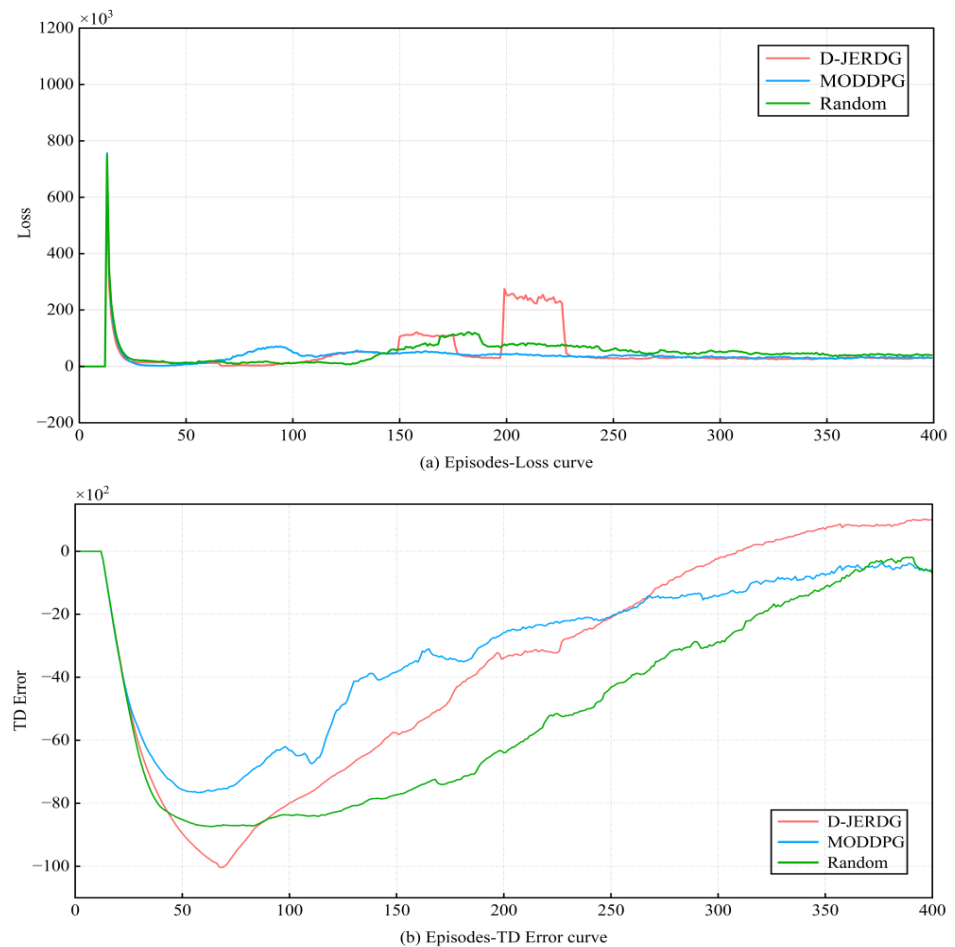


Figure 5. Learning curve with the episode = 400.

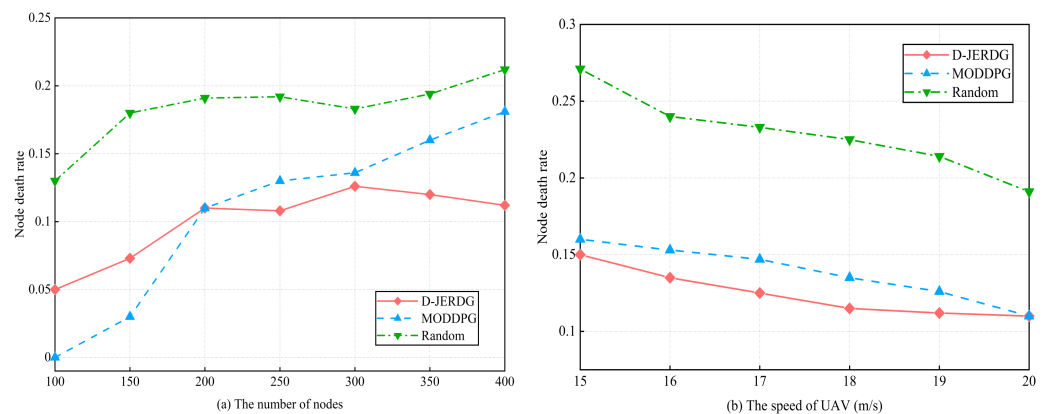




**Figure 6.** Learning curve with the episode = 400.

#### 5.4.1. Node Death Rate

In this section, we compared the performances of different algorithms in terms of node death rate. Specifically, in the MODDPG experiment, we designated the node with the least remaining energy as the target node. The experimental results are shown in Figure 7.



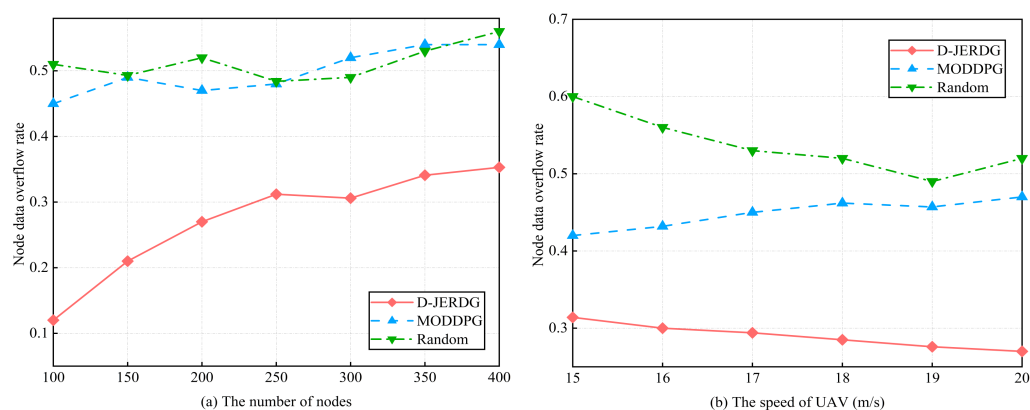
**Figure 7.** Experimental results of node death rate.

Figure 7a demonstrated a clear disparity in node death rates between D-JERDG and Random. Initially, MODDPG exhibited lower node death rates than D-JERDG for up to 200 nodes. This can be attributed to the advantage of the greedy-based node charging strategy in scenarios with a small number of nodes and low density. However, after

reaching 200 nodes, D-JERDGD demonstrated a lower node death rate than MODDPG, with a declining trend beyond 300 nodes. When the number of nodes is 250, compared with the MODDPG and random, D-JERDGD can decrease the node death rate by about 17% and 44%, respectively. This improvement can be attributed to the clustering algorithm employed in D-JERDGD, where several nodes closest to the CH node are selected as CM nodes. This approach ensures that the UAV covers more nodes during each hover, enhancing charging efficiency. The effectiveness of the proposed CH node selection mechanism was also validated, as it increased the likelihood of selecting CM nodes farther from the current CH node as new CH nodes. This enabled energy replenishment and reduced the node death rate of remote area nodes, making D-JERDGD more suitable for large-scale networks aiming to maintain network connectivity. From Figure 7b, we examined the variations in node death rates across different algorithms while considering different UAV flight speeds, with 200 nodes in the experiment. Overall, increasing the UAV flight speed resulted in reduced fluctuations in the node death rate, and D-JERDGD consistently outperformed MODDPG and random by maintaining lower node death rates. When the speed of the UAV was 15 to 20, compared with MODDPG and random, the average decline rate of D-JERDGD in terms of node death rate was about 10% and 45.5%, respectively. Furthermore, as the flight speed increased, the node death rate exhibited a significant decrease. This highlights the advantageous performance of D-JERDGD in terms of node death rate when compared with the other algorithms.

#### 5.4.2. Node Data Overflow Rate

In this section, we analyze the differences in node data overflow rates among different algorithms. In the MODDPG experiment, we designate the node with the highest data volume as the target node, and the experimental results are shown in Figure 8. In Figure 8a,b, with an increasing number of nodes, compared with MODDPG and random, the average decline rate of D-JERDGD in terms of node overflow rate is about 46.4% and 48.2%, respectively. With an increasing speed of the UAV, it is about 35.1% and 46%, respectively. This is because the dynamic routing protocol periodically selects CM nodes as CH nodes, allowing CH nodes to gather data from CM nodes and reduce node data overflow. However, for MODDPG and random, the one-to-one data collection approach leads to lower efficiency and inevitable node data overflow.

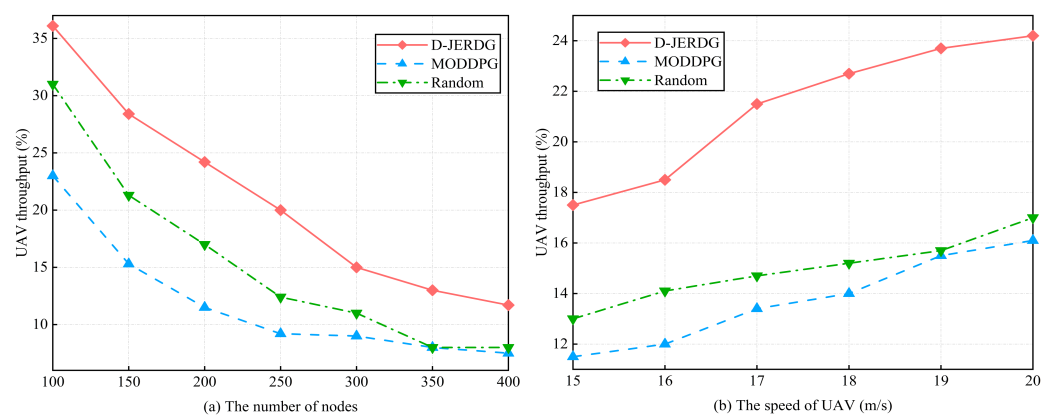


**Figure 8.** Experimental results of node data overflow rate.

#### 5.4.3. UAV Throughput

In this section, we compare the performances of different algorithms in terms of UAV throughput by changing the number of CH nodes and CM nodes in the network. The experimental results are shown in Figure 9. In Figure 9a, we observe that as the number of nodes increases, all algorithms experience a decrease in UAV throughput. However, D-JERDGD consistently achieves the highest throughput among the algorithms. Compared with MODDPG and random, the average growth rate of D-JERDGD in terms of UAV throughput

is about 79.3% and 43%, respectively. In D-JERDG, the use of a K-means-based clustering algorithm for CH node selection introduces randomness, allowing the UAV to quickly learn the positions of all nodes by obtaining different node coordinates in each training round. This enables D-JERDG to achieve a higher node access count within the same flight time. At 250 nodes, D-JERDG reaches a node access count of 50, indicating that it completes two rounds of CH node traversal. Beyond 300 nodes, D-JERDG throughput reaches a plateau and remains competitive with MODDPG. This can be attributed to the increased number of CH nodes requiring the UAV to spend more time hovering for data collection, leading to longer flight times and a decrease in node access count. In Figure 9b, we analyze the impact of different UAV flight speeds on throughput with a fixed number of 200 nodes. Overall, compared with MODDPG and random, D-JERDG outperforms MODDPG and random algorithms, and the throughput significantly increases as the flight speed rises. The average growth rate of D-JERDG in terms of UAV throughput is about 55.3% and 42.4%, respectively.



**Figure 9.** Experimental results of UAV throughput.

#### 5.4.4. Average Flight Energy Consumption

In this section, we analyze the differences in average UAV flight energy consumption among three algorithms. The experimental results are depicted in Figure 10. In Figure 10a, we observe that the UAV energy consumption in D-JERDG initially increases and then decreases as the number of nodes increases. Before reaching 200 nodes, the larger distances between CH nodes prompt the UAV to adapt its flight speed to minimize the flight time and efficiently collect data, resulting in an increasing energy consumption curve. However, after 200 nodes, as the number and density of CH nodes increase, the distances between them become shorter. Compared with MODDPG and random, the average decline rate of D-JERDG in terms of average flight energy consumption is about 4.2% and 4%, respectively. When the number of nodes is 350, D-JERDG can decrease the average flight energy consumption by about 10% and 9.8%, respectively. As a result, the UAV adjusts its flight speed by reducing it, which leads to longer flight times but decreases the overall UAV energy consumption. Compared with MODDPG and random, in Figure 10b, we examine the impact of different flight speeds on average energy consumption under the three algorithms. Overall, there is no significant difference among the algorithms (D-JERDG, MODDPG, and random) in terms of average energy consumption. However, as the flight speed increases, the average energy consumption increases as well. Notably, D-JERDG consistently maintains a lower average UAV flight energy consumption compared with MODDPG and random algorithms under different flight speed conditions.

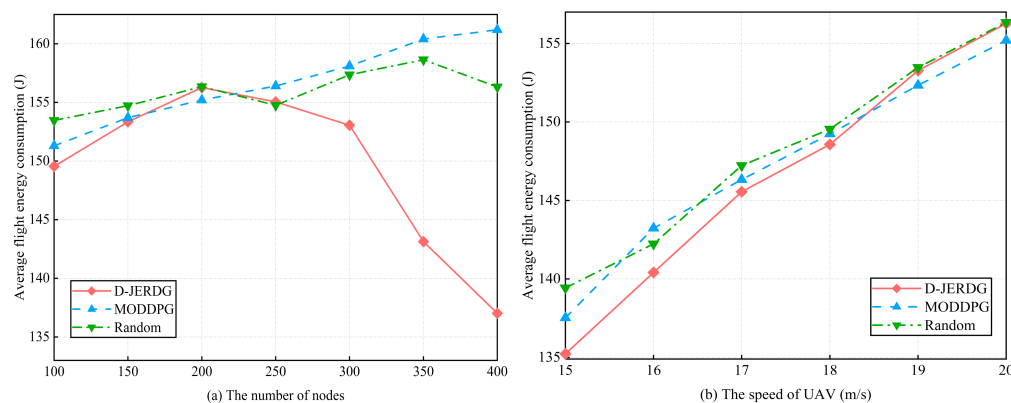


Figure 10. Experimental results of average flight energy consumption.

#### 5.4.5. Time Utilization Rate

In this section, we compare the differences in time utilization rates between the D-JERDG and MODDPG algorithms. The experimental results are presented in Figure 11. In Figure 11a, we observe that the time utilization of D-JERDG initially increases and then decreases as the number of CH nodes varies. This trend aligns with the analysis discussed earlier. In general, D-JERDG exhibits a similar pattern to MODDPG, and after 300 nodes, their time utilization becomes comparable. When the number of nodes is 150 to 350, the average growth rate of D-JERDG in terms of time utilization rates is about 23%. The reason behind this behavior lies in the D-JERDG method, which utilizes the SA algorithm to adjust the sequence of CH node visits, thereby reducing the UAV flight distance and time. Before reaching 300 nodes, with a smaller network size, the total flight time includes longer hover times. At 200 nodes, the time utilization reaches its peak, which aligns with the trend observed in the average UAV energy consumption curve. However, beyond 300 nodes, as the number of CH nodes increases, the flight time also increases. With the total flight time remaining constant, the time utilization decreases accordingly. In Figure 11b, we examine the impact of UAV flight speed on time utilization under both algorithms. Notably, D-JERDG consistently achieves higher time utilization compared with MODDPG, maintaining the highest level throughout the different flight speeds. As the speed of UAV increases, the average growth rate of D-JERDG in terms of time utilization rates is about 36%.

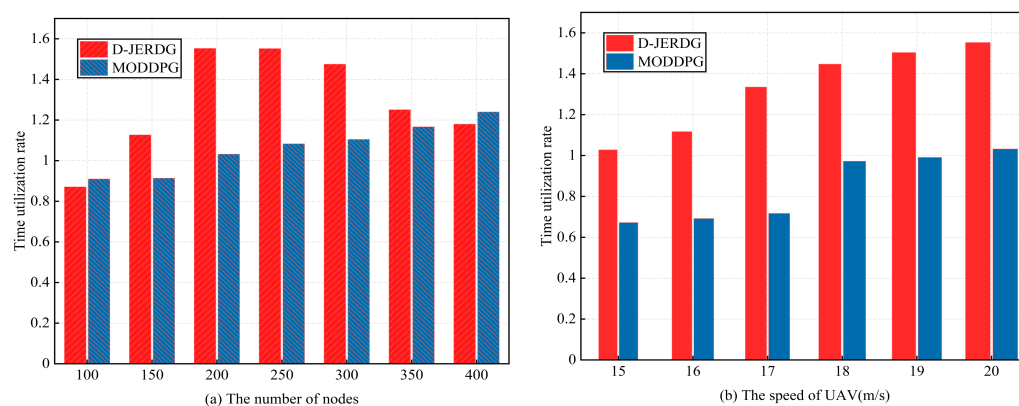


Figure 11. Experimental results of time utilization rate.

## 6. Conclusions

This study proposes a DRL-based method called D-JERDG for joint data collection and energy replenishment in UAV-WRSN. The main problems D-JERDG addresses are individual nodes' low data collection efficiency and the imbalance in node energy consumption.

D-JERDGD optimizes the network to tackle these issues by considering node inefficiency, UAV flight energy consumption, and UAV throughput. The clustering algorithm based on K-means and an improved dynamic routing protocol is used to cluster the nodes in the network. CH nodes are selected based on the remaining energy and geographical locations of the nodes within the clusters, effectively adapting to the dynamic nature of node energy consumption. Furthermore, a simulated annealing algorithm is employed to determine the visiting sequence, and the DRL model MODDPG is introduced to control the UAV for node servicing. Through extensive simulation results, the evaluation metrics of node inefficiency, UAV throughput, and average flight energy consumption are used to assess the performance of D-JERDGD. The results demonstrate that D-JERDGD achieves joint optimization of multiple objectives. It outperforms the existing MODDPG approach by significantly reducing node inefficiency, saving flight costs, and improving data collection efficiency. Moreover, multiple research studies suggest that multiagent systems have significant advantages in accomplishing complex tasks. Therefore, Multi-UAV-WRSN is expected to be a primary research focus in the future.

**Author Contributions:** Conceptualization, J.L.; Data curation, J.L. and Z.D.; Funding acquisition, Y.F.; Methodology, J.L. Project administration, Y.F. and N.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant No. 62062047, the major scientific and technological projects in Yunnan Province under Grant No. 202202AD080006, and the Yuxi Normal University under Grant No. 202105AG070010.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data are not publicly available due to it is not permitted.

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

## References

1. Kurs, A.; Karalis, A.; Moffatt, R.; Joannopoulos, J.D.; Fisher, P.; Soljacic, M. Wireless power transfer via strongly coupled magnetic resonances. *Science* **2007**, *317*, 83–86. [[CrossRef](#)] [[PubMed](#)]
2. Kurs, A.; Moffatt, R.; Soljačić, M. Simultaneous mid-range power transfer to multiple devices. *Appl. Phys. Lett.* **2010**, *96*, 044102. [[CrossRef](#)]
3. Yang, W.; Lin, C.; Dai, H.; Wang, P.; Ren, J.; Wang, L.; Wu, G.; Zhang, Q. Robust Wireless Rechargeable Sensor Networks. *IEEE/ACM Trans. Netw.* **2023**, *31*, 949–964. [[CrossRef](#)]
4. Rodić, A.; Mester, G. Virtual WRSN — Modeling and simulation of wireless robot-sensor networked systems. In Proceedings of the IEEE 8th International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, 10–11 September 2010; pp. 115–120. [[CrossRef](#)]
5. Deng, Q.; Ouyang, Y.; Tian, S.; Ran, R.; Gui, J.; Sekiya, H. Early Wake-Up Ahead Node for Fast Code Dissemination in Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* **2021**, *70*, 3877–3890. [[CrossRef](#)]
6. Schurgers, C.; Srivastava, M. Energy efficient routing in wireless sensor networks. In Proceedings of the 2001 MILCOM Proceedings Communications for Network-Centric Operations: Creating the Information Force (Cat. No.01CH37277), McLean, VA, USA, 28–31 October 2001; Volume 1, pp. 357–361. [[CrossRef](#)]
7. Goyal, N.; Dave, M.; Verma, A.K. Data aggregation in underwater wireless sensor network: Recent approaches and issues. *J. King Saud Univ.-Comput. Inf. Sci.* **2019**, *31*, 275–286. [[CrossRef](#)]
8. Dai, H.; Zhang, Y.; Wang, X.; Liu, A.X.; Chen, G. Omnidirectional Chargability with Directional Antennas. *IEEE Trans. Mob. Comput.* **2023**, *23*, 4483–4500. [[CrossRef](#)]
9. Dai, H.; Wang, X.; Lin, X.; Gu, R.; Shi, S.; Liu, Y.; Dou, W.; Chen, G. Placing Wireless Chargers With Limited Mobility. *IEEE Trans. Mob. Comput.* **2023**, *22*, 3589–3603. [[CrossRef](#)]
10. Zeng, Y.; Zhang, R.; Lim, T.J. Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Commun. Mag.* **2016**, *54*, 36–42. [[CrossRef](#)]
11. Pedditi, R.B.; Debasis, K. Energy Efficient Routing Protocol for an IoT-Based WSN System to Detect Forest Fires. *Appl. Sci.* **2023**, *13*, 3026. [[CrossRef](#)]
12. da Silva, R.I.; Del Duca Almeida, V.; Poersch, A.M.; Nogueira, J.M.S. Wireless sensor network for disaster management. In Proceedings of the 2010 IEEE Network Operations and Management Symposium—NOMS 2010, Osaka, Japan, 19–23 April 2010; pp. 870–873. [[CrossRef](#)]

13. Yang, J.; Wang, X.; Li, Z.; Yang, P.; Luo, X.; Zhang, K.; Zhang, S.; Chen, L. Path planning of unmanned aerial vehicles for farmland information monitoring based on WSN. In Proceedings of the 2016 12th World Congress on Intelligent Control and Automation (WCICA), Guilin, China, 12–15 June 2016; pp. 2834–2838. [\[CrossRef\]](#)
14. Liu, Y.; Pan, H.; Sun, G.; Wang, A. Scheduling Optimization of Charging UAV in Wireless Rechargeable Sensor Networks. In Proceedings of the 2021 IEEE Symposium on Computers and Communications (ISCC), Athens, Greece, 5–8 September 2021; pp. 1–7. [\[CrossRef\]](#)
15. Park, S.Y.; Jeong, D.; Shin, C.S.; Lee, H. DroneNet+: Adaptive Route Recovery Using Path Stitching of UAVs in Ad-Hoc Networks. In Proceedings of the GLOBECOM 2017—2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–7. [\[CrossRef\]](#)
16. Wang, Y.; Gao, Z.; Zhang, J.; Cao, X.; Zheng, D.; Gao, Y.; Ng, D.W.K.; Renzo, M.D. Trajectory Design for UAV-Based Internet of Things Data Collection: A Deep Reinforcement Learning Approach. *IEEE Internet Things J.* **2022**, *9*, 3899–3912. [\[CrossRef\]](#)
17. 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. [\[CrossRef\]](#)
18. Baek, J.; Han, S.I.; Han, Y. Energy-Efficient UAV Routing for Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 1741–1750. [\[CrossRef\]](#)
19. Delahaye, D.; Chaimatatanan, S.; Mongeau, M. Simulated Annealing: From Basics to Applications. In *Handbook of Metaheuristics*; Gendreau, M., Potvin, J.Y., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 1–35. [\[CrossRef\]](#)
20. Yu, Y.; Tang, J.; Huang, J.; Zhang, X.; So, D.K.C.; Wong, K.K. Multi-Objective Optimization for UAV-Assisted Wireless Powered IoT Networks Based on Extended DDPG Algorithm. *IEEE Trans. Commun.* **2021**, *69*, 6361–6374. [\[CrossRef\]](#)
21. Li, J.; Sun, G.; Wang, A.; Lei, M.; Liang, S.; Kang, H.; Liu, Y. A many-objective optimization charging scheme for wireless rechargeable sensor networks via mobile charging vehicles. *Comput. Netw.* **2022**, *215*, 109196. [\[CrossRef\]](#)
22. Dong, Y.; Li, S.; Bao, G.; Wang, C. An Efficient Combined Charging Strategy for Large-Scale Wireless Rechargeable Sensor Networks. *IEEE Sens. J.* **2020**, *20*, 10306–10315. [\[CrossRef\]](#)
23. Zhu, X.; Shen, L.; Yum, T.S.P. Hausdorff Clustering and Minimum Energy Routing for Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* **2009**, *58*, 990–997. [\[CrossRef\]](#)
24. Chand, S.; Singh, S.; Kumar, B. Heterogeneous HEED protocol for wireless sensor networks. *Wirel. Pers. Commun.* **2014**, *77*, 2117–2139. [\[CrossRef\]](#)
25. Wu, Q.; Sun, P.; Boukerche, A. A Novel Joint Data Gathering and Wireless Charging Scheme for Sustainable Wireless Sensor Networks. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [\[CrossRef\]](#)
26. Li, B.; Xiao, X.; Tang, S.; Ning, W. An Energy Efficiency-oriented Routing Protocol for Wireless Rechargeable Sensor Networks. In Proceedings of the 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2021; pp. 1–5. [\[CrossRef\]](#)
27. Han, G.; Yang, X.; Liu, L.; Zhang, W. A Joint Energy Replenishment and Data Collection Algorithm in Wireless Rechargeable Sensor Networks. *IEEE Internet Things J.* **2018**, *5*, 2596–2604. [\[CrossRef\]](#)
28. Zhao, M.; Li, J.; Yang, Y. A Framework of Joint Mobile Energy Replenishment and Data Gathering in Wireless Rechargeable Sensor Networks. *IEEE Trans. Mob. Comput.* **2014**, *13*, 2689–2705. [\[CrossRef\]](#)
29. Sah, D.K.; Amgoth, T. Renewable energy harvesting schemes in wireless sensor networks: A Survey. *Inf. Fusion* **2020**, *63*, 223–247. [\[CrossRef\]](#)
30. Qi, F.; Zhu, X.; Mang, G.; Kadoch, M.; Li, W. UAV Network and IoT in the Sky for Future Smart Cities. *IEEE Netw.* **2019**, *33*, 96–101. [\[CrossRef\]](#)
31. Xu, J.; Zeng, Y.; Zhang, R. UAV-Enabled Wireless Power Transfer: Trajectory Design and Energy Optimization. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 5092–5106. [\[CrossRef\]](#)
32. Liu, Y.; Pan, H.; Sun, G.; Wang, A.; Li, J.; Liang, S. Joint Scheduling and Trajectory Optimization of Charging UAV in Wireless Rechargeable Sensor Networks. *IEEE Internet Things J.* **2022**, *9*, 11796–11813. [\[CrossRef\]](#)
33. Wu, P.; Xiao, F.; Sha, C.; Huang, H.; Sun, L. Trajectory Optimization for UAVs' Efficient Charging in Wireless Rechargeable Sensor Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4207–4220. [\[CrossRef\]](#)
34. Zhao, C.; Wang, Y.; Zhang, X.; Chen, S.; Wu, C.; Teo, K.L. UAV Dispatch Planning for a Wireless Rechargeable Sensor Network for Bridge Monitoring. *IEEE Trans. Sustain. Comput.* **2023**, *8*, 293–309. [\[CrossRef\]](#)
35. Baek, J.; Han, S.I.; Han, Y. Optimal UAV Route in Wireless Charging Sensor Networks. *IEEE Internet Things J.* **2020**, *7*, 1327–1335. [\[CrossRef\]](#)
36. Lin, C.; Guo, C.; Du, W.; Deng, J.; Wang, L.; Wu, G. Maximizing Energy Efficiency of Period-Area Coverage with UAVs for Wireless Rechargeable Sensor Networks. In Proceedings of the 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Boston, MA, USA, 10–13 June 2019; pp. 1–9. [\[CrossRef\]](#)
37. Hu, H.; Xiong, K.; Qu, G.; Ni, Q.; Fan, P.; Letaief, K.B. AoI-Minimal Trajectory Planning and Data Collection in UAV-Assisted Wireless Powered IoT Networks. *IEEE Internet Things J.* **2021**, *8*, 1211–1223. [\[CrossRef\]](#)
38. Bouhamed, O.; Ghazzai, H.; Besbes, H.; Massoud, Y. A UAV-Assisted Data Collection for Wireless Sensor Networks: Autonomous Navigation and Scheduling. *IEEE Access* **2020**, *8*, 110446–110460. [\[CrossRef\]](#)

39. Liu, R.; Qu, Z.; Huang, G.; Dong, M.; Wang, T.; Zhang, S.; Liu, A. DRL-UTPS: DRL-Based Trajectory Planning for Unmanned Aerial Vehicles for Data Collection in Dynamic IoT Network. *IEEE Trans. Intell. Veh.* **2023**, *8*, 1204–1218. [[CrossRef](#)]
40. Li, K.; Ni, W.; Dressler, F. LSTM-Characterized Deep Reinforcement Learning for Continuous Flight Control and Resource Allocation in UAV-Assisted Sensor Network. *IEEE Internet Things J.* **2022**, *9*, 4179–4189. [[CrossRef](#)]
41. Shan, T.; Wang, Y.; Zhao, C.; Li, Y.; Zhang, G.; Zhu, Q. Multi-UAV WRSN charging path planning based on improved heed and IA-DRL. *Comput. Commun.* **2023**, *203*, 77–88. [[CrossRef](#)]
42. Liu, N.; Zhang, J.; Luo, C.; Cao, J.; Hong, Y.; Chen, Z.; Chen, T. Dynamic Charging Strategy Optimization for UAV-Assisted Wireless Rechargeable Sensor Networks Based On Deep Q-network. *IEEE Internet Things J.* **2023**, *11*. [[CrossRef](#)]
43. Wang, Y.; Feng, Y.; Liu, M.; Liu, N. Dynamic Spatiotemporal Charging Scheduling Based on Deep Reinforcement Learning for WRSN. *J. Softw.* **2023**, *35*, 1485–1501.
44. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
45. Heinzelman, W.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, Hawaii, 4–7 January 2000; Volume 2, p. 10. [[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.