

VoI and Energy-Aware AUV-Assisted Data Collection for Internet of Underwater Things

Jingzehua Xu^{*,†}, Zekai Zhang^{*,†}, Ziyuan Wang[†], Jingjing Wang[‡], Yong Ren[†]

^{*}Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China

[†]Department of Electronic Engineering, Tsinghua University, Beijing, 100083, China

[‡]School of Cyber Science and Technology, Beihang University, Beijing, China

Email: drwangjj@buaa.edu.cn

Abstract—In this study, an autonomous underwater vehicle (AUV) is considered to collect data in Internet of Underwater Things (IoUT) networks. The AUV is tasked with timely visits to sensor nodes (SNs) to collect data using a navigation-hover-communication protocol. Considering the AUV’s limited energy, the dynamic data upload demand of SNs and the diminishing Value of Information (VoI) during data transmission, efficient path planning for AUV is required. Therefore, we formulate a multi-objective optimization problem and employ the deep deterministic policy gradient (DDPG) algorithm to address it. Our objectives encompass the maximization of the sum data rate, the maximization of the sum VoI, and the minimization of the AUV’s energy consumption within specified task time constraints. To mitigate the challenges posed by sparse rewards, we enhance the DDPG algorithm with hindsight experience replay (HER). The simulation results show that the data collection policy trained by our proposed algorithm can converge quickly and has excellent generalization. When the communication range changes, it can still effectively reduce the AUV’s energy consumption while ensuring the quality and timeliness of the data collection task.

I. INTRODUCTION

Recently, IoUT networks have garnered considerable attention due to their great achievements in information collection, underwater search and rescue, military missions and other applications [1]. However, compared with terrestrial wireless communication networks, IoUT networks based on acoustic communication face the shortcomings of low bandwidth and long delay, and are not suitable for long-distance transmission of large amounts of data. Therefore, how to collect data efficiently and reliably in IoUT networks has become a challenge.

Traditional data collection schemes are mostly realized through multi-hop routing between SNs, which will cause additional energy loss of SNs and shorten the life cycle of IoUT networks [2]. Using AUVs as mobile sinks to visit SNs by establishing acoustic links to collect data has become the most potential solution. Most of the existing data collection schemes only take AUV’s sailing distance as the optimization objective [3]. However, these solutions perform poorly in practical applications because they ignore the characteristics of the SNs and do not focus on the quality and value of the data. Besides, the upload requirements of SNs are dynamic and different because they are located in different areas to perform different tasks, and once the data collection is not

timely, the old data will be overwritten by the new data and will lead to data loss [4]. On the other hand, the value of data decreases with time, so it is necessary to plan a reasonable path between the data collection points and the unloading point to avoid unnecessary loss of value [5]. Duan *et al.* adopted the branch bound method to optimize the AUV’s path by taking the sum VoI of the entire sensor network as an index [5]. Fang *et al.* discussed the timeliness of data collection under energy constraints of AUVs based on queuing theory and age of information [1]. However, neither work discusses the data upload priorities of SNs and how to maximize the sum of collected data. In [2], a hybrid data collection scheme is introduced, which comprehensively considers real-time data collection and energy efficiency of AUV, and selects cluster heads to transmit urgent data through multi-hop routing to reduce attenuation of VoI. Unfortunately, such traditional optimization solutions have struggled to fully utilize the AUV’s sensing capabilities to make real-time decisions in the complex underwater environment. Reinforcement learning has been widely applied to various complex tasks, thanks to the strong learning ability gained through interaction with the environment. In [4], the authors designed unmanned aerial vehicle path planning scheme based on reinforcement learning to achieve trade-offs between power supply and data collection of ground sensors.

To solve these problems, this paper propose VoI-energy-aware data collection scheme based on HER-DDPG, which comprehensively considered the data upload priorities of SNs, AUV’s energy consumption, the sum data rate, and the sum VoI of collected data, so as to plan the AUV’s collection path.

The paper is structured as follows. In Section II, the system model is given to elaborate the scenario, model assumptions and basic principles for data collection task. In Section III, we formulate a multi-objective optimization problem and propose HER-DDPG algorithm. In Section IV, we conduct simulation experiments to evaluate the performance of our proposed algorithm, with our conclusions presented in Section V.

II. SYSTEM MODEL

We consider a data collection system for the IoUT network shown in Fig. 1. At the sea bottom, M SNs are deployed, with data upload priorities varying based on regions and missions. During the mission, the AUV starts to visit the SNs within

[†] These authors contribute equally to this work.

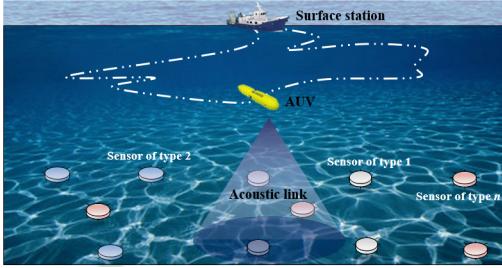


Fig. 1. Illustration of AUV-assisted data collection system.

a specific mission time T to collect data by establishing a short-range acoustic link. Then, the data is summarized to the surface station for unloading. Finally, the surface station transmits the data to shore base station for further processing through marine wireless communication.

A. Sensor Model

We define the set $S = \{1, 2, \dots, M\}$ and S_0 to represent the SNs and surface station, and $[x_i, y_i]$ to represent the position of SN i . SNs collect surrounding data in real time and store it in their data buffer waiting to be collected by AUV. The amount of data waiting to be collected in SN i at time t is $D_i(t)$, and its update equation is defined as [4]:

$$D_i(t + \Delta t) = D_i(t) + \lambda_i \cdot \Delta t, \quad (1)$$

where Δt is the update interval, $0 \leq t \leq T$, $D_i(t) \in [0, D_{\max}]$, D_{\max} is the maximum amount of data that can be stored by the SNs, which is constrained by hardware limitations. $\lambda_i(t)$ represents the data generation rate of SN i at time t , with varying parameters in the Poisson distribution for each SN.

The AUV must collect data from the SNs promptly to avoid data overflow, overwriting, or loss. However, the AUV can only collect data from one SN at a time, which requires considering the data upload priorities of different SNs. We define $q_i^{up}(t)$ to describe the data upload priority of sensor i , which is related to the proportion of the current data in the storage capacity and the data generation rate, specifically defined as Eq. (2):

$$q_i^{up}(t) = \lambda_i(t) \cdot \frac{D_i(t)}{D_{\max}}. \quad (2)$$

B. AUV Dynamics and Energy Consumption Model

Since the AUV collects data in a two-dimensional plane and its state varies discretely over time, the three-degree-of-freedom motion model is derived without loss of generality. The AUV's state at time t is $U(t) = \{p_x(t), p_y(t), v_x(t), v_y(t), a(t), \theta(t), \omega(t)\}$, the elements in the set represent current position, velocity, acceleration, yaw angle, and angular velocity. The motion of AUV in a short time interval can be regarded as uniform motion, then the position update equation is:

$$p_x(t + 1) = p_x(t) + (v_x(t) + v_x(t + 1))\Delta t/2, \quad (3a)$$

$$p_y(t + 1) = p_y(t) + (v_y(t) + v_y(t + 1))\Delta t/2. \quad (3b)$$

Similarly, the velocity update equation is:

$$v_x(t + 1) = v_x(t) + a(t) \cos(\theta(t + 1))\Delta t, \quad (4a)$$

$$v_y(t + 1) = v_y(t) + a(t) \sin(\theta(t + 1))\Delta t. \quad (4b)$$

Moreover, the angle update equation is:

$$\theta'(t + 1) = \theta(t) + \omega(t)\Delta t. \quad (5)$$

Considering that the angle range is $[-\pi, \pi]$, the equation (5) is modified as follows:

$$\theta(t + 1) = \begin{cases} \theta'(t + 1) - 2\pi & \theta'(t + 1) > \pi, \\ \theta'(t + 1) + 2\pi & \theta'(t + 1) < -\pi, \\ \theta'(t + 1) & \theta'(t + 1) \in [-\pi, \pi]. \end{cases} \quad (6)$$

It is assumed that the deceleration and acceleration of AUV are negligible, and the motion in a small time interval can be regarded as uniform motion. Ignoring the energy loss of all components outside the AUV's propeller, that is, all AUV's energy consumption is to overcome the resistance, and the resistance is equal to the thrust because the uniform motion. According to the hydrodynamic formula, the water resistance F_D of AUV is shown in Eq. (7), thus the sailing speed v can be solved as shown in Eq. (8):

$$F_D = (C\rho v^2 S)/2, \quad (7)$$

$$v = (2F_D/C\rho S)^{1/2}, \quad (8)$$

where C is the hydrodynamic coefficient related to propagation medium, the shape of the robot and other factors, and the value is 0.7 according to experience. ρ is water density. S is the AUV's cross-sectional area.

According to the empirical formula, the corresponding relationship between thrust F_T and propulsion power P is:

$$F_T = -0.0021P^2 + 0.6342P + 2.8372. \quad (9)$$

As the motor speed increases, the mechanical efficiency will gradually increase, the relationship between the sailing speed v and the working efficiency η of the propeller is:

$$\eta = F_T v / P = -0.081v^3 + 0.215v^2 - 0.01v + 0.541. \quad (10)$$

Fig. 2 illustrates the relationship between P and v . It is worth noting that in the following, the speed range of AUV is given as [1.3, 2] m/s, the hovering speed is 0 m/s, and the hovering power consumption $P_h = 90$ W.

C. Underwater Acoustic Communication Channel Analysis

AUV collects data from SNs through underwater acoustic communication channel, so we analyze this channel. In the shallow water acoustic propagation environment, the path loss PL of a signal with frequency f at the distance l is:

$$PL(l, f) = l^k a(f)^l, \quad (11)$$

where k is the spread factor, which takes a value of 1.5 in practice, and $a(f)$ is the absorption coefficient in dB/km, as calculated using the Thorp formula [6]:

$$10 \log(a(f)) = 0.11 \frac{f^2}{1+f^2} + 44 \frac{f^2}{4100+f^2} + 2.75 \times 10^{-4} f^2 + 0.003. \quad (12)$$

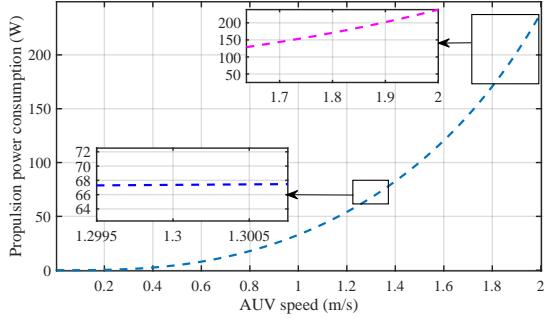


Fig. 2. The AUV's propulsion power consumption P versus speed v .

Underwater environmental noise $N(f)$ is composed of turbulence, ship, wind and thermal noises [7], denoted as $N_t(f)$, $N_s(f)$, $N_w(f)$ and $N_{th}(f)$ respectively:

$$N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f). \quad (13)$$

The noise components in equation (13) can be respectively described as:

$$\begin{cases} 10 \log N_t(f) = 17 - 30 \log f, \\ 10 \log N_s(f) = 30 + 20s + \log(f^{26}/(f+0.03)^{60}), \\ 10 \log N_w(f) = 50 + 7.5w^{1/2} + 20 \log(f/(f+0.4)^2), \\ 10 \log N_{th}(f) = -15 + 20 \log f, \end{cases} \quad (14)$$

where, s is the shipping activity factor, w is wind speed, $s \in [0, 1]$. Thus, narrow-band SNR $\gamma(l, f)$ is related to $PL(l, f)$ and $N(f)$ [8]:

$$\gamma(l, f) = \frac{1}{PL(l, f) \cdot N(f)}, \quad (15)$$

where, $\gamma(l, f)$ is related to distance l and frequency f , which makes channel capacity analysis very difficult. In order to simplify, we assume that there is an optimal frequency $f_o(l)$ for a given communication distance l , and the SNR at this frequency is $\gamma_o(l)$. We define a 3-dB frequency range $[f_L(l), f_U(l)]$, satisfying $\gamma(l, f_L(l)) = \gamma(l, f_U(l)) = \gamma_o(l) - 3dB$ [8]. Assuming that the sensor and AUV transmits information using a narrow-band signal with a center frequency f_c and bandwidth w falling within the 3dB frequency range, the SNR of the true channel can be replaced by the following switching function:

$$\tilde{\gamma}(l) = \begin{cases} \min \left\{ \gamma \left(l, f_c - \frac{w}{2} \right), \gamma \left(l, f_c + \frac{w}{2} \right) \right\}, & f \in w, \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

Assuming that the channel is an additive white Gaussian noise channel, the channel capacity $R(l)$ is [5]:

$$R(l) = w \log_2 \left(1 + \frac{P_T \tilde{\gamma}(l)}{w} \right). \quad (17)$$

D. VoI Definition

The importance of the data collected by the SNs varies, so does the initial value of the data they collect, which can be measured in VoI. The initial VoI for the data from SN i is denoted as E_i , and then the VoI decreases over time after the transmission begins and stops decaying until the end of the

task. $V_i(t)$ is used to track the VoI of data from SN i , defined as follows [5]:

$$V_i(t) = \begin{cases} \alpha E_i + (1 - \alpha) E_i f(t), & t \geq T_i, \\ 0, & \text{otherwise,} \end{cases} \quad (18)$$

where, α is the adjustment parameter used to adjust the weight of event importance and timeliness, T_i is the time for AUV to access sensor i , $f(t) = e^{-(t-T_i)/\beta}$ is the monotone decreasing function, $t > T_i$, β is the scaling factor.

Taking an example, when the AUV visits SN $i+1$ at time T_{i+1} , the VoI for the data collected from SN i is updated as:

$$V_i(T_{i+1}) = \alpha E_i + (1 - \alpha) E_i e^{-\frac{t_{i,i+1}}{\beta}}, \quad (19)$$

where $t_{i,i+1} = t_i^{\text{coll}} + t_i^{\text{sail}}$, t_i^{coll} is the data collection time at SN i , and t_i^{sail} is the travel time of AUV from SN i to $i+1$. They are respectively given by Eq. (20) and Eq. (21), where $l_{i,AUV}$ is the distance between SN i and AUV, $l_{i,i+1}$ is the distance between SN i and SN $i+1$, and $\bar{v}_{i,i+1}$ is the average speed. Then we have

$$t_i^{\text{coll}} = \frac{D_i(t)}{R(l_{i,AUV})}, \quad (20)$$

$$t_i^{\text{sail}} = \frac{l_{i,i+1}}{\bar{v}_{i,i+1}}. \quad (21)$$

When AUV leaves SN i and then visits n SNs in set $S_n = \{i+1, i+2, \dots, i+n\}$, $i \notin S_n$, and AUV finally returns to the surface station (the return time to the surface station is ignored), the VoI is updated as:

$$V_i = \alpha E_i + (1 - \alpha) E_i e^{-\sum_{k=i}^{i+n} \frac{t_{k,k+1}}{\beta}}. \quad (22)$$

III. PROBLEM FORMULATION AND ALGORITHM DESIGN

We formulate the problem and discussed optimization objectives and constraints in this section. Then, the data collection task is modeled as Markov decision process (MDP), and the corresponding reward function is designed. Finally, the HER-DDPG algorithm is proposed for AUV path planning to jointly optimize multiple objectives.

A. Problem Formulation

Our goal is to enable the AUV to sense the IoUT network environment for real-time path planning to maximize the sum data rate and sum VoI, while minimizing the AUV's energy consumption. The AUV visits the SNs successively according to the priority defined in Eq. (2). The communication coverage radius of the AUV is R_r . When the target node falls within the radius, the AUV starts to hover and collect data. Let k , $0 < k \leq K$ indicates the k -th time that the AUV hovers in a task, then the visited SN corresponding to the k -th hover point can be recorded as i^k . Then, the sum data rate R_{sum} , the sum VoI of the collected data V_{sum} and total energy consumption E_c within a given task period T are defined as:

$$R_{sum} = \sum_{k=0}^K R(l_{i^k,AUV}), \quad (23)$$

$$V_{sum} = \sum_{k=0}^K V_{ik}, \quad (24)$$

$$E_c = \int_0^T (\varepsilon(t)P_h + (1 - \varepsilon(t))\frac{P(t)}{\eta(t)})dt, \quad (25)$$

where P_h is the AUV's hovering power, $P(t)$ and $\eta(t)$ are respectively the propulsion power consumption and efficiency of AUV's thruster. These two terms are related to the speed $v(t)$ of AUV, so they are time-varying. In addition, $\varepsilon(t)$ is an indication label, $\varepsilon(t) = 1$ when the AUV is hovering at time t , otherwise 0.

The optimization problem is given as:

$$\max_{v(t), \theta(t)} (R_{sum}, V_{sum}, -E_c), \quad (26)$$

$$\text{s.t. } v(t) \in [0, v_{\max}], \quad (27)$$

$$\theta(t) \in [-\pi, \pi]. \quad (28)$$

B. Markov Decision Process Modeling

The AUV data collection process can be modeled as a MDP, defined by a quintuple [9]:

$$\Omega = (\mathbf{s}, \mathbf{a}, \mathbf{P}, \mathbf{R}, \mu), \quad (29)$$

where \mathbf{s} , \mathbf{a} , \mathbf{R} , represent state space, action space and reward function, \mathbf{P} is state transition probability distribution, μ is discount factor. We pay special attention to the \mathbf{s} , \mathbf{a} and \mathbf{R} of the AUV when designing the algorithm:

1) State Space $\mathbf{s}(t)$: In the data collection task, it is assumed that the AUV can observe its own location information $[p_x(t), p_y(t)]$, the relative distance $[d_x(t), d_y(t)]$ between AUV and target node, the cumulative number of out-of-bounds $N_o(t)$, the number of SNs with data loss $N_l(t)$:

$$\mathbf{s}(t) = \{d_x(t), d_y(t), p_x(t), p_y(t), N_o(t), N_l(t)\}. \quad (30)$$

2) Action Space $\mathbf{a}(t)$: In the process of data collection, the AUV makes action decisions by observing the state, $\mathbf{a}(t)$ is:

$$\mathbf{a}(t) = \{[v(t) \cos(\theta(t)), v(t) \sin(\theta(t))] \}. \quad (31)$$

3) Reward Function $\mathbf{R}(t)$: In reinforcement learning, agents rely on rewards to evaluate strategies for learning. For the multi-objective optimization problem mentioned above, we design the following rewards to encourage AUV to learn data collection strategies, and the reward design is as follows:

$$\mathbf{R}(t) = \omega_{dc}r_{dc}(t) + \omega_{dr}r_{dr}(t) + \omega_Vr_V(t) + \omega_{ec}r_{ec}(t) + \omega_ar_a(t). \quad (32)$$

When the AUV arrives near the specified target node and hovers around for data collection, reward $r_{dc}(t) = 500$ is obtained, and reward $r_{dr}(t)$ represents the sum data rate obtained up to time t . Reward $r_V(t)$ represents the sum VoI obtained until time t . Reward $r_{ec}(t)$ represents the sum energy consumption of AUV. In addition, auxiliary reward $r_a(t)$ is defined as Eq. (33). ω_{dc} , ω_{dr} , ω_V , ω_{ec} and ω_a are the weight factors of the corresponding items. Then

$$r_a(t) = -\omega_{dis}(d_x(t) + d_y(t)) - \omega_oN_o(t) - \omega_lN_l. \quad (33)$$

As can be seen from reward $r_a(t)$, the greater the penalty is when the AUV is farther away from the target node, the more times the AUV crosses the boundary, the more the penalty is, and the more the number of nodes that fail to be accessed by the AUV in time and cause data loss, the more the penalty is.

C. HER-DDPG Algorithm

In the data collection task, AUV will try to make the proper action $\mathbf{a}(t)$ according to its observation $\mathbf{s}(t)$ and strategy $\mu_\theta(\mathbf{a}(t) | \mathbf{s}(t))$ maximum discount reward $J(\theta)$. Based on the reward function $\mathbf{R}(t)$ designed above, the learning objective of reinforcement learning is:

$$\max_{\mu_\theta} J(\theta) = \max_{\mu_\theta} E \left[\sum_{t=1}^{T=\infty} \mathbf{R}(t) \right], \quad (34)$$

where μ_θ stands for the policy of the agent.

The DDPG algorithm is used to solve the learning objective. Assuming that the policy is deterministic, the action can be expressed as $\mathbf{a} = \mu_\theta(\mathbf{s})$. According to [10], we can derive the deterministic policy gradient theorem:

$$\nabla_\theta J(\mu_\theta) = E_{\mathbf{s} \sim v^{\pi_\beta}} \left[\nabla_\theta \mu_\theta(\mathbf{s}) \nabla_\mathbf{a} Q_\omega^\mu(\mathbf{s}, \mathbf{a})|_{\mathbf{a}=\mu_\theta(\mathbf{s})} \right], \quad (35)$$

where v is the distribution of states, π_β represents the behavior policy used to collect data, \mathbf{s} stands for the state of the agent, and $Q(\mathbf{s}, \mathbf{a})$ is the action value function.

To address the sparse rewards during the early stage of training, HER algorithm [11] is used to introduce the concept of goal relabeling into DDPG. Considering the dynamics of the system are the same between the multi-target tasks, the HER algorithm enables the agent to learn from unsuccessful experiences and better handle sparse reward problems. The details of the HER-DDPG are shown in Algorithm 1.

IV. SIMULATION RESULTS

In this section, we perform simulation experiments to showcase the HER-DDPG algorithm's excellent performance for the AUV-assisted data collection task. Firstly, the settings of the simulations are introduced in detail. Finally, the results of experiments are comprehensively analyzed.

A. Experiment Settings

In the simulation experiments, parameters are into two parts and considered respectively: system parameters and algorithm parameters. In IoT network, the number of SNs is set to 20, and these nodes have four types, that is, they have different $\lambda_i(t)$. The AUV and nodes are randomly distributed in a 80m×80m area at the beginning of the task. The data in the sensors is updated over time. Other major parameters of the system and algorithm are shown in Table I and Table II.

B. Simulation Results and Analysis

The simulation results of training AUV with HER-DDPG to perform data collection tasks are shown in Fig. 3-5. To highlight the adaptability of our proposed algorithm, we change the AUV coverage radius R_r from 3m to 5m to conduct different simulations. After 200 episodes of training, the average reward

Algorithm 1 HER-DDPG Algorithm

- 1: Initialize the replay buffer \mathcal{D} , Ornstein-Uhlenbeck random process \mathcal{N} , critic network Q_ω , policy network μ_θ , target network Q_{ω^-} , μ_{θ^-} and parameters ω , θ , $\bar{\omega}$, $\bar{\theta}$.
- 2: **for** episode $k = 1 \rightarrow M$ **do**
- 3: Reset the environment and set the goal ξ_t .
- 4: **for** time step $t = 1 \rightarrow T$ **do**
- 5: Execute action \mathbf{a}_t according to the policy μ_θ and exploration noise \mathcal{N}_t , and collect the next state \mathbf{s}_{t+1} from environment:

$$\mathbf{a}_t \sim \mu_\theta(\mathbf{s}_t \| \xi_t) + \mathcal{N}_t;$$
- 6: Calculate reward r_t and update goal to ξ_{t+1}
- 7: Store the tuple $(\mathbf{s}_t \| \xi_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1} \| \xi_t)$ into \mathcal{D} ;
- 8: Sample N tuples $\{(\mathbf{s}_i \| \xi_i, \mathbf{a}_i, \mathbf{r}_i, \mathbf{s}_{i+1} \| \xi_i)\}_{i=1,\dots,N}$ from \mathcal{D} ;
- 9: **if** episode $k < 0.1M$ **then**
- 10: Select a state \mathbf{s}' in above tuples and map it to a new goal ξ' :

$$\xi' = \phi(\mathbf{s}')$$
- 11: Calculate the new reward r' and replace the original tuple in the same trajectory by $(\mathbf{s}_i \| \xi', \mathbf{a}_i, r', \mathbf{s}_{i+1} \| \xi')$;
- 12: **end if**
- 13: **for** tuple $i = 1 \rightarrow N$ **do**
- 14: $y_i = r_i + \gamma Q_{\omega^-}(\mathbf{s}_{i+1}, \mu_{\theta^-}(\mathbf{s}_{i+1}))$
Minimize the objective loss to update current critic network:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - Q_\omega(\mathbf{s}_i, \mathbf{a}_i))^2$$
- 15: Calculate the policy gradient of tuples to update current policy network:

$$\nabla_\theta J \approx \frac{1}{N} \sum_{i=1}^N \nabla_\theta \mu_\theta(\mathbf{s}_i) \nabla_a Q_\omega(\mathbf{s}_i, \mathbf{a}) \Big|_{\mathbf{a}=\mu_\theta(\mathbf{s}_i)}$$
- 16: Update the target network:

$$\omega^- \leftarrow \tau \omega + (1 - \tau) \omega^-, \quad \theta^- \leftarrow \tau \theta + (1 - \tau) \theta^-$$
- 17: **end for**
- 16: **end for**
- 17: **end for**

curve is shown in Fig. 3, which proves the effectiveness and convergence of our proposed algorithm, indicating that the AUV has learned the optimal strategy for the data collection task. Fig. 4 shows that optimization objectives such as sum data rate, sum VoI, and AUV's energy consumption converge to the steady-state value as the number of training episodes changes under different conditions, indicating that AUV gradually grasps the optimal tradeoff when jointly optimizing these different objectives. In addition, Fig. 5 shows the numerical situation of SNs visited by AUV under different coverage

TABLE I
PARAMETERS OF THE SYSTEM

Parameters	Values
Experimental site size	$80m \times 80m$
Transmit power	30mW
Transmit frequency f	25kHz
Bandwidth w	1kHz
Spreading factor k	1.5
Shipping activity factor s	0.5
Number of sensor nodes M	20
Coverage radius of data collection R_r	3m
Weight to measure the trade α	0.7
Scaling factor for decreasing function of VoI β	10
Event importance E_i	[0,50]

TABLE II
PARAMETERS OF THE ALGORITHM

Parameters	Values
Reward weight of data collection ω_{dc}	500
Reward weight of data rate ω_{dr}	100
Reward weight of VoI ω_V	0.1
Reward weight of energy consumption ω_{ec}	-5
Reward weight of target distance ω_{dis}	-100
Reward weight of crossing boundary ω_o	-10
Reward weight of data overflow ω_l	-5
Number of training episodes	200
Reward discount γ	0.9
Learning rate for actor	0.001
Learning rate for critic	0.001
Soft updating rate τ	0.001

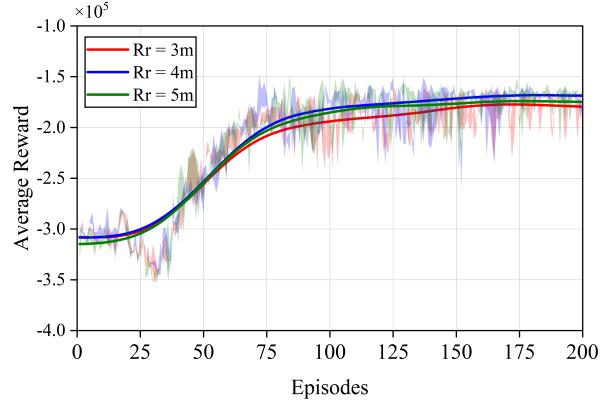


Fig. 3. Average reward versus training episodes.

radius R_r , verifying the rationality of the positive effect of a wider coverage radius.

Then, we use the optimal policy trained by HER-DDPG to guide AUV to complete the data collection task. After a period of operation, the hovering positions of AUV, the SNs visited by AUV and other SNs in the whole process are shown in Fig. 6, where purple star and green star respectively represent the hovering positions of AUV and the SNs visited by AUV. Other stars of different colors indicate SNs of different priority that are not visited. It can be seen that the AUV can select the optimal path to reach the SNs with the highest priority successively to complete data collection.

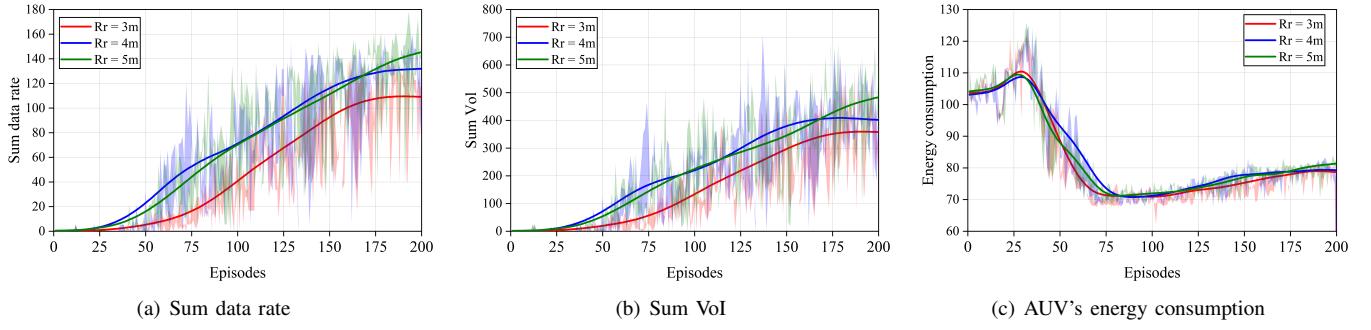


Fig. 4. The change of optimization objectives versus training episodes under different conditions.

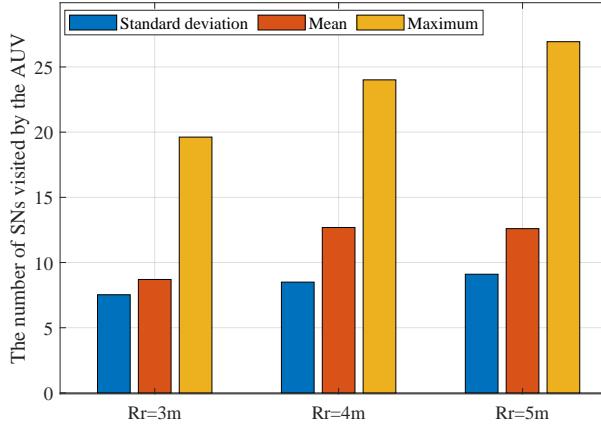


Fig. 5. Analysis of SNs visited by AUVs under different conditions.

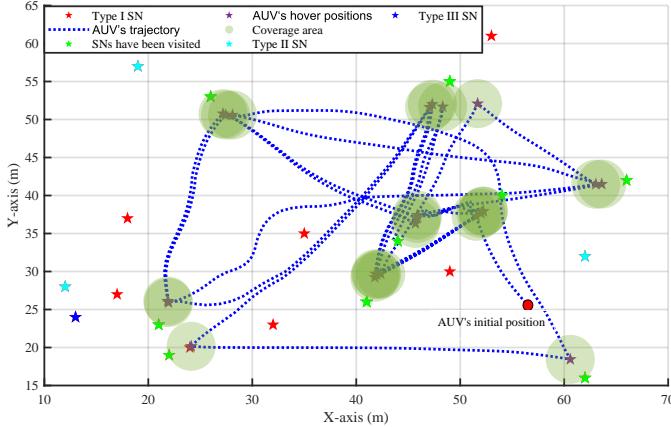


Fig. 6. The trajectory of the AUV in data collection task.

V. CONCLUSION

Aiming at the problems such as the difficulty and poor timeliness of underwater data collection, this paper analyzes underwater acoustic communication links and proposes an efficient AUV-assisted data collection scheme in IoUT network. A multi-objective optimization problem is established to maximize the sum data rate and sum VoI, minimize the AUV's energy consumption, and DDPG is adopted to solve it. To improve the sparse reward problem, we use HER to

improve DDPG and design HER-DDPG algorithm. Numerous simulation results demonstrate that our proposed scheme can optimize various objectives simultaneously and achieves balance.

REFERENCES

- [1] Z. Fang, J. Wang, C. Jiang, Q. Zhang, and Y. Ren, "AoI-inspired collaborative information collection for AUV-assisted internet of underwater things," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14559–14571, Oct. 2021.
- [2] J. Wang, S. Liu, W. Shi, G. Han, and S. Yan, "A multi-AUV collaborative ocean data collection method based on LG-DQN and data value," *IEEE Internet of Things Journal*, (DOI:10.1109/JIOT.2023.3322169), pp. 1–1, Oct. 2023.
- [3] G. Han, X. Long, C. Zhu, M. Guizani, Y. Bi, and W. Zhang, "An AUV location prediction-based data collection scheme for underwater wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6037–6049, June 2019.
- [4] Y. Yu, J. Tang, J. Huang, X. Zhang, D. K. C. So, and K.-K. Wong, "Multi-objective optimization for UAV-assisted wireless powered IoT networks based on extended DDPG algorithm," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 6361–6374, Sep. 2021.
- [5] R. Duan, J. Du, C. Jiang, and Y. Ren, "Value-based hierarchical information collection for AUV-enabled internet of underwater things," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9870–9883, Oct. 2020.
- [6] N. Morozs, W. Gorma, B. T. Henson, L. Shen, P. D. Mitchell, and Y. V. Zakharov, "Channel modeling for underwater acoustic network simulation," *IEEE Access*, vol. 8, pp. 136 151–136 175, Jul. 2020.
- [7] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, Oct. 2007.
- [8] Z. Fang, J. Wang, C. Jiang, J. Du, X. Hou, and Y. Ren, "Heterogeneous multi-AUV aided green internet of underwater things," in *IEEE International Conference on Communications*, Montreal, QC, Canada, Jun. 2021, pp. 1–6.
- [9] P. Jiang, S. Song, and G. Huang, "Attention-based meta-reinforcement learning for tracking control of AUV with time-varying dynamics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6388–6401, November 2022.
- [10] J. Yan, L. Zhang, X. Yang, C. Chen, and X. Guan, "Communication-aware motion planning of AUV in obstacle-dense environment: A binocular vision-based deep learning method," *IEEE Transactions on Intelligent Transportation Systems*, (DOI:10.1109/TITS.2023.3296415), pp. 1–17, Jul. 2023.
- [11] L. F. Vecchietti, M. Seo, and D. Har, "Sampling rate decay in hindsight experience replay for robot control," *IEEE Transactions on Cybernetics*, vol. 52, no. 3, pp. 1515–1526, Mar. 2022.