Efficient Data Collection Scheme for Multi-Modal Underwater Sensor Networks Based on Deep Reinforcement Learning

Shanshan Song, Jun Liu, *Member, IEEE*, Jiani Guo, Bin Lin, *Senior Member, IEEE*, Qiang Ye, *Senior Member, IEEE*, and Junhong Cui

***Abstract*—Autonomous Underwater Vehicles (AUVs) with multi-modal transmission can achieve high efficient data collection for underwater sensor networks. However, multi-modal transmission and trajectory planning impose great challenges on data collection in complex underwater environments. Most prior studies focus on design of multi-modal architecture, but lack of available implementation and consideration of AUVs' trajectory. Meanwhile, existing trajectory planning research cannot work well on data collection with multiple complex tasks simultaneously. In this paper, an efficient Data Collection scheme for Multi-modal underwater sensor networks based on Deep reinforcement learning (DCMD) is proposed to solve the above challenges. We first propose an optimal multi-modal transmission selection algorithm that provides an implementation to improve transmission efficiency. Then we propose a distributed multi-AUVs' trajectory planning algorithm based on deep reinforcement learning by AUVs' collaborations, considering transmission situation, ocean currents and underwater obstacles, to maximize collection rate and energy efficiency. In addition, we joint transmission and trajectory planning in a protocol to improve collection efficiency. Extensive experimental results show that DCMD achieves better performance on efficiency and reliability than four state-of-the-art methods, demonstrating its great advantage on collecting data for USNs.**

***Index Terms*—Underwater sensor networks, data collection, deep reinforcement learning, multimodal, trajectory planning.**

I. Introduction

UNDERWATER Sensor Networks (USNs) have been widely applied for underwater applications [1], [2], such as marine monitoring, military reconnaissance, hydrology surveys, oil and gas drilling. Sensor data collection is vital for USNs to carry out these applications [3].

In conventional USNs, data collection adopts acoustics as communication mean to upload sensor data [4], as shown in Fig. 1(a). However, as high energy consumption, large latency, and low transmission speed of underwater acoustic communication, USNs with natures of finite power are unsuitable to update a mass of sensing data. Recently, USNs employ an Autonomous Underwater Vehicle (AUV) to collect data with its mobility shown in Fig. 1(b), which greatly reduces energy consumption of sensor nodes for extending network lifetime [5], but efficiency and coverage range of data collection are still restricted by AUV's finite battery power. Multiple AUVs with multi-modal transmission can improve efficiency and geographical range of collection with their cooperation shown in Fig. 1(c).

However, data collection with multiple AUVs is still challenged by multi-modal transmission and trajectory planning in

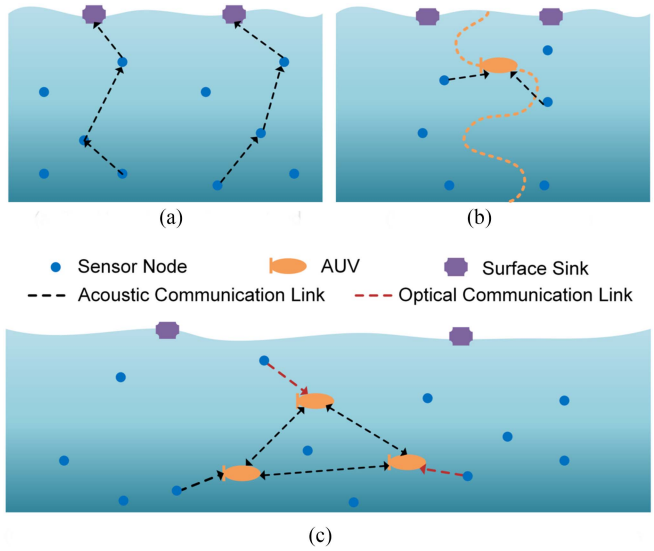


Fig. 1. Comparison of underwater data collection methods. (a) Traditional methods utilize sensor nodes to upload data by multi-hops, using acoustic communication. (b) Signal AUV collects USN's data with its mobility, using acoustic communication. (c) Multiple AUVs collect USN's data efficiently with collaborations, using multi-modal communication, e.g. acoustic communication, and optical communication.

Manuscript received 9 August 2022; revised 3 November 2022; accepted 21 December 2022. Date of publication 27 December 2022; date of current version 18 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 62101211, 61631008, and 61971206, in part by the National Key Research and Development Program of China under Grants 2021YFC2803000 and 2019YFE0111600, and in part by LiaoNing Revitalization Talents Program under Grant XLYC2002078. The review of this article was coordinated by Dr. Lin X. Cai. *(Corresponding author: Jun Liu.)*

Shanshan Song, Jiani Guo, and Junhong Cui are with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: songss@jlu.edu.cn; jnguo20@mails.jlu.edu.cn; jun-hong_cui@jlu.edu.cn).

Jun Liu is with the College of Computer Science and Technology, Jilin University, Changchun 130012, China, also with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China, and also with the Robotics Research Center, Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: liujun2019@buaa.edu.cn).

Bin Lin is with the College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China (e-mail: binlin@dlmu.edu.cn).

Qiang Ye is with the Department of Computer Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada (e-mail: qiangy@mun.ca).

Digital Object Identifier 10.1109/TVT.2022.3232391

TABLE I
NOTATIONS IN THIS ARTICLE

Notation	Description
SNR_o	Signal noise ration of optical communication
SNR_a	Signal noise ration of acoustic communication
N	Number of AUVs
K	Number of sensor nodes in USNs
x_t^i, y_t^i, z_t^i	AUV i's location at time t
$v_x^{i,t}, v_y^{i,t}, v_z^{i,t}$	AUV i's velocity at time t
h_t^i	AUV i's heading angle at time t
$\tilde{v}_x^{i,t}, \tilde{v}_y^{i,t}$	Ocean currents' velocity at the location of AUV i at time t
$\tilde{h}_t^i$	Ocean currents' heading angle at the location of AUV i at time t
o_t^i, a_t^i, r_t^i	AUV i's observation, action and reward at time t
s_t, o_t, a_t	All AUVs' state, observation, action at time t
S, O, A	State, observation, action space
p_t^i	AUV i's penalty at time t
$c_t(k)$	Collection rate of sensor node k at time t
e_T^i	AUV i's energy consumption at time T
γ	discounted factor
π_θ	control policy

TABLE II
COMPARISON OF MULTIPLE UNDERWATER COMMUNICATION MEANS

Means	Opticals	Magnetics	Acoustics
Data rate	Gbps	Mbps	Kbps
Wave speed	2.25×10^8m/s	3×10^7m/s	1.5×10^3m/s
Effective range	10-200m	10-100m	5-100km
transmission power	mW	MW	W
Characteristic	Scattering, line of sight, light noise	Conductivity, attenuation	Multi-path, Doppler, acoustic noise

complex underwater environments, as they two are fundamental and interact [7]. Although plenty of multi-modal transmission research and trajectory planning algorithms have been proposed, they are not specially designed for underwater data collection, and without consideration of the mutual influence between multi-modal transmission and trajectory planning.

Existing multi-modal transmission generally employs complementary acoustic, optical and magnetic transmission means to improve communication and collection efficiency [6]. However, these methods choose transmission mean without consideration of AUV's trajectory. When AUVs change acoustic to another transmission mean for collection, the trajectory changes simultaneously. Most prior research employs software-defined technology to address multi-modal scalability and flexibility issues, but most of them focus on the architecture, lack of implementation details and consideration of trajectory planning.

Meanwhile, existing trajectory planning approaches preset trajectory of AUVs for data collection based on offline optimization [8], [9]. However, they ignore unique underwater environments, such as ocean currents and underwater obstacles, which lead to large trajectory errors. Some methods can dynamically plan AUVs' collection path, such as artificial potential field, genetic, ant colony, and neural network-based algorithms, but they require a precise ocean current or mobility model, and definite pay-off function, which are not suitable for complex tasks. Our problem is complicated with many objectives to achieve simultaneously, including high-efficient data collection rate, energy consumption, multi-modal transmission, obstacle avoidance, cooperation among AUVs, etc.

Recent research employs Deep Reinforcement Learning (DRL) to solve complex decision-making issues [10]. Some methods use DRL to plan AUVs trajectory [11], [12], [13], but they are designed to maximize coverage area or target tracking, cannot be directly applied for underwater data collection because the different task objectives.

This paper proposes a Data Collection scheme for Multi-modal underwater sensor networks based on Deep reinforcement learning (DCMD), which utilizes multi-modal transmission and AUVs' collaborations to achieve higher-efficient data collection with lower-cost. The main contributions of this paper are summarized as follows:

- This is the first work designing an efficient data collection scheme for USNs that joints multi-modal transmission and trajectory planning simultaneously.
- A novel multi-modal transmission selection algorithm is proposed to achieve high quality of transmission service for underwater data collection. The optimization for selecting transmission mean and mode is formulated as an Integral Linear Programming (ILP) problem, then we solve it considering the influence of AUVs' trajectory, distance, and water turbidity on Signal Noise Ratio.
- Then a novel distributed deep reinforcement learning-based multiple AUVs trajectory planning algorithm is proposed specifically designed for underwater data collection, with consideration of multi-modal transmission situation, underwater obstacles and ocean currents, to achieve high-efficient collection rate and energy consumption. In addition, we joint transmission selection and trajectory planning in a protocol, that focuses on maximizing efficiency of collection, energy and transmission.
- Extensive simulations results demonstrate that DCMD achieves better collection efficiency than the state-of-the-art baselines in [11], [12], [13], and Random.

II. RELATED WORK

A. Multi-Modal Transmission Network

Recently, several research pays particular attention to multi-modal underwater communication means, because a single one is insufficient for high-efficient data collection [15], [16]. We present multiple underwater communication means, include acoustics, opticals, and magnetics in Table II. As shown in Table II, the data rate of optical communication is Gbps with wave speed 2.25×10^8m/s, and the data rate of magnetic communication is Mbps with wave speed 3×10^7m/s, which are much higher than the one of acoustic communication. The high data rate of opticals and magnetics can obviously decreases transmission latency and improve data collection performance,

but they are only suitable for short-distance communication (opticals 10-200 m, magnetics 10-100 m), while acoustics can reach 5-100 km. In addition, optical communication may be disabled even over short distance when water is turbid. Therefore, multiple underwater communication means are on demand for data collection.

However, traditional USNs' architecture only fits for a single communication mean, which is not scalable and flexible [17]. This imposes great challenges for adopting multiple communications, induces barriers to integrate heterogeneous underwater communication devices.

The paradigm of Software-Defined Networking (SDN) has been emergent as a revolutionary technology, which enables USNs' architecture to be software-based, service-oriented and programmable. Vasilescu et al. first proposed an underwater sensor networks architecture for long-term monitoring of coral reefs and fisheries, which integrates optical and acoustic communications [18]. Then Akyildiz et al. proposed a software-defined architecture to facilitate the development of next-generation underwater communication systems [19]. It can provide truly differentiated and scalable networking services.

However, the specific mechanism of these architectures still need to be refined, such as transmission mean selection, data transition scheduling. Lin et al. proposed a software-defined beaconing framework for software-defined architecture, which integrates two categories of defined beacons to synchronize network information and execute network operations [20], but it is centralized and without consideration of trajectory planning.

B. Trajectory Planning

Previous studies focus on trajectory planning of AUVs for underwater data collection to improve network lifetime and collection efficiency [21]. Some methods employ variants of traveling salesman problem to plan AUVs' trajectory in static or fixed environments [15]. However, in dynamic underwater environments, such as obstacles or ocean currents with temporal and spatial change, they may not work well.

Some studies focus on dynamic real-time trajectory planning [22], which mainly employ ant colony [23], fuzzy algorithm [24], genetic optimization [25], artificial potential field [26], and machine learning [27]. However, they may fail to find feasible paths because local minimum or oscillate traps for complex underwater tasks, including collection rate, energy consumption, multi-modal transmission, obstacle avoidance, cooperation among AUVs, etc.

Recently, reinforcement learning brings new solution for AUVs trajectory planning, utilizing its feasibility in complex underwater environments. Chu et al. proposed a deep reinforcement learning path planning method based on double deep Q network (IDDQN) to improve AUV's path planning capability in unknown environments [11]. Li et al. proposed a path planning model for AUVs based on deep reinforcement learning (APPD) using WL interpolation surface to model the seabed [12]. Although these machine learning methods are feasible to suit dynamic ocean environments, they need sonar image to build underwater map, or only provide discrete action for AUVs. Yang et al. proposed a cooperative multi-agent reinforcement learning based tracking algorithm following a centralized training with distributed execution (CT-DE) manner [13]. However, they are not specially designed for underwater data collection. Importantly, cooperation between AUVs are not exploited, which affects collection efficiency of them.

III. System Model and Problem Formulation

A. System Model for Data Collection

We define $N \triangleq i = 1, 2, \ldots N$ AUVs and $K \triangleq k = 1, 2, \ldots K$ sensor nodes in USNs. The task of N AUVs is to collect K sensor nodes' data by controlling their moving velocity and orientation in an efficient manner. We aim to maximize the collection efficiency and minimize the energy consumption of AUVs. The energy consumption of an AUV comes from its movement and transmission. We assume AUV i's energy consumption e_t^i at time t is denoted as:

$$e_t^i = \epsilon_{m,t}^i + \epsilon_{c,t}^i, \tag{1}$$

where $\epsilon_{m,t}^i$ and $\epsilon_{c,t}^i$ are the moving energy consumption and transmitting energy consumption for AUV i at time t, respectively. The transmitting energy consumption of an AUV depends on whether it communicates and its communication mode at time t. We set $\epsilon_{c,t}^i = 0$ if AUV i without transmitting. The moving energy consumption of an AUV depends on its moving state. Until time t, the energy consumption of AUV i is $\sum_{\tau=1}^{t} e_\tau^i$

During collection, AUV i moves with its velocity $v_x^{i,t}, v_y^{i,t}, v_z^{i,t}$, heading angle h_t^i, and its movement is affected by ocean currents. AUVs typically cruise at fixed depth for collection. As the depth can be easily obtained with a depth meter, the z-axis velocity of an AUV is easy to obtain. In our model, we decide x and y axes velocity and orientation of AUVs and ocean currents. AUV i's movement is also affected by ocean currents. The velocity of the ocean current is $\tilde{v}_x^{i,t}, \tilde{v}_y^{i,t}$, and its heading angle is $\tilde{h}_t^i$. There are obstacles in the underwater environment.

B. Problem Formulation

We consider two metrics for performance evaluation of DCMD simultaneously: collection rate c_t and energy efficiency η_t, The collection rate is defined as:

$$c_t = \frac{\sum_{k=1}^{K} c_t(k)}{K}, \tag{2}$$

where $c_t(k)$ represents the value whether sensor node k' has been collected. If collected, $c_t(k) = 1$. Otherwise, $c_t(k) = 0$.

The energy efficiency is denoted as:

$$\eta_t = \frac{c_t K}{\widetilde{e}_t}, \tag{3}$$

where $\widetilde{e}_t$ is normalized and averaged from $e_t = \sum_{i=1}^{N}(\sum_{\tau=1}^{t} e_\tau^i)$, which is consumed by all AUVs' movement and transmission until time t, and c_t is also the collection rate by all AUVs until time t.

IV. Distributed Multi-Modal Transmission-and-Trajectory Protocol

A. Optimal Multi-Modal Transmission Selection

1) Optimal Object: We aim to minimize the energy consumption e_t^i of AUV i with multi-modal transmission means $s = 1, 2, \ldots S$ and modes $y = 1, 2, \ldots Y$, because different transmission means and modes transmit data with different power. For example, as the data rate of optical communication is Gbps, it costs lower energy with shorter transmission delay than acoustic communication, as in Table II. Meanwhile, one transmission mean can have multiple transmission modes. For example, the DSP-based acoustic modem [28] has SISO, SIMO and MIMO three transmission modes to satisfy different demands of transmission distance and speed. We formulate it as an integral linear programming problem, aiming to optimize the energy consumption of AUV i by appropriate transmission mean s and mode y, as:

$$\min \sum_{\tau=1}^{t} \left(\epsilon_{m,t}^i + \left(\sum_{s=1}^{S} x_{s,t} P_s^y \right) \right), \quad \text{s.t.} \begin{cases} \sum_{s=1}^{S} x_{s,t} = 1 & s = 1, 2, \ldots, S \\ x_{s,t} \in \{0, 1\} & s = 1, 2, \ldots, S \end{cases} \tag{4}$$

where $\epsilon_{m,t}^i$ represents energy consumption of AUV i's mobility. $\epsilon_{m,t}^i$ depends on the resistance of AUV i when traveling with ocean currents, which is proportional to the square of AUV's velocity at time t. P_s^y represents energy consumption of AUV i's communication with transmission mean s and mode y. Transmission modes represent transmission speeds and transmission power to some extend. For example, optical transmission costs mW, and acoustic transmission costs W according to its transmission mode. $x_{s,t}$ represents the adopted transmission mean of AUV i at time t with an integer values of 0 or 1. When mean s is adopted at time t, then $x_{s,t} = 1$, otherwise $x_{s,t} = 0$.

2) Implementation Details: Because Signal Noise Ratio (SNR) is the most important factor in the decision of transmission mean and mode, we evaluate the influence of trajectory, distance and water turbidity on SNR, then we choose the suitable transmission mean and mode by consideration of distance and SNR to maximize the optimal object mentioned above it. Note that, as the instability of magnetic communication, we employ acoustic and optical communication as our alternate means for transmission.

Trajectory is the AUV's locations for each time, it depends on AUV's velocity and orientation. The trajectory of an AUV is the basis of distance calculation.

Distance between sender i and receiver j is calculated by locations, denoted as d_{ij}.

Water turbidity severely effects on quality of underwater optical communication medium. It is evaluated by extinction coefficient $c(\lambda)$. which effects on the SNR of optical communicaiton. There are mainly four types of water turbidity: pure, clean, coastal, and turbid water. In complex underwater environments, optical signal suffers from absorption and scattering [33]. $c(\lambda)$ is governed by the absorption $a(\lambda)$ and scattering coefficient $b(\lambda)$, formulated as:

$$c(\lambda) = a(\lambda) + b(\lambda), \tag{5}$$

λ is light wavelength. The absorption coefficient $a(\lambda)$ is [34]:

$$a(\lambda) = a_w(\lambda) + a_c(\lambda) C_c + a_f C_f e^{-k_f \lambda} + a_h C_h e^{-k_h \lambda}, \tag{6}$$

where $a_w(\lambda)$ is the absorption coefficient of pure sea water. The absorption coefficient of pure sea water is the smallest. $a_c(\lambda)$ is the absorption coefficient of chlorophyll. C_c, C_f and C_h are the concentrations of chlorophyll, fulvic and humic acids, respectively. a_f and a_h are the specific absorption coefficients of fulvic and humic acid. k_f and k_h are constants.

The scattering coefficient $b(\lambda)$ is as follows:

$$b(\lambda) = b_w(\lambda) + b_s(\lambda) C_s + b_l(\lambda) C_l, \tag{7}$$

where $b_w(\lambda)$ is the scattering coefficient of pure sea water. We consider two cases scattering: small particles and large particles. $b_s(\lambda)$ is the specific scattering coefficients for small particles, and $b_l(\lambda)$ is for large. C_s and C_l are the concentrations for small and large particles. a_f and a_h are the specific absorption coefficients of fulvic and humic acid. k_f and k_h are constants.

SNR is vital for transmission means and modes. For short-range communication, SNR helps to choose from optical and acoustic communication. We denote Υ_o and Υ_a for SNR of optical communication and acoustic communication, respectively, which are formulated as:

$$\Upsilon_o = 10 \lg \left(\frac{S_o}{N_o} \right), \quad \Upsilon_a = 10 \lg \left(\frac{S_a}{N_a} \right), \tag{8}$$

where S_o is efficient received optical power, and N_o is ambient light noise for optical communication. S_a is received acoustic power, and N_a is environment acoustic noise.

The efficient received power of optical signal can be formulated as [35]:

$$S_o = \varepsilon^2 \tilde{S}^2, \tag{9}$$

where ε is the conversion efficiency from optical to electrical signals. $\tilde{S}$ is the optical power of receiver, which is evaluated by distance d_{ij} of node i and j, and extinction coefficient $c(\lambda)$, as follows:

$$\tilde{S} = E \varphi_s \varphi_r e^{-c(\lambda) d_{ij}} \frac{R_r}{2\pi d_{ij}^2}. \tag{10}$$

It is obviously that the distance between sender and receiver affects the received optical power $\tilde{S}$, and the sender's optical transmission power E changes by send and received optical efficiency φ_s and φ_r. R_r is the receiver's aperture area.

For the noise N_o in optical environments, shot noise σ_s^2 and thermal noise σ_t^2 are generally considered as noise source. Therefore, the noise is formulated as:

$$N_o = \sigma_s^2 + \sigma_t^2. \tag{11}$$

The shot noise σ_s^2 can be denoted as:

$$\sigma_s^2 = 2m\varepsilon B(P_{r,s}) + 2m I_c I_b B, \tag{12}$$

where m is the electronic charge, I_c is background current, and I_b is noise bandwidth factor. The noise bandwidth is B. Another

source of N_o is σ_t^2, denoted as:

$$\sigma_t^2 = \frac{8\pi k T_k}{V}\eta A_r I_b B^2 + \frac{16\pi^2 k T_k F}{g_m}\eta^2 A_r^2 I_n B^3, \quad (13)$$

where k is the Boltzmann's constant, T_k is the absolute temperature, V is the open-loop voltage gain, η is the fixed capacitance of photo detector per unit area, F is the FET channel noise factor, and g_m is the FET transconductance.

By the formulas above, Υ_o can be derived as:

$$\Upsilon_o = 10\lg\left(\frac{\varepsilon^2\left(E\varphi_s\varphi_r e^{-c(\lambda)d_{ij}}\frac{R_r}{2\pi d_{ij}^2}\right)^2}{2m\varepsilon B(\tilde{S})+\mu}\right), \quad (14)$$

where μ is the representation of σ_t^2 and the rest part of σ_s^2, denoted as:

$$\mu = 2mI_c I_b B + \frac{8\pi k T_k}{V}\eta A_r I_b B^2 + \frac{16\pi^2 k T_k F}{g_m}\eta^2 A_r^2 I_n B^3. \quad (15)$$

The received power of acoustic signal S_a can be formulated as:

$$S_a = \frac{\tilde{E}}{A(d_{ij}, f)}, \quad (16)$$

where $\tilde{E}$ is the acoustic power of sender, $A(d_{ij}, f)$ is the path loss, formulated by:

$$A(d_{ij}, f) = d_{ij}^u a(f)_{ij}^d, \quad (17)$$

where u is spreading factor, and $a(f)$ is absorption coefficient which is determined by the frequency of acoustic signal f.

For the noise N_a in acoustic environments, we mainly consider currents, surface ship, wind, and thermal noise, and formulates it as:

$$N_a = N_c + N_s + N_w + N_t, \quad (18)$$

where N_c, N_s, N_w, N_t are currents, surface ship, wind, and thermal noise, respectively. They are affected by acoustic signal frequency.

By the formulation above, we can get SNR for optical and acoustic communications. The process to select transmission mean and mode is in algorithm 1. If d_{ij} is over the threshold of optical transmission range Γ, acoustic communication is the only selection, that is $x_{2,t} = 1$. Considering the multiple modulation modes of an acoustic communication device, we further decide specific mode by comparing the SNR of acoustic communication Υ_a and multiple acoustic modes' threshold intervals ϕ_a, which reflect channel condition. Different transmission speeds represents different communication modes. Otherwise, distance d_{ij} is less than threshold Γ, we consider the SNR of optical communication Υ_o, if Υ_o is greater than minimum SNR of optical communication $S_{\min}$, we can choose optical communication $x_{1,t} = 1$.

Algorithm 1: Selection for Transmission Mean and Mode.

Input: Distance d_{ij} between node i and node j at time t, Υ_a, and Υ_o
Output: Communication mean s, and its mode y at time t

```
if d_ij > Γ then
    x_{2,t} = 1, select acoustic communication
    Then compare Υ_a with φ_a(m)
    for i < n do
        if d_ij/Υ_a ∈ φ_a(m) then
            y ← m, select mode m for acoustic communication
        end if
        i ← i + 1
    end for
else
    if Υ_o > S_min then
        x_{1,t} = 1, select optical communication
    else
        x_{2,t} = 1, y = min(m)
    end if
end if
return s, y
```

B. Distributed Multi-AUVs Trajectory Planning Based on Deep Reinforcement Learning

We then propose a distributed multi-AUVs path planning algorithm tightly coupled with transmission situation based on Deep Reinforcement Learning (DRL) to guarantee high-efficiency data collection and energy consumption for USNs. Each AUV has its own inherited control logic to determine their trajectories for data collection in a fully distributed manner. We specifically design the state, observation, action space, and reward for AUV's DRL model.

1) Model Design: The task of AUVs is to travel to collect K sensor nodes' data by controlling its velocity and orientation in an cooperative manner. AUV i observes environment o_t^i to determine its action a_t^i at timeslot t.

Observation Space: Considering the local observation capacity, observation space o_t^i consists of: AUV i' locations x_t^i, y_t^i, AUVs' velocity $v_x^{i,t}, v_y^{i,t}$, heading angle h_t^i, and energy consumption e_t^i. Due to the great effect on trajectory planning, transmission situation Υ_o and Υ_a is added as an important observation. As ocean currents affect AUV's energy efficiency, velocity $\tilde{v}_x^{i,t}, \tilde{v}_y^{i,t}$ and heading angle $\tilde{h}_t^i$ of ocean currents are also considered. AUV i observes $o_t^i| = x_t^i, y_t^i, v_x^{i,t}, v_y^{i,t}, h_t^i, e_t^i, \Upsilon_o, \Upsilon_a, \tilde{v}_x^{i,t}, \tilde{v}_y^{i,t}, \tilde{h}_t^i$ at each timeslot. Therefore, observation space is formulated as $O \triangleq o_t^i | i \in N, t = 1, 2, \ldots, T$.

State Space: Because multiple AUVs apply their collaborations to collect data to improve efficiency, the information of collected sensor nodes is essential for each AUV to determine their trajectories in a distributed manner. Although our problem is a POMDP, the environment is fully observable for AUVs, and state space includes all AUVs' observations with the collection rate of all sensor nodes. The collected rate measures whether a specific sensor node has been collected by any AUV in the past time t. If a sensor node accesses the close range of any AUV and has been collected data by this AUV, we refer the

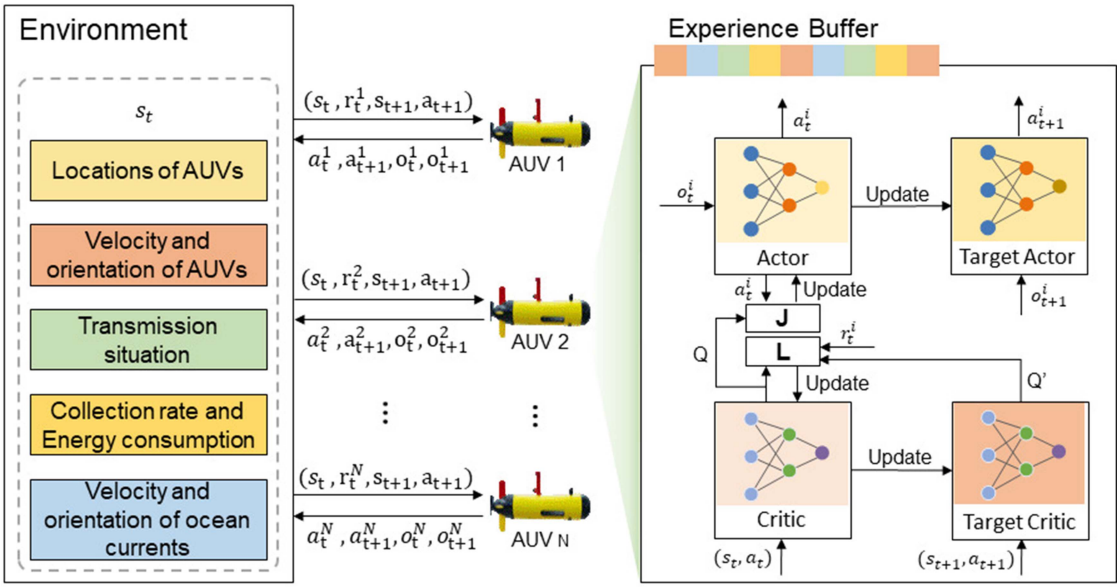


Fig. 2. Model Training of an AUV.

sensor node as "collected", otherwise not collected. Combining collection rate and observation space, the state space is $S \triangleq s_t = O \cup c_t(k) | \forall k = 1, 2, \ldots K$.

Action Space: For our scenario, the action of an AUV mainly means its mobility, which is measured by two core variables: velocity and orientation, which are both continuous. Therefore, the action of AUV i is $a_t^i = v_x^{i,t}, v_y^{i,t}, h_t^i$, and $a_t = a_t^i | i \in N$. The action space of all AUVs is $A \triangleq a_t | t = 1, 2, \ldots T$.

Reward Function: A reward function is defined to achieve maximizing collection rate, minimizing energy consumption. Meanwhile, we consider ocean currents' influence on AUV's mobility to reduce energy consumption. Obstacle avoidance are taken into account as AUV's important function. The reward function has penalty and reward part.

For penalty, each AUV is prohibited from traveling outside the coverage range of USNs or losing connectivity to all remaining AUVs within its max communication range R, which ensure the collaborations between AUVs. AUVs have penalty as hitting an obstacle to guarantee its safety. The penalty is denoted as:

$$\zeta_t^i = \begin{cases} \zeta_o, & \text{hit obstacles} \\ \zeta_l, & \text{travel outside or lose connection} \end{cases}. \tag{19}$$

The reward mainly considers AUVs' collection rate and energy consumption, thus reward function is formulated as:

$$r_t^i = \frac{\rho_1 \sum_{k=1}^{K} \triangle c_t(k)}{\sum_{i=1}^{N} \triangle e_t^i} - \rho_2 [(v^{i,t} - v^{i,t-1})^2 + (h_t^i - h_{t-1}^i)^2] - \zeta_t^i, \tag{20}$$

where $\triangle c_t(k) = c_t(k) - c_{t-1}(k)$ is the collection rate increment of sensor node k, and $\triangle e_t^i = e_t^i - e_{t-1}^i$ is the energy consumption during time $t-1$ and t. ρ_1 and ρ_2 represent weights of their corresponding reward and penalty. $\frac{\sum_{k=1}^{K} \triangle c_{tr}(k)}{\sum_{i=1}^{N} \triangle e_t^i}$ measures the increment of the accumulative reward. By r_t^i, we have $r_t = r_t^i | i \in N$.

2) Training and Testing: During training, each AUV acquires to obtain global environment s_t and all AUVs' action a_t by AUVs' communication. AUVs cooperate to transmit their local observation to others. The training process is shown in Fig. 2.

Each AUV has four neural networks: actor network, target actor network, critic network, target critic network. Critic network evaluates the value of action by all AUVs' state s_t and actions a_t, and actor network modify the probability of the selected action based on the values of critic network. Target critic network can be updated by critic network to evaluate next time value of action by all AUVs' state s_{t+1} and actions a_{t+1}

We first initialize N critic networks $Q^i(\cdot)$ and N actor networks $\mu^i(\cdot)$ with parameters θ^{Q^i} and θ^{μ^i} for all the AUVs. Then N target critic networks and N target actor networks are initialized, which are the copy of their corresponding actor and critic networks with parameters $\theta^{Q'^i} = \theta^{Q^i}$ and $\theta^{\mu'^i} = \theta^{\mu^i}$, as shown in Fig. 3.

Then we update the critic network and actor network separately. The critic network of AUV i minimizes the loss function $L(\theta^{Q^i})$, then update as:

$$L(\theta^{Q^i}) = \frac{1}{M} \sum_{j=1}^{M} (j_j^i - Q^i(s_j, a_j | \theta^{Q^i}))^2, \tag{21}$$

where M is a randomly sampled mini-batch from experience reply buffer. B is the size of experience reply buffer stored B transitions, including state, action, and reward. y_j^i is the target value of critic's target network, formulated as:

$$y_j^i = r_j^i + \gamma Q'^i \left(s_{j+1}, a_{j+1} | \theta^{Q'^i} \right) |_{a_{j+1}^i = \mu'^i (o_{j+1}^i | \theta^{\mu'^i})}. \tag{22}$$

The parameters of target critic network Q'^i is updated as:

$$\theta^{Q'^i} = \tau \theta^{Q^i} + (1 - \tau) \theta^{Q'^i}. \tag{23}$$

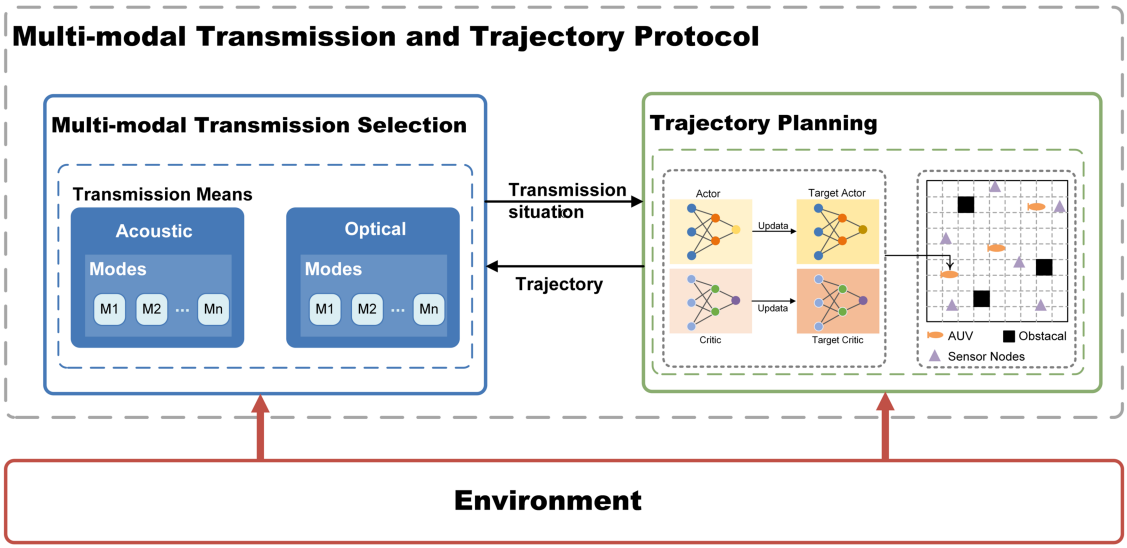


Fig. 3. Protocol Joint Multi-modal Transmission and Trajectory Planning for USNs Data Collection.

Then the critic network feeds back to actor network. Each AUV's actor network weights θ^{μ^i} are updated by the policy gradient method, formulated as:

$$\nabla_{\theta^{\mu^i}} J(\theta^{\mu^i}) \approx \frac{1}{M}\sum_{j=1}^{M} \nabla_{\theta^{\mu^i}} \mu^i(o_j^i|\theta^{\mu^i}) \nabla_{a_j^i} Q^i(s_j, a_j)|_{a_j^i=\mu^i(o_j^i|\theta^{\mu^i})}, \quad (24)$$

where o_j^i represents AUV i's own observation. The parameters of target actor network $\mu^{'i}$ are updated as:

$$\theta^{\mu^{'i}} = \tau\theta^{\mu^i} + (1-\tau)\theta^{\mu^{'i}}. \quad (25)$$

We demonstrate the process for training in algorithm 2. As the training process needs global state and action information, we design the training process is a combination of online and offline.

During testing, the decision of each AUV is distributed, as each AUV only needs to observe its local environment o_t^i to obtain its following action a_t^i, which avoids communication overheads to obtain global observation.

Algorithm 2: Training for AUV i.

Input: State s_t
Output: Action a_t^i

```
Initialize critic and actor network Q^i(s_t, a_t|θ^{Q^i}) and
  μ^i(o_t^i|θ^{μ^i}), with weights θ^{Q^i} and θ^{μ^i}, respectively.
Copy critic and actor network as target critic and target
  actor network Q'^i(·) and μ'^i(·), respectively.
Set episode ← 0, t ← 0
for episode < E do
  for t < T do
    AUV i takes action a_t^i and gets reward r_t^i with
      consideration of hitting obstacles, traveling outside
      and losing connectivity
    update new state s_{t+1}
    Store (s_t, a_t, r_t, s_{t+1}) in experience replay buffer
    Set s_t ← s_{t+1}
    Get M random samples (s_t, a_t, r_t, s_{t+1}) in
      experience replay buffer and set target value y_j^i
    Update critic network and actor network
    Update target critic network and target actor
      network
  end for
end for
return a_t^i
```

C. Protocol Joint Multi-Modal Transmission and Trajectory Planning

The design of protocol joint multi-modal transmission and trajectory planning for USNs data collection is shown in Fig. 3. Transmission and trajectory are fundamental and interact for data collection. In Fig. 3, the transmission selection part decides transmission mean and mode by Section A, provides transmission situation for the trajectory planning part in Section B. Meanwhile, the trajectory planning in Section B decides AUVs velocity and orientation, and returns trajectory for the transmission selection part in Section A. The two parts make decision based on the observation of underwater environments. Note that, for implementation, the transmission selection part employs Network Function Virtualization (NFV) and Software-Defined Radio (SDR) technology to achieve flexible, diverse, and extensible underwater multi-modal communications.

V. Performance Analysis

As the multi-modal transmission selection consumes very few computing resources, we focus on the computational complexity of trajectory planning in DCMD.

For an AUV's training process, the computational complexity depends on the neuron number of input, hidden and output layers [36]. We consider the four neural networks (actor, critic, target actor, and target critic networks) with i neurons in the input layer, h neurons in the hidden layer and o neurons in the output layer. The complexity of the four back-propagation networks is due entirely to the training process of an AUV, which is iterative. A backward pass which involves the error computations is O(iho). As our model employs experience reply buffer with size B, the total complexity of an AUV is $O(Biho)$, and the total complexity of N AUV is $O(BNiho)$.

The sizes of the inputs of actor networks is $i = 11N$, which is composed of observation space. The size of the inputs of critic network is $i = 3 + 11N$, which is composed of observation and action space. The sizes of the outputs of the critic and actor networks are $o = 4$ and $o = 3$, respectively, as the outputs of critic is action and Q function. The computational complexity of $O(Biho)$ is $O(BN)$. Therefore, the computational complexity of the training process for all AUVs is $O(BN^2)$.

VI. Simulation Studies

A. Simulation Settings

We initially deploy 20 sensor nodes in a 10 km $\times$ 10 km area at random (assuming sensor nodes on the seabed), and assume 3 AUVs and random obstacles in this area. Each AUV has two transmission means, include acoustic and optical communications. The true velocity $\hat{v}$ of an AUV consists two parts: the sailing velocity $v_x^{i,t}, v_y^{i,t}$ and the velocity from the effect of ocean current $\tilde{v}_x^{i,t}, \tilde{v}_y^{i,t}$. We can have $\hat{v}^{i,t} = v^{i,t} + \tilde{v}^{i,t}$. The simulations are tested on a Monte Carlo set of 100 channels.

For transmission, AUVs interact with each other by the acoustics with communication range 1 km, which considers that the size of interactive data is small. AUVs collect data by the acoustics with communication range 500 m, which considers that large data transmission with long-distance is impractical. The transmission rate of acoustics depends on the mode of acoustic communication, according to DSP acoustic modem. The acoustic transmission environment is that, the mean frequency of acoustic signal f is 24KHz, spreading factor u is 0.5, absorption coefficient $a(f)$ depends on the value of f, as $10\lg a(f) = 0.11\frac{f^2}{1+f^2} + 44\frac{f^2}{4100+f} + 2.75g10^{-4}f^2$+0.003. The optical communication range is 100 m, and its transmission rate is 20 M. We assume the optical transmission environment is in clear ocean water turbidity, that absorption $a(\lambda)$ is 0.114, scattering coefficient $b(\lambda)$ is 0.037, and extinction coefficient $c(\lambda)$ is 0.151. Both acoustic and optical communication have retransmission mechanism to ensure the reliability of transmission.

We utilize the ocean current data from true underwater environments, which is downloaded from National Marine Information Center of China (http://global-tide.nmdis.org.cn). The data consists of time, velocity and direction of ocean current [37]. We randomly select the data from Changjiang Estuary ($31°21'N122°1$), China randomly. Note that, similar results are obtained for other ocean current data. We employ smoothing spline algorithm to fit the 24 real data points of ocean current

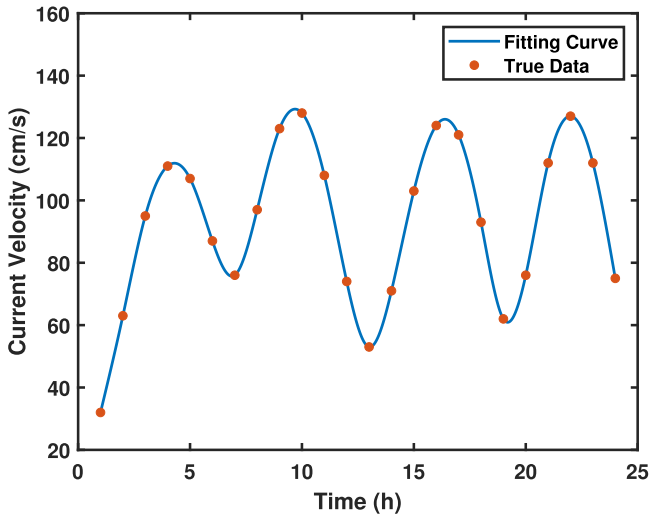


Fig. 4. The fitted ocean current velocity and real data points.

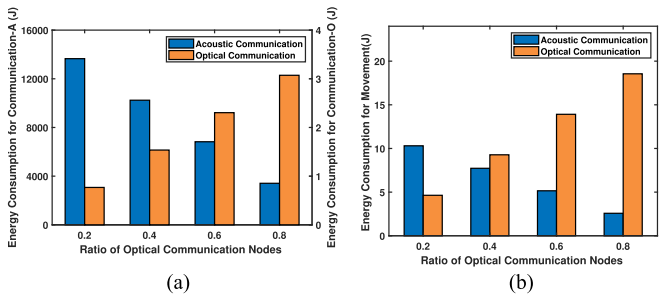


Fig. 5. Transmission and mobility Energy consumptions with multi-modal transmission means. (a) Transmission. (b) Mobility.

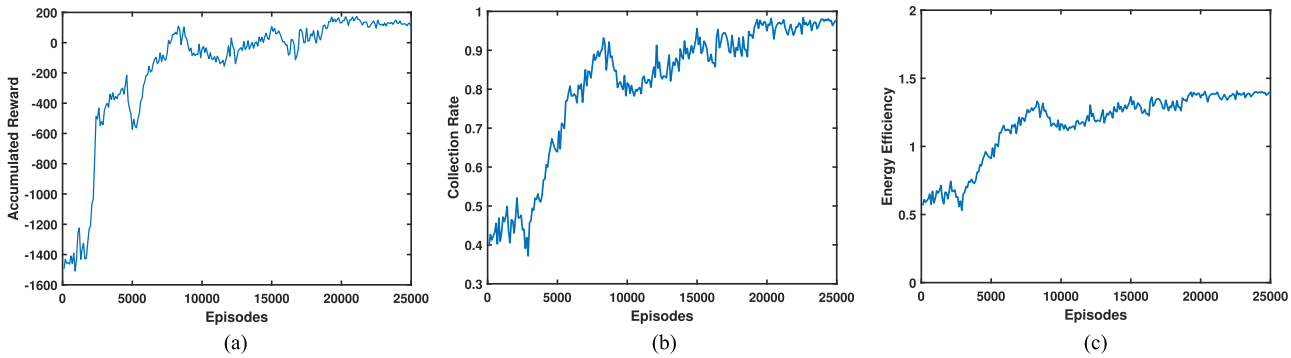
in Changjiang Estuary, on March 17, 2022, randomly, by which acquire 35000 groups of current velocity from 0.5 to 1 h. The date and time above are both selected at random. The fitted current velocity (fitting curve) and real data points (true data) are shown in Fig. 4. In Fig. 4, we can find that the true red data points are all on the fitting curve, which shows a good performance for the fitting of current velocity.

For trajectory planning, we conduct plenty of experiments to find the most appropriate set of hyperparameters. 25000 episodes are prepared for training, and each of episode has 50 timeslots. During testing period, we test each episode 100 times, take the average and the best one out of each episode. We set $p_o = 3$, and $p_l = 10$, which can best train the model. We set $\tau = 0.01$, discount factor is 0.95, neuron number of hidden layer is 64, and experience replay buffer size $1M$.

B. Multi-Modal Transmission Selection

We simulate energy consumption with multi-modal underwater communications for USNs, which considers the specific optical and acoustic communication conditions. The energy consumption with transmission and movement are shown in Fig. 5(a) and (b).

In Fig. 5(a), we can observe that as number of optical communication nodes goes up, the transmission energy consumption

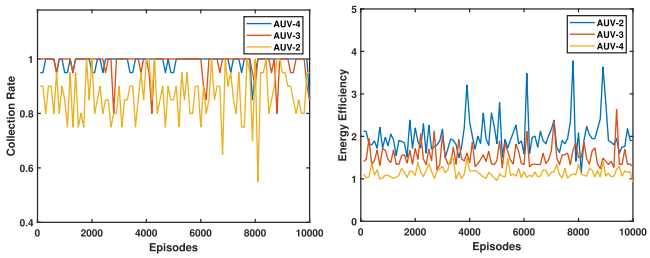


(a) (b) (c)

Fig. 6. Accumulated reward, collection rate and energy efficiency during training. (a) Accumulated reward. (b) Collection rate. (c) Energy efficiency.

TABLE III
BENEFIT AND COST OF NEURAL NETWORKS WITH DIFFERENT HIDDEN LAYER NODES

Neuron number	64	96	128	160
Training time(s)	**2867.5**	3608.5	4222.5	4730.3
Collection rate	**97.3%**	96.6%	96.9%	98.6%
Energy efficiency	**1.53**	1.3	**1.8**	1.3



(a) (b)

Fig. 7. Collection rate and energy efficiency with different number of AUVs. (a) Collection rate. (b) Energy efficiency.

of all nodes drops down, which shows the transmission energy efficiency improves. For example, when 80% of the nodes use optical communication, the transmission energy consumption for all AUVs is about 3413 J, compared to 13653 J when 20% of the nodes use optical communication.

As we know, when AUVs adopt optical communication, they need to sail near to nodes. Then we analyze the benefit of multi-modal communication and movement in Fig. 5(b). Compared with communication, the movement energy consumption of AUVs are small. When 80% nodes use optical communication, the movement energy consumption for all AUVs is about 14.9 J, compared to 21.1 J when 20% nodes use optical communication.

C. Training Convergence for Trajectory Planning

We first simulate extensively to find appropriate hyperparameters of the DNN. The results of different numbers of neurons in actor, critic, and their target networks during training are shown in Table III.

In Table III, we can observe that the neuron number in each layer of a 2 fully-connected hidden layers affect collection rate and energy efficiency. The energy efficiency with different neurons numbers is fluctuating. The neural network with 128 hidden nodes has the best energy efficiency, but its collection rate is lower than neuron number 64, and training time is much larger than 64. Although the energy efficiency of 64 neuron nodes' energy efficiency is not best, its training time is smallest, and its collection rate is very high. Considering the benefit and cost of neural number, we set neuron number of hidden layer 64.

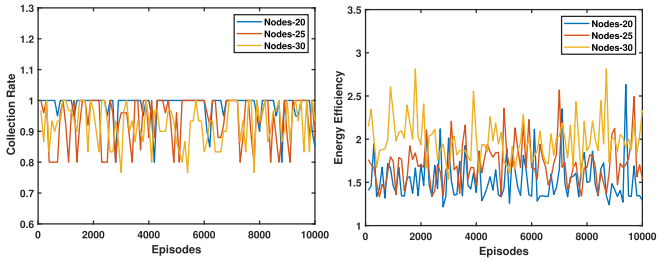
Then we simulate the convergence of the trajectory planning algorithm in Fig. 6. As shown in Fig. 6(a), although the reward is fluctuating in 25000 episodes for training, the trend of it is increasing overall. During the first 10000 episodes, AUVs travel randomly for training, and communication lose, traveling outside, and obstacle collision all have penalties, which make the accumulated reward is negative at first 10000 episodes. As the training time increases, AUVs try to maintain communication, travel inside, avoid obstacles, and consider collection rate and energy consumption affected by ocean currents. Thus, AUVs have positive reward which goes up. At about 20000 episodes, the rewards keep high and stable.

In Fig. 6(b) and (c), we show the results of collection rate and energy efficiency for training. In Fig. 6(b), the collection rate rises linearly over time and eventually reaches a high collection rate for USNs, which maintains fluctuating a little bit over than 90%. The max collection rate is 98.4%. Meanwhile, the trend of energy efficiency in Fig. 6(c) is the same with collection rate, which increases first before 10000 episodes, and reaches a stable and high value after about 20000 episodes.

The results in Table III and Fig. 6 verify that AUVs are well trained with good policy.

D. Finding Appropriate Hyperparameters for Path Planning

We evaluate performance of DCMD with comparison of different number of AUVs. Fig. 7(a) show the comparison results of 2, 3, 4 AUVs for collection rate, respectively. We can observe that the collection rate of 3 and 4 AUVs are both significantly better than 2 AUVs. Although the collection rate of 3 AUVs is close to 4 AUVs, its energy efficiency in Fig. 7(b) is higher than 4 AUVs. Therefore 3 AUVs are appropriate for our simulation coverage area.



(a) (b)

Fig. 8. Collection rate and energy efficiency with different number of sensor nodes. (a) Collection rate. (b) Energy efficiency.

The performance comparison of different number of sensor nodes in USNs is shown in Fig. 8. Fig. 8(a) and (b) show the comparison results of 20, 25, 30 sensor nodes for collection rate and energy efficiency, respectively. We can observe that the collection rate of 20, 25, 30 nodes fluctuate between 0.8 and 1, but overall all perform well with 10000 episodes. The average collection rate with 20, 25, 30 nodes is 95.67%, which verify the stability of DCMD. The energy efficiency goes up with increment of sensor nodes number, because energy consumption is higher for more sensor nodes collection.

E. Comparing With State-of-The-Art Baselines

We compare DCMD with four state-of-the-art methods, include IDDQN [11], APPD [12], CT-DE [13], and Random. IDDQN has double actor and critic neural networks with continuous action space of each AUV. But each AUV only considers its own state, not global environments and states, which has disadvantage on performing distributed collaborative tasks. CT-DE uses multi-agent deep deterministic policy gradient (MADDPG) with one actor neural network and one critic neural network. APPD adopts greedy solution that may lead to maximum immediate reward r_t^i in a distributed manner, without considering the long term. Random is a baseline method that each AUV randomly decides its action, includes velocity and orientation, not considering the tasks for collection.

1) Comparison Results With Different Number of Sensor Nodes: In Fig. 9, we show the results of DCMD and baseline methods with different number of sensor nodes. We test with 10000 episodes. Fig. 9(a), (b), (c) show that DCMD performs significantly better on collection rate than CT-DE, APPD, and random methods with 20, 25, and 30 nodes. Although the collection rate of IDDQN in Fig. 9(a), (b), (c) is near to DCMD, the average collection rate of DCMD is 0.98, 0.95, 0.94, which is better than IDDQN 0.90, 0.93, 0.90, respectively.

Then in Fig. 9(d), (e), (f), we compare energy efficiency of DCMD with state-of-the-art baseline methods with 20, 25, and 30 nodes. In Fig. 9(d), the average energy efficiency of DCMD is 1.53, which is better than IDDQN 1.37, APPD 1.13, and random 1.47. Although the average energy efficiency of CT-DE is the same as DCMD, its collection rate is only 0.64, which is really too low to collect data for USNs. Therefore, DCMD performs best than state-of-the-art methods with higher collection rate and efficiency. The same results are in Fig. 9(e), (f).

TABLE IV
OBSTACLE COLLISION NUMBERS OF DCMD AND BASELINE METHODS WITH EPISODES

Episodes	DCMD	IDDQN	CT-DE	APPD	Random
0 - 2000	**7.6**	13.3	1.45	1.9	1.5
2000 - 4000	**7.9**	14.3	1.8	0.7	1.8
4000 - 6000	**8.5**	13.4	1.45	1.8	1.1
6000 - 8000	**7.45**	14.2	1.6	1.4	2
8000 - 10000	**8**	14.1	1.65	1.25	2.2

In addition, we compare the collision number of DCMD and state-of-the-art baseline methods in Table IV. In Table IV, we can easily observe that, the obstacle collision number of IDDQN is nearly double of DCMD, because IDDQN only consider its own state, and sacrifices obstacle collision numbers to improve collection rate, which is unfitting. The collision numbers of random, APPD, and CT-DE are small, but their collection rates are too low and instability in Fig. 9(a), (b), (c).

2) Comparison Results With Different Number of AUVs: In Fig. 10, we show the results of DCMD and baseline methods with different number of AUVs. We test with 10000 episodes. Fig. 10(a), (b), (c) show that DCMD performs better on collection rate than CT-DE, APPD, and random methods with 2, 3, and 4 AUVs. Although the collection rate of IDDQN in Fig. 10(a), (b), (c) is near to DCMD, the average collection rate of DCMD is 0.95, 0.98, 0.99, which is better than IDDQN 0.82, 0.90, 0.90, respectively.

Then in Fig. 10(d), (e), (f), we compare energy efficiency of DCMD with state-of-the-art baseline methods with 2, 3, and 4 AUVs. In Fig. 10(d), the average energy efficiency of DCMD is 2.16, which is better than IDDQN 1.93, APPD 0.62, and random 1.04. Although the average energy efficiency of CT-DE 2.55 is the highest, its collection rates is only 0.41 in Fig. 10(a), which is really too low to collect data for USNs. The same results are in Fig. 10(e), (f). Therefore, DCMD performs best than state-of-the-art methods with higher collection rate and efficiency.

3) Comparison of Average Results With Different Number of Nodes and AUVs: We analyze the average collection rate and energy efficiency with different sensor nodes and AUV in Table V. As in Table V, we can observe that the average collection rate of DCMD is the best with value 0.97. The average collection rate of IDDQN is close to DCMD with value 0.89. The average collection rates of CT-DE, APPD and random are to low, which cannot be applied for underwater data collection. Average energy efficiency of CT-DE is the highest, with average values of 1.89, because low collection rate means the number of collected nodes is low, which leads to energy consumption is too little to make energy efficiency high.

In Table V, we can also observe that the average collection rate and energy efficiency of DCMD and IDDQN are both relative stable, while CT-DE, APPD and random vary significantly. That because underwater environments have the characteristics of time and space variability, APPD and random make it hard to find an optimal solution all the time, and CT-DE only have two neural networks for training, which makes its

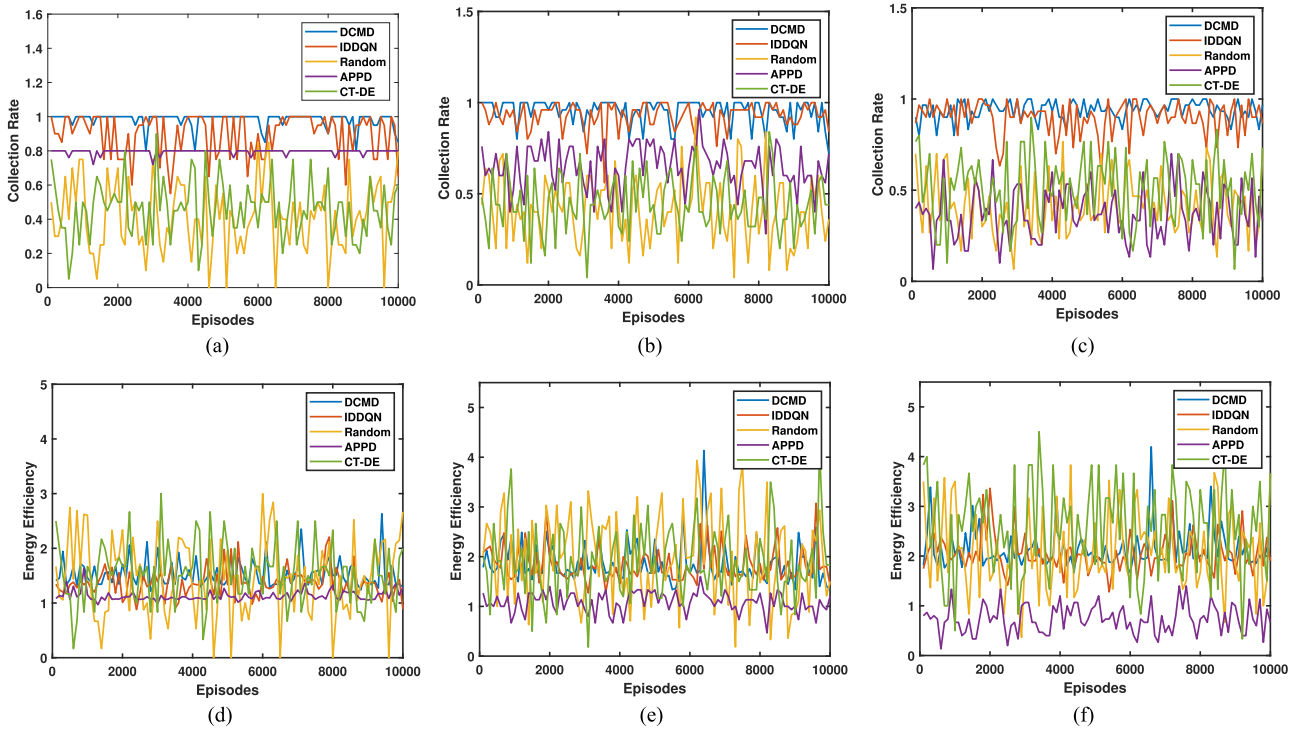


Fig. 9. Collection rate and energy efficiency of DCMD and baseline methods with different number of sensor nodes. (a) Collection rate with 20 nodes. (b) Collection rate with 25 nodes. (c) Collection rate with 30 nodes. (d) Energy efficiency with 20 nodes. (e) Energy efficiency with 25 nodes. (f) Energy efficiency with 30 nodes.

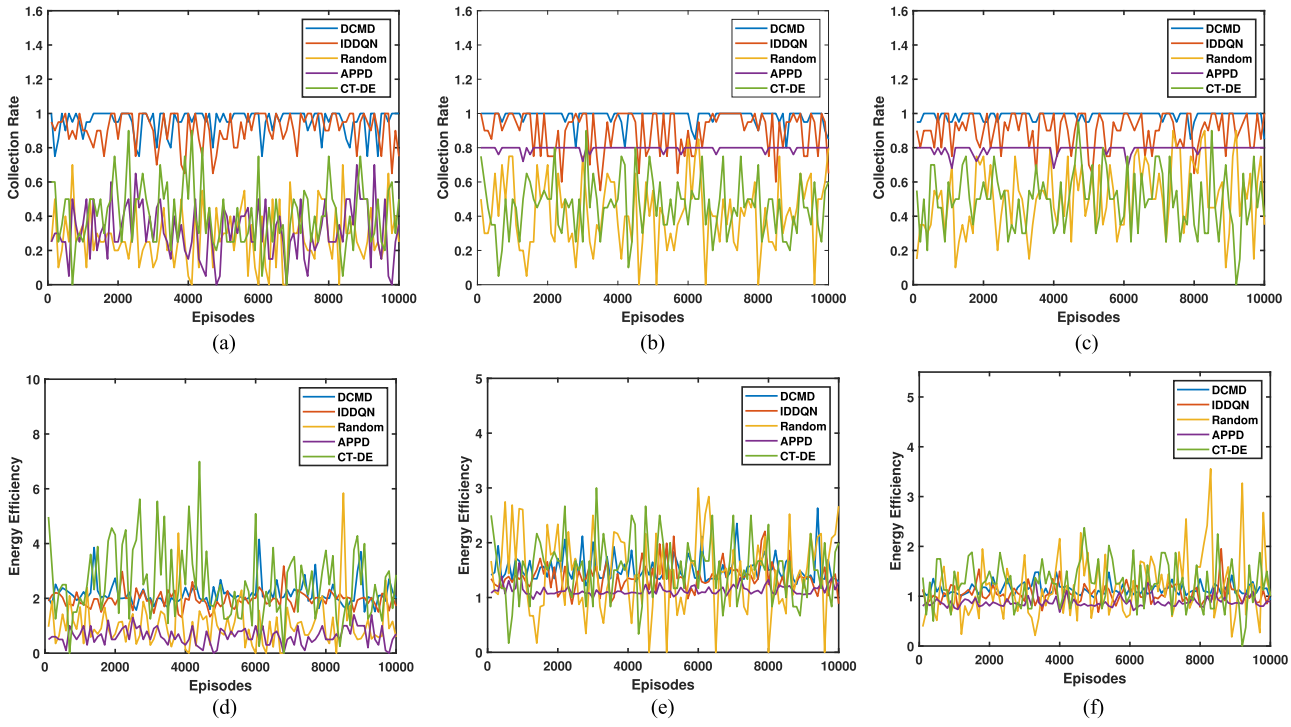


Fig. 10. Collection rate and energy efficiency of DCMD and baseline methods with different number of AUVs. (a) Collection rate with 2 AUVs. (b) Collection rate with 3 AUVss. (c) Collection rate with 4 AUVs. (d) Energy efficiency with 2 AUVs. (e) Energy efficiency with 3 AUVs. (f) Energy efficiency with 4 AUVs.

TABLE V
AVERAGE COLLECTION RATE AND ENERGY EFFICIENCY OF DCMD AND BASELINE METHODS WITH DIFFERENT NUMBER OF AUVS AND SENSOR NODES

Collection rate	DCMD	IDDQN	CT-DE	APPD	Random	**Energy efficiency**	DCMD	IDDQN	CT-DE	APPD	Random
nodes=20	**0.98**	0.90	0.64	0.63	0.46	**nodes=20**	**1.53**	1.37	**1.53**	1.13	1.47
nodes=25	**0.95**	0.93	0.44	0.65	0.42	**nodes=25**	1.86	1.84	1.90	1.08	**2.00**
nodes=30	**0.94**	0.90	0.51	0.37	0.42	**nodes=30**	2.13	2.04	**2.56**	0.74	2.1
AUV=2	**0.95**	0.82	0.41	0.31	0.28	**AUV=2**	2.16	1.93	**2.55**	0.62	1.04
AUV=3	**0.98**	0.90	0.64	0.63	0.46	**AUV=3**	**1.53**	1.37	**1.53**	1.13	1.47
AUV=4	**0.99**	0.90	0.49	0.79	0.51	**AUV=4**	1.14	1.05	**1.24**	0.88	**1.24**
Average	**0.97**	0.89	0.52	0.66	0.43	**Average**	1.73	1.60	**1.89**	0.93	1.55

training accuracy cannot catch up with DCMD and IDDQN. DCMD has higher collection rate and energy efficiency than IDDQN because DCMD considers global observation information during training, while IDDQN only considers a single AUV's own observation information. Therefore, comprehensive consideration, DCMD performs best than other state-of-the-art baselines.

VII. CONCLUSION

In this paper, we present DCMD to collect data of multi-modal underwater sensor networks by multiple AUVs. The main ideas of DCMD is to utilize multi-modal transmission and AUVs' collaborations to achieve high-efficiency data collection, energy consumption, and transmission. We propose an optimal transmission selection algorithm to maximize benefit of communication. The AUVs' trajectory planning algorithm based on deep reinforcement learning is proposed to improve collection efficiency by collaborations, with consideration of transmission situation and the unique underwater environments: obstacles and ocean currents. The transmission selection and trajectory planning are combined in one protocol. Our experimental results show that DCMD achieves better performance by multiple AUVs than other benchmark methods.

In the future, we plan to further improve performance of the AUVs' trajectory planning, which to reduce computing and storage demands of AUVs.

REFERENCES

[1] G. Han, X. Long, C. Zhu, M. Guizani, and W. Zhang, "A high-availability data collection scheme based on multi-AUVs for underwater sensor networks," *IEEE Trans. Mobile Comput.*, vol. 19, no. 5, pp. 1010–1022, May 2020.

[2] S. Song, J. Liu, J. Guo, J. Wang, and J. H. Cui, "Neural-network based AUV navigation for fast-changing environments," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9773–9783, Oct. 2020.

[3] H. Zheng, W. Ning, and W. Jie, "Minimizing deep sea data collection delay with autonomous underwater vehicles," *J. Parallel Distrib. Comput.*, vol. 104, pp. 99–113, 2017.

[4] X. Wei, H. Gu, X. Wang, X. Wang, and M. Qiu, "Reliable data collection techniques in underwater wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 404–431, Jan.–Mar. 2022.

[5] G. A. Hollinger et al., "Underwater data collection using robotic sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 5, pp. 899–911, Jun. 2012.

[6] C. Lin, G. Han, J. Du, Y. Bi, and K. Fan, "A path planning scheme for AUV flock-based internet-of-underwater-things systems to enable transparent and smart ocean," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9760–9772, Oct. 2020.

[7] T. Wang, D. Zhao, S. Cai, W. Jia, and A. Liu, "Bidirectional prediction-based underwater data collection protocol for end-edge-cloud orchestrated system," *IEEE Trans. Ind. Inform.*, vol. 16, no. 7, pp. 4791–4799, Jul. 2020.

[8] 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 Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6037–6049, Jun. 2019.

[9] C. A. Nan et al., "6G service-oriented space-air-ground integrated network: A survey - sciencedirect," *Chin. J. Aeronaut.*, vol. 35, no. 9, pp. 1–18, 2022.

[10] C. Zhou et al., "Deep reinforcement learning for delay-oriented IoT task scheduling in SAGIN," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 911–925, Feb. 2021.

[11] Z. Chu, F. Wang, T. Lei, and C. Luo, "Path planning based on deep reinforcement learning for autonomous underwater vehicles under ocean current disturbance," *IEEE Trans. Intell. Veh.*, early access, Feb. 28, 2022, doi: 10.1109/TIV.2022.3153352.

[12] Z. Li and X. Luo, "Autonomous underwater vehicles (AUVs) path planning based on deep reinforcement learning," in *Proc. 11th Int. Conf. Intell. Control Inf. Process.*, 2021, pp. 125–129.

[13] Z. Yang, J. Du, Z. Xia, C. Jiang, A. Benslimane, and Y. Ren, "Secure and cooperative target tracking via auv swarm: A reinforcement learning approach," in *Proc. IEEE Glob. Commun. Conf.*, 2021, pp. 1–6.

[14] C. Liu, Z. Chen, and Y. Zhan, "Energy-efficient distributed mobile crowd sensing: A deep learning approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1262–1276, Jun. 2019.

[15] X. Zhuo, M. Liu, Y. Wei, G. Yu, and R. Sun, "AUV-aided energy-efficient data collection in underwater acoustic sensor networks," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10010–10022, Oct. 2020.

[16] N. Cheng et al., "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, May 2019.

[17] J. Liu, J. Wang, S. Song, J. Cui, and B. Li, "MMNET: A multi-modal network architecture for underwater networking," *Electronics*, vol. 9, no. 12, 2020, Art. no. 2186.

[18] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke, "Data collection, storage, and retrieval with an underwater sensor network," in *Proc. 3rd Int. Conf. Embedded Networked Sensor Syst.*, New York, NY, USA, 2005, pp. 154–165.

[19] I. F. Akyildiz, P. Wang, and S. C. Lin, "Softwater: Software-defined networking for next-generation underwater communication systems," *Ad Hoc Netw.*, vol. 46, pp. 1–11, 2016.

[20] C. Lin, G. Han, M. Guizani, Y. Bi, J. Du, and L. Shu, "An SDN architecture for AUV-based underwater wireless networks to enable cooperative underwater search," *IEEE Wireless Commun.*, vol. 27, no. 3, pp. 132–139, Jun. 2020.

[21] M. Khan, S. H. Ahmed, Y. Z. Jembre, and D. Kim, "An energy-efficient data collection protocol with AUV path planning in the Internet of Underwater Things," *J. Netw. Comput. Appl.*, vol. 135, pp. 20–31, 2019.

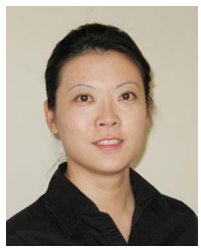
[22] M. Samir, S. Sharafeddine, M. Assi, M. Nguyen, and A. Ghrayeb, "UAV trajectory planning for data collection from time-constrained IoT devices," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 34–46, Jan. 2020.

[23] Y. N. Ma, Y. J. Gong, C. F. Xiao, Y. Gao, and J. Zhang, "Path planning for autonomous underwater vehicles: An ant colony algorithm incorporating alarm pheromone," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 141–154, Jan. 2019.
[24] C. Xiong, D. Chen, D. Lu, Z. Zeng, and L. Lian, "Path planning of multiple autonomous marine vehicles for adaptive sampling using voronoi-based ant colony optimization," *Robot. Auton. Syst.*, vol. 115, pp. 90–103, 2019.
[25] M. Elhoseny, A. Shehab, and X. Yuan, "Optimizing robot path in dynamic environments using genetic algorithm and bezier curve," *J. Intell. Fuzzy Syst.*, vol. 33, no. 4, pp. 2305–2316, 2017.
[26] C. Cheng, D. Zhu, S. Bing, Z. Chu, J. Nie, and Z. Sheng, "Path planning for autonomous underwater vehicle based on artificial potential field and velocity synthesis," in *Proc. IEEE 28th Can. Conf. Elect. Comput. Eng.*, 2015, pp. 717–721.
[27] R. J. Wai and A. S. Prasetia, "Adaptive neural network control and optimal path planning of UAV surveillance system with energy consumption prediction," *IEEE Access*, vol. 7, pp. 126137–126153, 2019.
[28] Y. Hai, W. Lei, S. Zhou, Z. Shi, and Z. Hao, "DSP based receiver implementation for OFDM acoustic modems," *Phys. Commun.*, vol. 5, pp. 22–32, 2012.
[29] P. Xie, J. H. Cui, and L. Lao, "VBF: Vector-based forwarding protocol for underwater sensor networks," in *Proc. 5th Int. IFIP-TC6 Conf. Netw. Technol., Serv., Protocols; Perform. Comput. Commun. Netw.; Mobile Wireless Commun. Syst.*, 2006, pp. 1216–1221.
[30] Y. Hai, Z. J. Shi, and J. H. Cui, "DBR: Depth-based routing for underwater sensor networks," in *Proc. Int. IFIP-TC6 Netw. Conf. Adhoc Sensor Netw.*, 2008, pp. 72–86.
[31] P. Xie, Z. Zhou, Z. Peng, J.-H. Cui, and Z. Shi, "Void avoidance in three-dimensional mobile underwater sensor networks," in *Wireless Algorithms, Systems, and Applications*, Ser. Lecture Notes in Computer Science, Ed., vol. 5682, 2009, pp. 305–314.
[32] R. Z. Zhou and J. H. Cui, "Energy efficient multi-path communication for time-critical applications in underwater sensor networks," in *Proc. 9th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Hong Kong, China,, 2008, pp. 221–230.
[33] V. I. Haltrin, "Chlorophyll-based model of seawater optical properties," *Appl. Opt.*, vol. 38, no. 33, 1999, Art. no. 6826.
[34] L. Prieur and S. Sathyendranath, "An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials1," *Limnol. Oceanogr.*, vol. 26, no. 4, pp. 671–689, 1981.
[35] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004.
[36] M. Sipper, "A serial complexity measure of neural networks," IEEE Int. Conf. Neural Netw., 1993, pp. 962–966.
[37] S. Song, L. Jiu, J. Guo, C. Zhang, T. Yang, and J. Cui, "Efficient velocity estimation and location prediction in underwater acoustic sensor networks," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2984–2998, Feb. 2022.

Shanshan Song received the B.S. and M.S. degrees in computer science and technology from Jilin University, Changchun, China, in 2011 and 2014, respectively, and the Ph.D, degree in management science and engineering from Jilin University in 2018. She was a Postdoctoral Researcher with the Department of Computer science and technology, Jilin University. She is currently an Associate Professor with the Department of Computer science and technology, Jilin University. Her main research interests include underwater data collection, localization, and navigation. She is the WUWNet' 2021 EDAS Chair.

Jun Liu (Member, IEEE) received the B.Eng. degree in computer science from Wuhan University, Wuhan, China, in 2002, and the Ph.D. degree in computer science and engineering from the University of Connecticut, Storrs, CT, USA. He is currently a Professor with the School of Electronic and Information Engineering with Beihang University, Beijing, China. His main research interests include underwater acoustic networking, time synchronization, localization, network deployment, and also interested in operating system, and cross layer design.

Jiani Guo received the B.S. degree in computer science and technology from Beijing Jiaotong University, Beijing, China. She is currently working toward the Ph.D. degree with the College of Computer science and technology with Jilin University, Changchun, China. Her research interests include MAC protocols design and localization for underwater acoustic networks.



Bin Lin (Senior Member, IEEE) received the B.S. and M.S. degrees from Dalian Maritime University, Dalian, China, in 1999 and 2003, respectively, and the Ph.D. degree from Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2009. She is currently a Full Professor and the Dean of Communication Engineering with the College of Information Science and Technology, Dalian Maritime University. From 2015 to 2016, she has been a Visiting Scholar with George Washington University, Washington, DC, USA. Her research interests include wireless communications, network dimensioning and optimization, resource allocation, artificial intelligence, maritime communication networks, edge/cloud computing, wireless sensor networks, and Internet of Things. She is an Associate Editor for the *IET Communications*.

Qiang Ye (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2016. Since September 2021, he has been an Assistant Professor with the Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada. His research interests include network slicing for 5G networks, edge intelligence for autonomous vehicular networks, artificial intelligence for future networking and protocol design, and performance analysis for the Internet of Things. He was a Technical Program Committee Members for several international conferences, including the IEEE GLOBECOM'20, VTC'17, VTC'20, and ICPADS'20. He is the Editor of the *International Journal of Distributed Sensor Networks* (SAGE Publishing) and *Wireless Networks* (SpringerNature), and the Area Editor of the Encyclopedia of Wireless Networks (SpringerNature).

Junhong Cui received the B.S. degree in computer science from Jilin University, Changchun, China, in 1995, the M.S. degree in computer engineering from the Chinese Academy of Sciences, Beijing, China, in 1998, and the Ph.D. degree in computer science from the University of California, Los Angeles, Los Angeles, CA, USA. She was on the Faculty of the Computer Science and Engineering Department with the University of Connecticut, Storrs, CT, USA. She is currently the Professor with the College of Computer Science and Technology with Jilin University, Changchun, China. Her research mainly focuses on algorithm and protocol design in underwater sensor networks. She is the Guest Editor of *ACM Mobile Computing and Communications Review* and *Elsevier Ad Hoc Networks*. She co-founded the first ACM International Workshop on UnderWater Networks (WUWNet 2006) and is currently the WUWNet Steering Committee Chair.