Efficient Collision-Free Data Collection for Underwater Acoustic Sensor Networks: A Hierarchical DRL Approach

Hao Chen, Jiani Guo, Bowen Zhang, Shanshan Song*, *Member, IEEE*, Qiang(John) Ye, *Senior Member, IEEE*, and Miao Pan, *Senior Member, IEEE*

***Abstract*—Autonomous underwater vehicles (AUVs) have become a promising solution for data collection in underwater acoustic sensor networks (UASNs), and deep reinforcement learning (DRL) has been widely applied to enhance collection performance. However, our preliminary experiments indicate that existing DRL-based studies still face two critical challenges: 1) Collection blind spots. The sparse collection rewards and the requirement for energy-efficient trajectory planning during data collection jointly restrict AUVs' ability to explore and collect data from all sensor nodes (SNs), ultimately resulting in some SNs remaining uncollected. 2) Collection collisions. Simultaneous data collection by multiple AUVs can lead to packet collisions and collection failures, further decreasing the collection rate. To address these challenges, we propose a hierarchical DRL-based collision-free data collection scheme (HCDC). Specifically, we leverage a hierarchical DRL framework to decompose the multi-AUV-assisted data collection (MADC) problem into a high-level global target selection (GTS) and a low-level local trajectory planning (LTP) subproblems. For GTS, we design a multi-agent GTS (MA-GTS) algorithm to assign the next target SN for collection to each AUV. The MA-GTS incorporates both global and local rewards to collaboratively optimize the overall energy consumption while avoiding individual penalties. Based on the assigned target SN, a deep deterministic policy gradient-based LTP (DDPG-LTP) algorithm is proposed to conduct AUV trajectory planning, utilizing intrinsic rewards to enhance learning efficiency and eliminate collection blind spots. Furthermore, to avoid packet collisions, we analyze the conditions for collision-free data collection and propose an adaptive back-off slot (ABS) algorithm to schedule AUVs' collection slots. With the collision-free slots, DDPG-LTP dynamically adjusts AUVs' velocities to ensure collision-free collection while reducing energy consumption. Extensive simulation results demonstrate that HCDC can achieve better collection rate and energy efficiency than state-of-the-art schemes.**

***Index Terms*—Underwater acoustic sensor networks, data collection, trajectory planning, collision-free.**

Manuscript received April 19, 2021; revised August 16, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFC2803000; in part by the National Natural Science Foundation of China under Grant 62471201 and Grant 62501250; in part by the Postdoctoral Science Foundation of China under Grant 2025M771509; in part by the Postdoctoral Fellowship Program of CPSF under Grant Number GZC20250178. *(Shanshan Song is the corresponding author.)*

Hao Chen, Jiani Guo, Bowen Zhang, and Shanshan Song are with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: haochen22@mails.jlu.edu.cn; jnguo@jlu.edu.cn; zhangbw23@mails.jlu.edu.cn; songss@jlu.edu.cn).

Qiang(John) Ye is with the Department of Electrical and Software Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: qiang.ye@ucalgary.ca).

Miao Pan is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204 USA (e-mail: mpan2@uh.edu).

I. Introduction

As the demand for marine exploration and resource utilization continues to grow, underwater acoustic sensor networks (UASNs) have become increasingly vital in various underwater applications, such as environmental monitoring [1], marine search [2], and resource exploration [3]. In these applications, numerous sensor nodes (SNs) are deployed to collect and transmit sensing data to sink nodes via multi-hop communication [4] [5], which results in high and unbalanced energy consumption across the network. To achieve energy-efficient data collection and balanced energy consumption, autonomous underwater vehicles (AUVs) have been widely employed to assist data collection in UASNs [6] [7]. However, a single AUV is constrained by its finite battery power and slow cruise velocity, leading to long-delay data collection. To improve collection efficiency, multi-AUV-assisted data collection has become a promising solution.

For multi-AUV-assisted data collection, the primary objectives are to improve the collection rate and decrease the system cost (e.g. AUV sailing distance and energy consumption). To achieve this, AUV trajectory planning has been extensively studied [8] [9], as the trajectory determines whether an AUV can reach a collection area and significantly affects the efficiency of the collection system. Once an AUV reaches the area, the collection is conducted through underwater acoustic communication between the AUV and the deployed SNs. Therefore, data communication is critical for further enhancing the collection efficiency [10]. The AUV trajectory planning and communication need to be jointly optimized to achieve the aforementioned objectives, while dealing with challenges such as ocean currents and underwater obstacles. Deep Reinforcement Learning (DRL) has been widely applied to trajectory planning and communication optimization in multi-AUV-assisted data collection [11] [12], as it can self-learn and adapt to dynamic environments. However, existing DRL-based studies still face the following two challenges.

Collection blind spots: In a DRL-based collection process, AUVs only receive collection rewards after moving multiple steps to reach a designated collection area, resulting in sparse rewards. Especially during the initial random exploration phase, AUVs struggle to accumulate positive reward reward samples. Consequently, insufficient positive samples hinder learning to collect data from all SNs, leading to collection blind spots. Moreover, the collection process also needs to

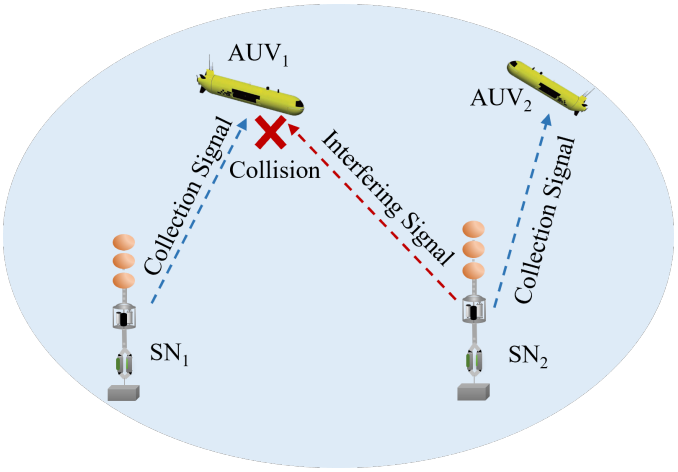


Fig. 1: Packet collisions resulting from simultaneous collection by multiple AUVs.

consider sailing energy consumption and avoiding obstacles, which introduce additional rewards. These rewards are inherently in conflict with the collection objective and are generated far more frequently than the sparse collection rewards, further hindering the exploration of uncollected areas. Although some studies have mitigated the collection blind problem through reward shaping [13] [14], manually designing dense rewards often has inherent limitations: 1) The introduced rewards need to be specifically designed for specific tasks to guide the agent in completing the tasks, lacking generalizability; 2) The introduced dense rewards may conflict with the original objectives, causing AUVs to repeatedly obtain rewards in local regions while refusing to explore the other regions.

Packet collisions: In the communication process after AUVs reaching the collection area, most existing methods focus on the communication resource allocation between AUVs and SNs to ensure the energy efficiency and reliable collection [15] [16]. However, potential packet collisions may occur during simultaneous data collection by multiple AUVs. As Fig. 1 shows, AUV_1 and AUV_2 simultaneously collect data from SN_1 and SN_2, respectively. The data sent from SN_2 interferes with the data sent from SN_1, causing a packet collision at AUV_1. Such unexpected collisions can lead to collection failures and reduced collection rate. Packet collisions are particularly severe in underwater environments due to the narrow bandwidth and long preambles [17].

In this paper, we propose a hierarchical DRL-based multi-AUV-assisted collision-free data collection scheme (HCDC). To address the problem of collection blind spots, we formulate and decompose the overall multi-AUV-assisted data collection (MADC) problem into a high-level global target selection (GTS) subproblem and a low-level local trajectory planning (LTP) subproblem. Accordingly, we develop a hierarchical algorithm framework, consisting of a multi-agent GTS (MA-GTS) algorithm and a deep deterministic policy gradient-based LTP (DDPG-LTP) algorithm, to solve the GTS and the LTP subproblems, respectively. The MA-GTS is designed to assign collection targets and determine the sequence for optimizing the collection rate and AUV energy consumption, which employs global and local rewards to collaboratively optimize the overall energy consumption while avoiding individual penalty. Based on the collection targets obtained from the MA-GTS, the DDPG-LTP is employed to optimize the trajectory to efficiently reach each target's collection location. Leveraging such hierarchical framework, the low-level DDPG-LTP can obtain intrinsic rewards based on the collection targets, avoiding collection blind spots. Furthermore, to avoid collection collision, we analyze the condition for collision-free data collection, and introduce an adaptive back-off slot (ABS) algorithm to schedule AUVs' collision-free collection slots. With the collision-free slots, the DDPG-LTP not only adjusts AUV velocities to ensure collision-free collection but also reduces energy consumption. The main contributions of this paper are summarized as follows:

- For multi-AUV-assisted data collection, we decompose the MADC problem into the high-level GTS and the low-level LTP subproblem. Correspondingly, we develop a hierarchical DRL framework, consisting of the MA-GTS and DDPG-LTP algorithms, respectively. Such hierarchical framework obtains intrinsic rewards based on the collection target, avoiding collection blind spots.
- To avoid collection collision, we analyze the condition required for collision-free data collection and design an ABS algorithm to schedule AUVs' collection slots.
- We design global reward and local reward for the MA-GTS to collaboratively optimize overall energy consumption while avoiding individual penalties, and introduce collision-free collection slots for the DDPG-LTP to avoid packet collisions.
- We conduct simulations to evaluate the performance of the HCDC and several state-of-the-art methods. Our results demonstrate that HCDC achieves better energy efficiency and collection rates than the state-of-the-art schemes.

The rest of this paper is organized as follows: we discuss the related work in Section II. In Section III, we describe the considered multi-AUV-assisted data collection network, present the MADC problem formulation, and explain its decomposition into the high-level GTS subproblem and the low-level LTP subproblem. Section IV presents the details of HCDC, including MA-GTS, ABS, and DDPG-LTP algorithms. In Section V, we provide the simulation results, followed by conclusions provided in Section VI.

II. Related Work

Early data collection schemes primarily concentrate on collaborative transmission among SNs. Han *et al.* [18] classified SNs into different virtual data sets and utilized hierarchical strategies of dynamic layer and static layer to optimize data transmission. Guo *et al.* [19] proposed a cross-layer MAC protocol to improve the data collection efficiency by integrating Geo-routing protocols and OFDM technology simultaneously. Due to the flexible and autonomous characteristics of AUVs, AUV-assisited data collection has been extensively studied in recent years. For AUVs with spiral trajectories, a malfunction discovery and repair mechanism is introduced to ensure the

high availability of the data collection scheme [20]. Khan *et al.* [21] investigated an energy-efficient AUV-assisted clustering scheme based on fixed AUV trajectories. The aforementioned schemes do not fully leverage the flexible mobility of AUVs. Instead, they overly rely on multi-hop transmission among SNs, resulting in significant and uneven energy consumption, which greatly reduces the network's lifespan.

The trajectory design of AUVs has a significant impact on the performance of data collection system. Hence, a series of treatises were dedicated to optimize AUV trajectory, aiming for reducing AUV cruising distance or energy consumption. Considering the communication constraints between the AUV and devices, Zhuo *et al.*9072166 formulated AUV trajectory planing as a special traveling salesperson problem (TSP) to minimize the AUV travel time. Similarly, Wei *et al.* [22] transformed the problem of traversing all anchor nodes into a TSP in an AUV-assisted magnetic induction and acoustic hybrid network. To determine an optimal path that maximizes the data value delivered to the aggregation devices, Gjanci *et al.* [23] define a heuristic greedy and adaptive AUV path-finding algorithm that drives the AUV to collect data from nodes depending on the value of their data. Liu *et al.* [24] used an improved genetic algorithm to minimize the sailing time of the AUV as much as possible. However, these studies primarily focus on target selection while assuming an ideal AUV cruising environment, overlooking the impact of complex underwater conditions (including ocean currents and underwater obstacles) on AUV trajectory planning.

To address the AUV trajectory planning challenges in complex underwater environment, existing studies have explored a variety of algorithms. Yang *et al.* [25] proposed an underwater path planning method based on proximal policy optimization to plan the AUV's movement direction, with an information encoding module developed to extract local obstacle features. Considering the maneuverability of under-actuated AUVs under ocean current disturbances, Chu *et al.* [26] developed a dynamic and composite reward function to guide the AUV to its destination while avoiding obstacles. For the multi-modal underwater data collection network, Song *et al.* [27] proposed a novel distributed DRL-based multiple AUV trajectory planning algorithm. For the collection task of unknown SNs location, Jiang *et al.* [28] proposed a MAPPO-based algorithm that uses a target uncertainty map to guide the AUV swarm toward areas with higher probabilities of SN existence. Fang *et al.* [29] comprehensively considered AUV trajectory optimization and energy consumption, and introduced the age of data to ensure data freshness. However, these studies do not address the collection blind spots problem. Some studies design auxiliary continuous rewards to guide AUVs toward their target SNs [13] [14], which can mitigate blind spots problem but cannot completely avoid it.

During the collection phase, reliable communication is fundamental to achieve a high collection rate. Fang *et al.* [30] designed a two-stage algorithm for joint optimization of the communication resource allocation and trajectory planning in AUV-assisted date collection network. To meet the diverse requirements of advanced underwater applications, Hou *et al.* [31] provided on-demand computing services and proposed a system-level optimization model that integrates environment-aware trajectory planning, communication resource allocation, computation offloading, and data caching. However, these studies primarily focus on resource allocation between AUVs and SNs, while overlooking the severe packet collisions in underwater environments. Zhuo *et al.* [32] designed a packet scheduling scheme to transmit data packets from AUVs to the base station while avoiding collisions with regular data transmissions among SNs. To reduce packet collision in the network and decrease the delay of data collection, Wang *et al.* [33] proposed a MAC protocol for network clustering and AUV interaction communication. However, these protocols require frequent control packet exchange between AUVs and SNs, making them unsuitable for multi-AUV-assisted data collection.

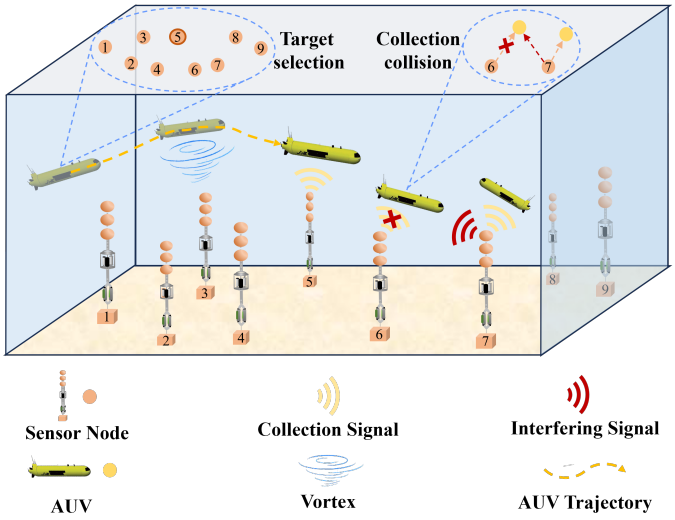


Fig. 2: Multi-AUV-assisted data collection network.

III. System Model and Problem Formulation

A. Network Scenario

We consider an multi-AUV-assisted data collection system for UASN, as shown in Fig. 2. which consists of an AUV swarm denoted by $\mathcal{M} = 1, 2, ..., M$, and several static SNs denoted by $\mathcal{N} = 1, 2, ..., N$. Each AUV is equipped with an acoustic communication device to collect sensor data and facilitate collaborative information sharing. In our scenario, multiple AUVs can share their local state to collaborate to improve collection efficiency through periodic broadcasts, where the communication overhead is acceptable with the advancement of broadcast protocols [34] [35]. In each slot, AUVs collaboratively select their collection targets and determines velocity for their movement. After movement, a set of AUVs reach the designated collection location and send a ACTIVE packet to initiate the data collection period, while other AUVs remain silent. After successfully collecting sensor data, AUV will store data until it returns to the base station to deliver data. The main notations used throughout this paper are summarized in Table I.

TABLE I: Notations and Symbols

Notations	Description
$M/\mathcal{M}$	number/set of AUVs
$N/\mathcal{N}$	number/set of SNs
$\tau/T/\mathcal{T}$	Length/number/set of slots
L	Packet size
R	Transmission rate
T_p	Packet transmission delay
ζ	Conversion efficiency of electricity
C_d	Drag coefficient
$h_{i,t}$	AUV i's collection target in slot t
$\boldsymbol{v}_{i,t}$	AUV i's velocity in slot t
$\boldsymbol{L}_{i,t}$	AUV i's location in slot t
$e_{i,t}$	AUV i's movement energy consumption in slot t

B. AUV Mobility Model

AUV trajectory planning is significantly affected by complex ocean currents. To tackle the unknown current, AUVs are equipped with the horizontal acoustic doppler current profiler (H-ADCP). For one thing, H-ADCP can measure the current velocities in a horizontal line up to hundreds of meters in front of AUVs. For another, the ocean current is relatively stable in short time and small area. Therefore, the AUV can leverage H-ADCP to perceive current information, avoiding location offsets and reducing energy consumption through trajectory optimization.

Due to the earth's rotation effect, the impact of the horizontal current field remains dominant in AUVs' motion. Thus, the ocean current field is modeled by the 2-D Navier-Stokes equation given by:

$$\frac{\partial \omega}{\partial t} + (\boldsymbol{v}_c \nabla)\omega = \nu \Delta \omega, \tag{1}$$

where $\boldsymbol{v}_c = [\boldsymbol{v}_{c,x}, \boldsymbol{v}_{c,y}]$ represents the velocity field and ω denotes the vorticity of current. Besides, ν is the viscosity of the fluid, while ∇ and Δ are the gradient and Laplacian operators, respectively. For simulation, we approximate the Eq. (1) as:

$$\boldsymbol{v}_c(\boldsymbol{L}_{i,t}) = \frac{A_0 \left(1 - e^{-\frac{\|\boldsymbol{L}_{i,t} - \boldsymbol{L}_0\|_2^2}{{r_0}^2}}\right)}{2\pi \|\boldsymbol{L}_{i,t} - \boldsymbol{L}_0\|_2^2} \cdot [x - x_0, y - y_0], \tag{2}$$

where $\boldsymbol{v}_c(\boldsymbol{L}_{i,t}) = [x, y]$ and $L_0 = [x_0, y_0]$ denote the locations of AUV i and the center of the Lamb vortex, respectively. Furthermore, A_0 and r_0 represent the strength and radius of the vortex. Given AUV i's velocity $\boldsymbol{v}_{i,t}$ at slot t, its location can be updated as follows:

$$\boldsymbol{L}_{i,t+1} = \boldsymbol{L}_{i,t} + \boldsymbol{v}_{i,t}\tau. \tag{3}$$

where τ is the slot length. Based on the AUV i's location, we can obtain the current velocity. Then the relative velocity can be denoted as follows:

$$\boldsymbol{v}_r(\boldsymbol{L}_{i,t}) = \boldsymbol{v}_{i,t} - \boldsymbol{v}_c(\boldsymbol{L}_{i,t}). \tag{4}$$

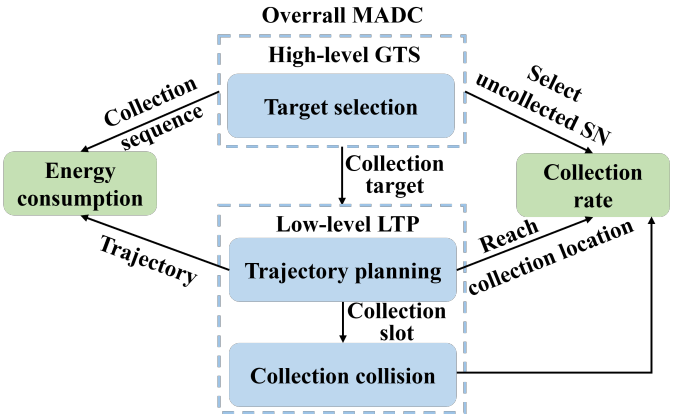


Fig. 3: The overall MADC is decomposed into a high-level GTS and a low-level LTP. The high-level GTS select the target SNs to be collected, while the low-level LTP plans trajectory based on the collection target to reach the designated collection location and achieve collision-free data collection.

where $\boldsymbol{v}_c(\boldsymbol{L}_{i,t})$ is the current velocity, and $\boldsymbol{v}_{i,t}$ is AUV's absolute velocity. Based on the computational fluid dynamics (CFD) methods, the drag force of AUV i can expressed as:

$$F_{i,t} = \frac{1}{2}\rho_L A C_d \|\boldsymbol{v}_r(\boldsymbol{L}_{i,t})\|_2^2, \tag{5}$$

where C_d denotes the drag coefficient, A is the cross-sectional area of AUV moving along the current direction, and ρ_L is the density of seawater. Moreover, we obtain the movement energy consumption of AUV as:

$$e_{i,t} = \frac{1}{\zeta} F_{i,t} \|\boldsymbol{v}_r(\boldsymbol{L}_{i,t})\|_2 \tau, \tag{6}$$

where ζ is the conversion efficiency of electricity.

C. Problem Formulation and Decomposition

1) Overall MADC Problem: The optimization objectives of the overall MADC problem include maximize collection rate and minimize AUVs' energy consumption. The total collection rate collected by AUV swarm can be expressed as:

$$R_c = \frac{\sum_{i=1}^{M}\sum_{t=1}^{T} f(h_{i,t}) g_{i,t}}{N}. \tag{7}$$

where $h_{i,t} \in \mathcal{N}$ is AUV i's collection target in slot t, $f(h_{i,t})$ is a binary function that $f(h_{i,t}) = 1$ if $h_{i,t}$ has not been collected before, otherwise $f(h_{i,t}) = 0$, and $g_{i,t}$ is a binary value that $g_{i,t} = 1$ if AUV i reach $h_{i,t}$'s collection location, otherwise $g_{i,t} = 0$. During data collection period, the AUVs' total energy consumption can be expressed as:

$$E = \sum_{i=1}^{M}\sum_{t=1}^{T} e_{i,t}. \tag{8}$$

For maximizing R_c and minimize E, we jointly optimize the AUVs' velocity strategy $\boldsymbol{V} = \{\boldsymbol{v}_{i,t}, i \in \mathcal{M}, t \in \mathcal{T}\}$ and target

selection strategy $\boldsymbol{H} = \{h_{i,t}, i \in \mathcal{M}, t \in \mathcal{T}\}$. As a result, the MADC problem can be define as:

$$\min_{\boldsymbol{V},\boldsymbol{H}} F = \{-R_c, E\}. \tag{9}$$

$$\text{s.t. } 0 \leq \|\boldsymbol{v}_{i,t}\|_2 \leq v_{max}, \forall i \in \mathcal{M}, \forall t \in \mathcal{T} \tag{9a}$$

$$0 \leq \sum_{t=1}^{T} e_{i,t} \leq E_{max}, \forall i \in \mathcal{M} \tag{9b}$$

$$\tag{9c}$$

where v_{max} and E_{max} are the maximum velocity and battery capacity of AUV, respectively. Constraint (9a) limits the max velocity of AUV according to the actual situation. Constraint (9b) bounds that each AUV's total energy consumption should not exceed the battery capacity, preventing the AUV from failing to return safely to the base station.

In the over MADC problem, AUVs typically takes multiple steps to reach the designated collection location and collect data. In other words, collection targets are long-term goals that remain consistent over extended periods, whereas AUV trajectories require frequent adjustments to adopt complex ocean currents and avoid obstacles effectively. Therefore, as illustrated in Fig. 3, we decompose the overall MADC problem into a high-level GTS subproblem and a low-level LTP subproblem, which are responsible for target selection and AUV trajectory planning, respectively.

2) High-level GTS Problem: After an AUV selects a target SN, it may require multiple time slots to reach the designated collection location. Therefore, the high-level GTS problem is executed in a large timescale, determining the global collection sequence by selecting target SNs. The collection sequence directly influences the traversal distance of AUVs and thus their energy consumption, while the selection of uncollected SNs is essential for improving the collection rate. Consequently, the high-level GTS problem is designed to optimize the collection rate and AUVs' energy consumption by choosing appropriate collection targets.

3) Low-level LTP Problem: Given the collection target, the low-level LTP dynamically adjusts AUV's velocity in real time to: a) plan an energy-efficient and safe trajectory to the designated location; and b) avoid collection collisions. The velocity determines the AUV's movement trajectory, which directly affects its movement energy consumption. Furthermore, velocity and trajectory determine whether the AUV can reach the collection location and its collection slot. The collection slot further influences the occurrence of collection collisions. Consequently, both the feasibility of reaching the collection location and the occurrence of collection collisions affect the collection rate. Moreover, the energy consumption of the trajectory determined by the low-level problem directly impacts the optimization results of the high-level GTS problem.

IV. Hierarchical DRL-based Collision-free Data Collection Scheme

In this section, we present HCDC scheme to address the problem described in Section III-C. Then, the high-level MA-GTS algorithm for target selection, the ABS algorithm based on collision-free condition, and the low-level DDPG-LTP algorithm are introduced for trajectory planning, respectively.

A. HCDC Overview

In the context of data collection, AUVs only receive collection rewards after successfully collection data through multiple movement steps. Such sparse collection reward result in AUVs lacking sufficient positive samples to learn an efficient collection policy. Moreover, the frequent reward introduced by other objectives further hinder AUVs from exploring uncollected areas. The sparse reward and insufficient exploration ultimately lead to collection blind spots. Furthermore, ensuring collision-free data collection is also crucial for improving collection rate. To overcome the above challenges, we decompose the MADC problem into high-level GTS and low-level LTP subproblems and propose HCDC scheme, as shown in Fig. 4. The proposed HCDC scheme consists of three components: high-level MA-GTS algorithm for collection target selection, ABS algorithm for collision-free collision, and low-level DDPG-LTP algorithm for trajectory planning.

The high-level MA-GTS algorithm employs a centralized training with decentralized execution manner to coordinate multiple AUVs's collection target selection. In MA-GTS, each AUV independently selects its target SN based on the observed states of SNs and other AUVs. Notably, the action of MA-GTS do not directly interact with the environment but instead serve as a goal to guide the lower-level DDPG-LTP algorithm toward collection location. Extrinsic reward for MA-GTS are provided by environment after DDPG-LTP completes its execution and terminates. In addition, the extrinsic reward is divided into global reward and local reward. The global reward is designed to cooperative optimize total energy efficiency, while local reward aims to prevent individual AUVs from selecting already collected SNs.

Given the collection target, AUVs collect sensing data upon reach the designated collection location. In such process, collision avoidance is critical as packet collision lead to collection failures and a reduced collection rate. Especially, packet collisions are particularly severe in underwater environments due to the narrow bandwidth and long preambles. To address this, we schedule AUVs that will collision to reach the collection area in different slots, thereby avoiding collisions. Specifically, we first analyze the conditions required for collision-free data collection. Based on the conditions, we design ABS to schedule AUVs' collection slots, ensuring collision-free collection.

Based on the target SN and collection slot, the low-level DDPG-LTP algorithm adaptively adjusts it velocity to reach collection location with in collision-free slot, while achieving energy-efficient and safe trajectory planning in complex ocean environments. The specified target SN reduces the observation space of DDPG-LTP, thereby decreasing the training complexity. Since DDPG-LTP does not directly receive rewards from the external environment, we carefully designed an intrinsic reward to guide AUV toward the collection location. The AUVs move to a new location after executing the movement trajectory generated by DDPG-LTP. Based on the new system

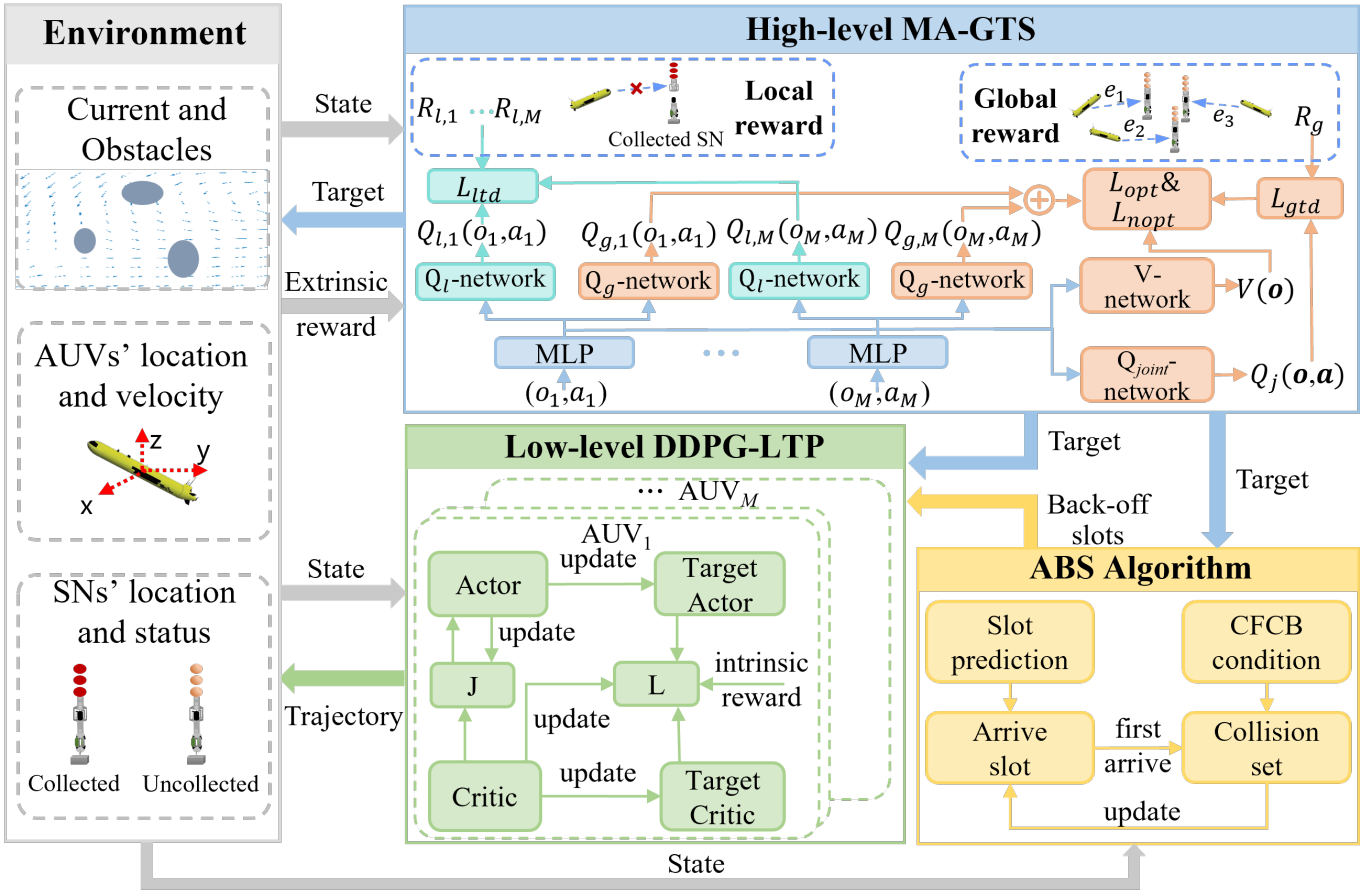


Fig. 4: Framework of HCDC.

state (AUV's new location), the ABS module recalculates the back-off slots and feeds this information back to DDPG-LTP for generating the next action. This process is continuously repeated throughout the AUVs' movement, thereby forming a feedback closed-loop mechanism between ABS and DDPG-LTP.

B. High-level MA-GTS Algorithm

Given the discrete action space of the high-level GTS subproblem, a deep Q-network is adopted to select each AUV's collection target SN. However, the dimension of the joint action space increase exponentially as the AUV number increases. To prevent the "curse of dimensionality" for action space in the multi-AUV environment, the MARL is applied in high level subproblem.

MA-GTS leverages a centralized training, decentralized execution (CTDE) framework, allowing for efficient learning in such environments. It decomposes the joint Q-value function into individual Q-value functions for each agent, while maintaining coordination among the agents through shared parameters. This method significantly reduces the complexity of learning in large multi-agent systems and helps AUVs achieve coordinated decision-making without experiencing the curse of dimensionality. Detail designs of high level algorithm are listed as follows.

1) State and Observation Design: The state space of of MA-GTS mainly consists of the locations of all AUVs and SNs, which are essential for completing the data collection task. Additionally, the state space includes whether each SN has been collected, which helps AUVs avoid selecting already-collected SNs. As for the observation space, it contains all the information in the state space, with the key difference being that the location information is transformed into the relative location vector of each SN with respect to the AUV.

2) Action Design: The local action for each AUV is to select a collection target SN, with the choice influenced by factors such as the AUV's current location, the status of nearby SNs, and coordination with other AUVs.

3) Extrinsic Reward Design: Designing an accurate reward function is crucial to not only reflect the objective of the task but also to ensure faster and more stable convergence. The high-level MA-GTS directly receives extrinsic reward form environment, which incorporate both global and local rewards: the energy consumption of the AUV and the selection penalty.

The energy consumption is calculated based on the movement energy required by the AUV from selecting a collection target to completing the collection. This process is globally optimized through AUV collaboration and task assignment. Therefore, this term is considered a global reward and is

defined as follows:

$$R_g = \sum_{i=1}^{M} \sum_{t=t_s}^{t_e} e_{i,t}, \tag{10}$$

where t_s and t_e are the start and end slots of the collection process for the current target SN.

The selection penalty is applied when AUV i mistakenly selects an already-collected SN as its target, which can be expressed as:

$$R_{l,i} = (1 - f(h_{i,t}))P_r, \tag{11}$$

where $h_{i,t} \in \mathcal{N}$ is AUV i's collection target in slot t, $f(h_{i,t})$ is a binary function that $f(h_{i,t}) = 1$ if $h_{i,t}$ has not been collected before, otherwise $f(h_{i,t}) = 0$. P_r is the constant penalty value for repeated collection. This penalty reflects an individual decision error of the AUV, which does not affect the decisions of other AUVs. Consequently, it is treated as a local reward to avoid interference with the training of other AUVs and to accelerate the convergence speed.

4) Train Design: In MA-GTS, we employ and train four neural networks: individual global action-value network $Q_{g,i}$ and local action-value network $Q_{l,i}$ for each AUV i, joint action-value network Q_{joint}, and state-value network V. These networks are trained in a centralized manner. After training, each AUV i uses its own individual action-value networks $Q_{g,i}$ and $Q_{l,i}$, and takes an action during decentralized execution by maximizing the sum of $Q_{g,i}$ and $Q_{l,i}$. The joint action-value network is trained to estimate the true action value by minimizing the global TD-error as follows:

$$L_{gtd} = \mathbb{E}\left[\left(Q_{joint}(\boldsymbol{o}, \boldsymbol{a}) - y\right)^2\right], \tag{12}$$

where $\boldsymbol{o}$ is the concatenation of AUVs' observations, $\boldsymbol{a}$ is the concatenation of AUVs' actions, and y represents the target joint Q-value, formulated as:

$$y = R_g + \gamma Q_{joint}(\boldsymbol{o}', \bar{\boldsymbol{a}}'). \tag{13}$$

where $\bar{\boldsymbol{a}}' = [\bar{a}_1', ..., \bar{a}_M']$ and $\bar{a}_i' = \arg\max_{a_i'}[Q_{g,i}(o_i', a_i') + Q_{l,i}(o_i', a_i')]$. The individual local action-value network can be trained by minimizing the local TD-error as follows:

$$L_{ltd} = \mathbb{E}\left[\sum_{i=1}^{M}(Q_{l,i}(o_i, a_i) - (R_{l,i} + \gamma \max_{a_i'} Q_{l,i}(o_i', a_i'))^2)\right], \tag{14}$$

Then, We fix Q_{joint} and use it to guide the training of $Q_{g,i}$ and V, ensuring compliance with the individual-global-max (IGM) condition [36]. Therefore, the loss function for training of $Q_{g,i}$ and V can be donated as follows:

$$L_{opt} = \left(\sum_{i=1}^{M} Q_i(o_i, \bar{a}_i) - Q_{joint}(\boldsymbol{o}, \bar{\boldsymbol{a}}) + V(\boldsymbol{o})\right)^2, \tag{15}$$

$$L_{nopt} = \left(\min\left[\sum_{i=1}^{M}(Q_i(o_i, a_i) - Q_{joint}(\boldsymbol{o}, \boldsymbol{a}) + V(\boldsymbol{o}), 0\right]\right)^2, \tag{16}$$

where $\bar{\boldsymbol{a}} = [\bar{a}_1, ..., \bar{a}_M]$ and $\bar{a}_i = \arg\max_{a_i}[Q_{g,i}(o_i, a_i)]$. Finally, we combine three loss functions in a weighted manner as follows:

$$L = L_{gtd} + L_{ltd} + \alpha L_{opt} + \beta L_{nopt}, \tag{17}$$

where α and β are the weight coefficients.

C. ABS Algorithm

Given the collection target, AUVs collect sensing data upon reach the designated collection location. In such process, packet collision lead to collection failures and a reduced collection rate. To this end, we analyze the conditions required for collision-free data collection, then design ABS to schedule AUVs' collection slots, ensuring collision-free collection.

1) Collision-free Condition: As illustrated in Fig. 5, once the AUV reaches the collection area, it send a ACTIVE packet to its target SN, which is used to activate the data transmit period. Then, the SN replies to AUV with a DATA packet. During this process, packet collisions may occur if other AUVs collect simultaneously. Due to the long propagation delay in UASNs, packet collision depend not only on their sending time but also on the propagation delay and packet length. For simplicity, we first analyze a simplified scenario consisting of a sender, a receiver, and an interfering sender. Specifically, we assume that sender i transmits a packet of size L_i to receiver j at time t_i, while interfering sender k transmits a packet of size L_k at time t_k. The packet transmission delay is calculated as follows:

$$T_{pkt} = \frac{L}{R} + T_{pre} \tag{18}$$

where L is the packet size, R is the transmission rate, and T_{pre} is the preamble delay. Let $d_{i,j}$ be the distance between i and j, and then the propagation delay can be donated as:

$$t_{i,j} = \frac{d_{i,j}}{c}, \tag{19}$$

where c is the constant sound speed, which is reasonable since the variation of sound speed is relatively small in practical ocean environments.

Packet collision occurs when the packets send by nodes i and k reach node j with overlapping time. As Fig. 6 shows, if node k send a packet during the collision window, node j fails to receive the packet from node i. Therefore, the collision-free condition $CF(j, i, k, t_i, t_k)$ for node j to receive node i's packet without collision from node k can be expressed as:

$$t_i + t_{i,j} - t_{j,k} - T_{pk} \leq t_k \leq t_i + t_{i,j} - t_{j,k} + T_{pi}, \tag{20}$$

where T_{pk} and T_{pi} represent the transmission delay of the packets send by node i and node k, respectively.

For AUV i and AUV j, packet collisions are avoided as long as they do not reach their respective collection areas in the same slot. Therefore, we analyze the collision-free collection condition under which AUV i and AUV j can simultaneously reach their collection areas. Specifically, both AUVs send an ACTIVE packet to start collection period at time t_0, as illustrated in Fig. 5. During this process, all ACTIVE and DATA packets must satisfy the CF condition. The collision-free collection condition $CFC(i, j)$ for AUV i to successfully collect data can be expressed as:

$$\begin{gathered} CF(h_i, i, j, t_0, t_0) \textbf{ AND} \\ CF(h_i, i, h_j, t_0, t_0 + t_{j,h_j} + T_{ACT}) \textbf{ AND} \\ CF(i, h_i, j, t_0 + t_{i,h_i} + T_{ACT}, t_0) \textbf{ AND} \\ CF(i, h_i, h_j, t_0 + t_{i,h_i} + T_{ACT}, t_0 + t_{j,h_j} + T_{ACT}), \end{gathered} \tag{21}$$

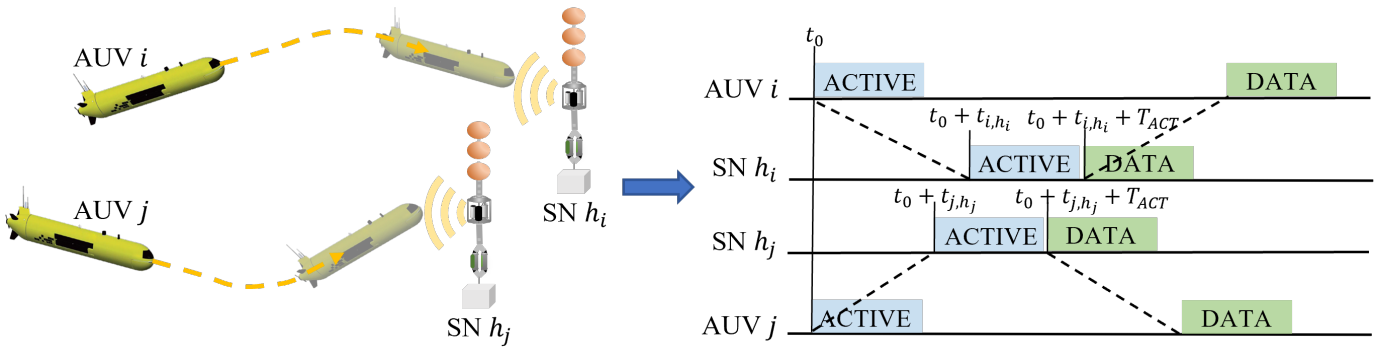


Fig. 5: AUV movement and packet exchange process during data collection.

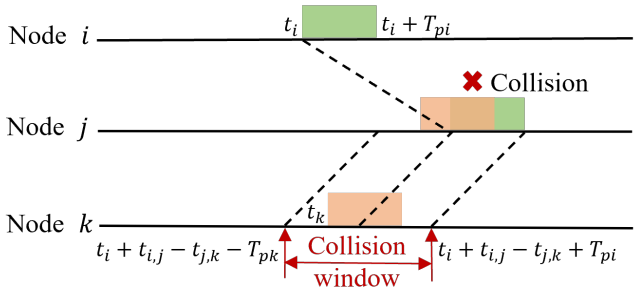


Fig. 6: Collision window of node k when node i is transmitting data to node j.

where T_{ACT} is the transmission delay of ACTIVE packet, h_i and h_j is the collection target of AUV i and AUV j, respectively. Eq. (21) indicate that both ACTIVE and DATA packet during AUV i's collection process are successfully transmitted without interference from AUV j and SN h_j. Finally, the condition $CFCB(i,j)$ for both AUV i and j to achieve collision-free data collection can be expressed as:

$$CFC(i,j) \textbf{ AND } CFC(j,i). \tag{22}$$

2) Adaptive Back-off Slot: During data collection, if AUV i and AUV j attempt to collect data within the same slot but fail to satisfy the $CFCB$ condition, which indicates that collection collision would occur, the AUVs must dynamically adjust their cruising velocities to avoid arriving at their respective collection locations simultaneously. To this end, we propose an ABS Algorithm to calculates the minimum required back-off slots for collision-free collection. Furthermore, we employ a DRL-based trajectory optimization scheme, which guides each AUV to reach its collection location within a collision-free slot, as detailed in Section IV-D.

For each AUV i, we define its collision set $\mathcal{C}_i \triangleq \{j \in \mathcal{M} \setminus \{i\} \mid \neg CFCB(i,j)\}$, which includes all AUVs that cannot collect simultaneously with AUV i. Furthermore, we employ a slot prediction model f_ϕ to estimate slot number $S_{wb,i}$ required for AUV i to reach its collection location without any back-off:

$$S_{wb,i} = f_\phi(o_i), \tag{23}$$

where o_i is the AUV i's observation, which will be further described in Section IV-D. As previously mentioned, AUV i need to back-off some slots to avoid simultaneous collection with AUVs in $\mathcal{C}_i$, ensuring collision-free collection.

To minimize the back-off delay for each AUV as much as possible, we utilize the current collection slot $S_{ab,i}$ after back-off $S_{b,i}$ slots and d_{i,h_i} to evaluate AUV i's collection priority $p(i)$. The priority of AUV i over AUV j, i.e., $p(i) > p(j)$, can be defined as:

$$S_{ab,i} < S_{ab,j} \textbf{ OR } (S_{ab,i} = S_{ab,j} \textbf{ AND } d_{i,h_i} < d_{j,h_j}), \tag{24}$$

where d_{i,h_i} is the distance between AUV i and its collection target. Based on this priority relationship, we model the AUVs and their collection order as a directed acyclic graph (DAG) G, where the vertices are AUVs, and the edges are AUVs' priority order. Specifically, there is an edge directed from i to j if and only if $j \in \mathcal{C}_i$ and $p(i) > p(j)$. In this way, AUV j's collection slot can be determined based on the collection slots of all predecessor AUV.

However, once AUV i delay its collection due to its predecessor, we need to reconstruct the edges between AUV i and its successor, as the variation in $S_{ab,i}$ results in a corresponding change in its priority $p(i)$. To avoid redundant reconstruction and improve computational efficiency, we implement an on-demand graph construction approach instead of building the complete DAG initially. Specifically, AUV i will not reconstruct edges to its successor nodes if the following condition holds:

$$p(i) > p(j), \forall j \in \mathcal{C}_i. \tag{25}$$

This is because that AUV i has no predecessors and its collection slot cannot be delayed. Thus, we only determine the collection slot for AUV i that satisfy Eq. (25) and construct edges to its successors. Furthermore, we employ a priority queue to efficiently identify the highest-priority AUV, avoiding circular comparisons between each AUV and its conflict set members.

In summary, we initialize $S_{ab,i} = S_{wb,i}$ to record AUVs' temporary collection slot, and set $S_{b,i} = -1$ to denote unvisited state. The algorithm then iteratively processes the highest-priority AUV i: calculate its back-off slot $S_{b,i}$ and update the collection slot for all unvisited AUV $j \in \mathcal{C}_i$. The detail of ABS algorithm is presented in algorithm 1.

Algorithm 1 Adaptive back-off Slot Algorithm

Input: $\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_M$
Output: $\mathrm{S}_b = [S_{b,1}, S_{b,2}, ..., S_{b,M}]$.

```
Initialize a priority queue Q
Initialize a map S_ab = [S_ab,1, S_ab,2, ..., S_ab,M]
for each AUV i ∈ M do
    Compute d_{i,h_i} and S_wb,i = f_φ(o_i)
    Add (i, d_{i,h_i}, S_wb,i) into Q
    S_ab,i ← S_wb,i
    S_b,i ← −1
end for
while Q is not empty do
    (i, d, s) ← Q.pop()
    if s < S_ab,i then
        continue
    end if
    S_b,i ← s − S_wb,i
    for each AUV j ∈ C_i do
        s_new ← s + 1
        if s_new > S_b,j AND S_b,j = −1 then
            S_ab,j ← s_new
            Add (j, d_{j,h_j}, s_new) into Q
        end if
    end for
end while
```

D. Low-level DDPG-LTP Algorithm

Based on the target SN and collection slot, we propose the low-level DDPG-LTP algorithm for each AUV to collect the target SNs. We specify the low level algorithm with the following designs.

1) State and Observation Design: Given the target assigned to each AUV, the AUV only needs to focus on the location of its target node. Therefore, the state space for each AUV consists of the location of the AUV and its corresponding target SN. The observation space, similarly, includes the relative location of the target SN with respect to the AUV. To avoid collection collision, the state and observation spaces incorporate the AUV's back-off slot calculated by the ABS algorithm based on the current system state. Additionally, considering the complexity of underwater environments, such as ocean currents and obstacles, the state and observation spaces also incorporate information about underwater obstacles and ocean currents. In particular, each AUV can only measure the ocean current velocity at its current location in real time using the H-ADCP. The lack of global ocean current information leads to a partially-observable environment, making DRL-based methods particularly suitable for this problem.

2) Action Design: For our scenario, the action space primarily consists of the AUV's movement velocity. The AUV needs to dynamically adjust its velocity to balance data collection efficiency with energy consumption and avoid collection collisions.

3) Intrinsic Reward Design: To achieve efficient data collection, we have designed a intrinsic reward function that encourages AUVs to approach the collection locations as quickly as possible while ensuring collision-free data collection and minimizing energy consumption. Additionally, avoiding collisions with obstacles is a critical aspect that must be guaranteed.

1) *Efficient Collection:* In order to enable the AUV to reach the collection location as quickly as possible and improve collection efficiency, a fixed reward is given when the AUV reaches the collection location. Conversely, we use the Euclidean distance from the current location to the target collection location to evaluate the action. Therefore, AUV i's collection reward R_c^i can express as follows:

$$R_c^i = \begin{cases} C, & \text{reach collection location} \\ \|\boldsymbol{L}_{c,i} - \boldsymbol{L}_{i,t}\|_2, & \text{otherwise,} \end{cases} \tag{26}$$

where C is the fixed collection reward, $\boldsymbol{L}_{c,i}$ is AUV i's collection location, and $\boldsymbol{L}_{i,t}$ is AUV i's current location.

2) *AUVs' Movement Energy Consumption:* To reduce AUVs' movement energy consumption, we introduce R_e^i into the reward function, which can be expressed as follows:

$$R_e^i = e_{i,t}. \tag{27}$$

3) *Collection Collision Avoidance:* AUV i adjusts its cruising velocity according to the back-off slots $S_{b,i}$ to avoid collection collisions. The collision avoidance reward can be formulated as follows:

$$R_{ca}^i = -S_{b,i} * \|\boldsymbol{v}_{i,t}\|_2, \tag{28}$$

where $\boldsymbol{v}_{i,t}$ is AUV i's velocity.

4) *Penalty Item:* Obstacle avoidance is considered an essential function of AUVs to ensure safety. The penalty is denoted as:

$$R_p^i = \rho_h P_h, \tag{29}$$

where ρ_h is binary values that denote whether AUV hits obstacles, P_h is the constant penalty value for hitting obstacles.

In summary, the reward function of AUV i is formulated as the weighted sum of the aforementioned components:

$$R_i = \lambda_1 R_c^i + \lambda_2 R_e^i + \lambda_3 R_{ca}^i + \lambda_4 R_p^i. \tag{30}$$

4) Train Design: For DDPG-LTP, we use a simple extension of actor-critic policy gradient methods. Specifically, each AUV i has four neural network: a current actor network π_i parameterized by θ_i; a target actor network $\pi_i^{'}$ parameterized by $\theta_i^{'}$; a current critic network Q_i parameterized by ω_i; a target critic network $Q_i^{'}$ parameterized by $\omega_i^{'}$. Then, each AUV can update these networks separately according to the following equations.

For each AUV i, the current critic network minimizes the loss function $L(\omega_i)$, which can be described as follows:

$$L(\omega_i) = \mathbb{E}\left[\left(Q_i(o_i, a_i) - y_i\right)^2\right], \tag{31}$$

TABLE II: Values of Main Parameters

Scenario Parameters	**Values**
Conversion efficiency of electricity ζ	0.8
Density of seawater ρ_L	1050kg/m^3
Drag coefficient C_d	0.117
AUVs' max velocity v_{max}	6kn
Sensor data size L	1000B
Algorithm Parameters	**Values**
Number of hidden layers	3
MA-GTS's neuron number of each hidden layer	(1024, 512, 256)
MA-GTS's neuron number of each hidden layer	(256, 128, 64)
MA-GTS's weight coefficient for L_{opt} α	0.5
MA-GTS's weight coefficient for L_{opt} β	0.5
DDPG-LTP's weight coefficient for R_c λ_1	1
DDPG-LTP's weight coefficient for R_e λ_2	10^{-5}
DDPG-LTP's weight coefficient for R_{ca} λ_3	0.5
DDPG-LTP's weight coefficient for R_p λ_4	1
Learning rate	0.001
Batch size M	1024
Discount factor γ	0.975
Soft update rate τ	0.01
Maximal steps per episode T	50

where o_i is AUV i's observation, and a is the action of AUV i in observation o_i, y_i is the target Q value of target critic network, formulated as:

$$y_i = R_i + \gamma Q_i^{'}(o_i^{'}, \pi_i^{'}(o_i^{'})). \tag{32}$$

Then, the critic network feeds back to actor network. Each AUV's current actor network weights θ_i can be updated according to the policy gradient method, formulated as:

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}\left[\nabla_{\theta_i} \pi_i(a_i|o_i) \nabla_{a_i} Q_i(o_i, a_i)|_{a_i=\pi_i(o_i)}\right]. \tag{33}$$

The parameters of target network $Q_i^{'}$ and $\pi_i^{'}$ are periodically copied from the current network Q_i and π_i to enhance training stability.

V. Simulation Results

In this section, we present simulation results to show the performance of our proposed scheme and answer the following questions:

- *RQ1*. Can HCDC avoid collection collision among multiple AUVs?
- *RQ2*. Does HCDC exhibit any preference toward specific SNs, and can it effectively eliminate collection blind spots?
- *RQ3*. Can HCDC exhibit better performance compared to the state-of-the-art schemes across diverse parameter settings?

A. Simulation settings

Based on the settings in our previous works [27] [37] and in other related field experiments [38], we randomly deploy 3 AUVs, 25 SNs in a three-dimensional area of 20 km $\times$ 20 km with a depth of 2000 m, where all SNs are placed on the seabed. Each AUV and SN are equipped with an acoustic communication unit for data collection, the preamble delay T_{pre} is set to 0.5s, and the communication rate R is set to 1500bps. The ocean current field is generated by the superposition of multiple vortices introduced in Section III-B, with different centers, strengths, and radii. The vortex centers are uniformly distributed across the collection area, while their strengths and radii follow a normal distribution. To ensure that the AUVs can adapt to varying environmental settings without relying on any specific current configuration, the current field is regenerated at the beginning of each training episode. The other parameters are summarized in Table II.

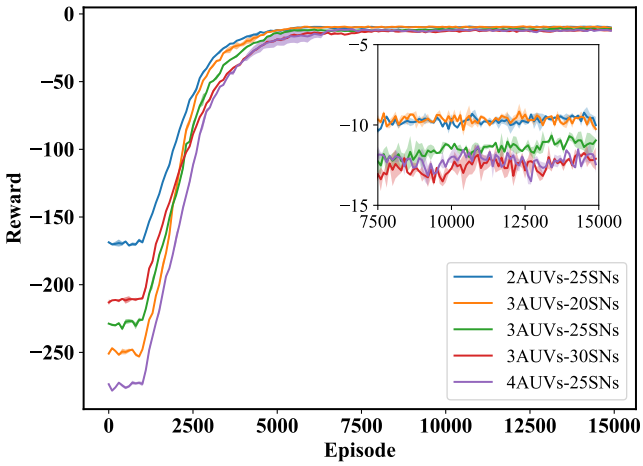
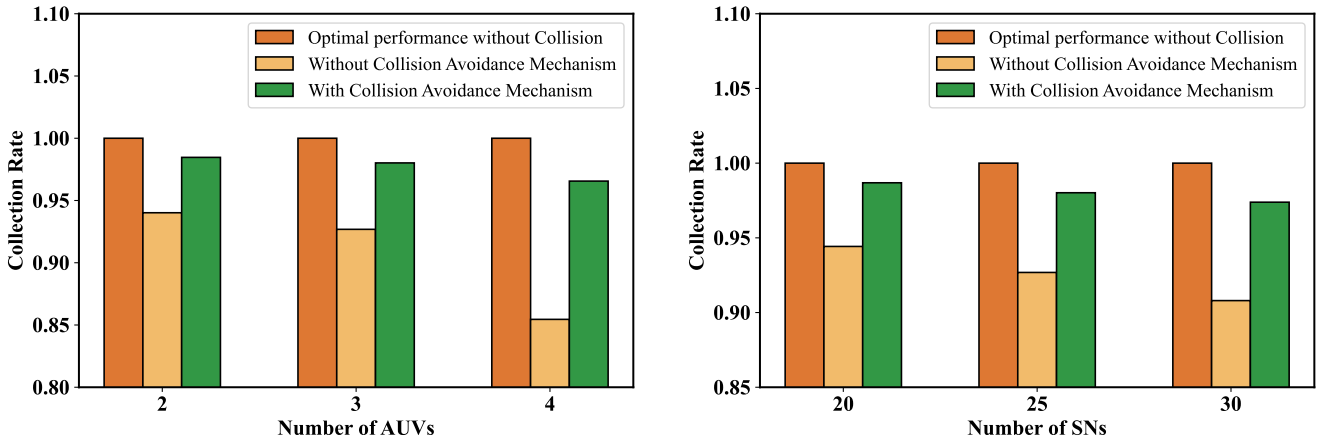


Fig. 7: Convergence curves of HCDC with varying numbers of AUVs and SNs.

B. Model Training

To evaluate the convergence of the proposed HCDC algorithm, we repeat multiple experiments in different collection scenarios with varying numbers of SNs and AUVs. Fig. 7 illustrates the learning curves of reward for 15,000 episodes. It can be observe that the reward increase rapidly during the early stages of training in all scenarios. This is because the low-level DDPT-LTP accumulates a large number of samples, enabling it to learn an effective trajectory planning policy. Meanwhile, the high-level MA-GTS focuses on selecting uncollected SNs to avoid penalties. These factors jointly contribute to the sharp increase in reward during the initial training phase. After approximately 3,000 episodes, MA-GTS begins to guide multiple AUVs in collaboratively learning the collection order of SNs to optimize the overall movement energy consumption, which leads to a slower increase in reward.

Before the training begins, the reward exhibit significant variation across different scenarios. Specifically, scenarios with more AUVs or fewer SNs tend to receive lower rewards due to a higher likelihood of selecting already-collected SNs, which leads to penalties. This is particularly evident during the early training stage, where penalty item dominates the reward function. After the training converges, the reward is primarily determined by the AUVs' total movement energy consumption. As a result, scenarios with fewer AUVs and SNs



(a) Collection rate with varying numbers of AUVs.

(b) Collection rate with varying numbers of SNs.

Fig. 8: Comparison of collection rate with and without collision avoidance under varying numbers of AUVs and SNs.

tend to receive higher reward, since less energy consumption is required to complete the data collection tasks.

C. Collection Collision Avoidance Capability of HCDC (RQ1)

In order to verify that HCDC can effectively avoid collection collision, we compare the collection rate with and without the collision avoidance mechanism in different scenarios with varying numbers of SNs and AUVs. Moreover, we construct an idealized virtual environment without any collection collisions, and using the resulting collection rate as the optimal performance upper bound.

The collection rate comparison of different numbers of AUV and SN is shown in Fig. 8. We can observe that the collection rate with the collision avoidance mechanism is significantly higher than that without it, indicating that the proposed HCDC algorithm can effectively avoid collection collisions among multiple AUVs in different scenarios. Although the collection rate with the collision avoidance mechanism is close to that of the idealized environment without any collection collisions, it still cannot achieve completely collision-free data collection. This is because, although the algorithm can accurately determine the backoff slots, the specific backoff trajectories are determined by the DRL algorithm, whose black-box nature cannot guarantee complete collision avoidance. Therefore, improving the interpretability and reliability of DRL to better adapt to complex underwater environments remains an important direction for our future work.

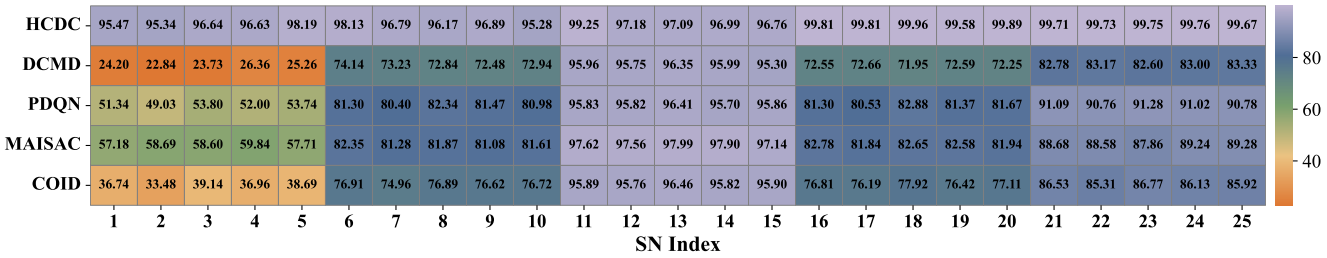
Fig. 8(a) shows the comparison result of 2, 3 and 4 AUVs. We can observe that the collection rate decreases as the number of AUVs increases. This is because a higher number of AUVs leads to a greater likelihood that multiple AUVs attempt to collect data simultaneously, thereby raising the probability of collection collisions and failures, which ultimately reduces the overall collection rate. In particular, the collection rate tends to decrease significantly as the number of AUVs increases. This is because, assuming that each AUV independently and uniformly chooses whether to collect data in a given time slot, the probability that only one AUV performs data collection in that slot decreases exponentially with the number of AUVs.

Fig. 8(b) shows the comparison result of 20, 25 and 30 SNs. Although the variation of collection rate is not very pronounced, a slight decreasing trend can still be observed as the number of SNs increases. This is because denser SN deployments lead to more frequent collection actions by AUVs. In other words, the probability that an AUV performs data collection in a given time slot increases, which significantly increases the probability of collection collision.

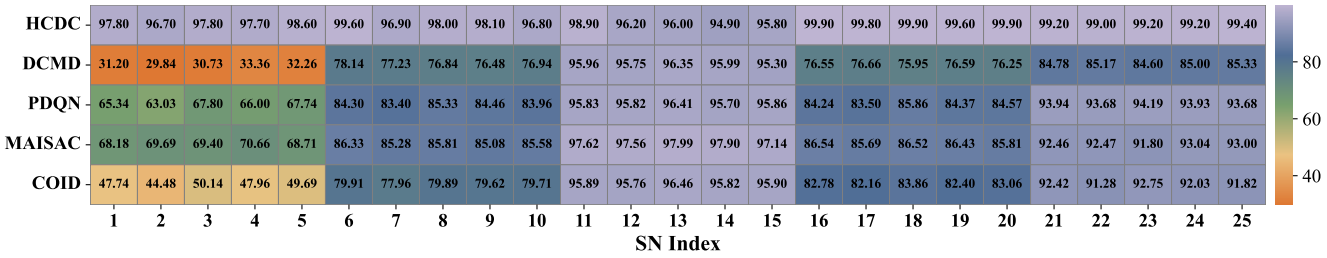
D. Collection Blind Spot Elimination (RQ2)

To verify that HCDC can effectively eliminate collection blind spots, we compare each SN's collection rate and the standard deviation of these collection rates under both large-scale (20km $\times$ 20km) and small-scale (10km $\times$ 10km) deployment areas. In this set of experiments, we compare HCDC with the following state-of-the-art schemes:

- *DCMD* [27] proposes a distributed DRL scheme specifically designed for underwater data collection. However, it overlooks the collection blind spots problem, relying solely on AUVs' random exploration to accelerate learning and improve sampling efficiency.
- *PDQN* [13] employs a parametrized deep Q-Network with a hybrid action space, enabling it to simultaneously determine collection target and movement trajectory. Moreover, it designs a continuous and effective reward based on the "goal-oriented, continuous rewards" principle to guide the AUVs to complete the collection task.
- *MAISAC* [14] also designs auxiliary rewards to guide AUVs toward their target SNs. In addition, it extends the soft actor-critic algorithm to a multi-agent version using decentralized training and decentralized execution, enabling AUVs to be trained in parallel and independently to perform their tasks in an unknown and dynamic environment.



(a) Each SN's collection rate in large area.



(b) Each SN's collection rate in small area.

Fig. 9: Each SN's collection rate across different methods.

- *COID* [39] introduces real-world data into the ocean environment model for AUV trajectory planning, effectively enhancing its adaptability to realistic underwater environment. However, it focuses on AUVs' trajectory to the target SNs, without considering the coordination of the task allocation and collection sequencing among multiple AUVs.

To investigate the relationship between an SN's data collection rate and the number of exploration steps required for successful data collection, we do not deploy SNs entirely at random. Instead, all SNs are organized into several groups, with each group randomly distributed within a confined region located at a predetermined distance from the AUVs' initial position. This design is based on the assumption that the number of exploration steps required to collect data from an SN is approximately proportional to its distance from the AUVs' initial position.

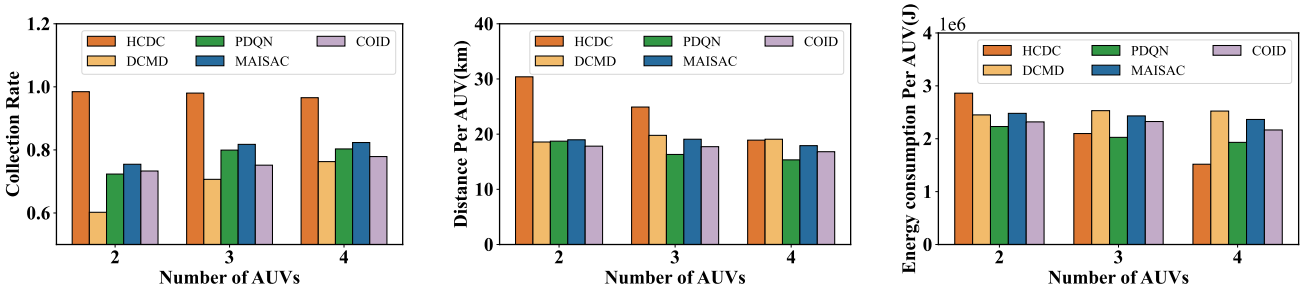
Fig. 9 shows the collection rate of each SN, while Table III presents the standard deviation of collection rates across different methods. It can be observed that each scheme exhibits varying collection performance across different SNs, and the variance of collection rates is used to quantitatively measure such preference toward specific SNs. As shown in Fig. 9, the collection rates of multi-AUV-based schemes are higher than those of single-AUV-based methods, while DCMD being the only exception. Moreover, DCMD shows the largest variation in collection rates across SNs. This is because DCMD completely ignores the collection-blind-spot problem and relies solely on random exploration to learn collection policies, which leads to a large number of SNs being left uncollected. In contrast, the other multi-AUV-based schemes fully leverage inter-AUV cooperation, resulting in significantly higher collection rates. Both COID, PDQN and

TABLE III: The standard deviation of each SN's collection rate across different methods.

Scenario	HCDC	DCMD	PDQN	MAISAC	COID
Large area	1.67	24.69	15.59	13.29	20.49
Small area	1.48	22.45	10.84	9.76	17.30

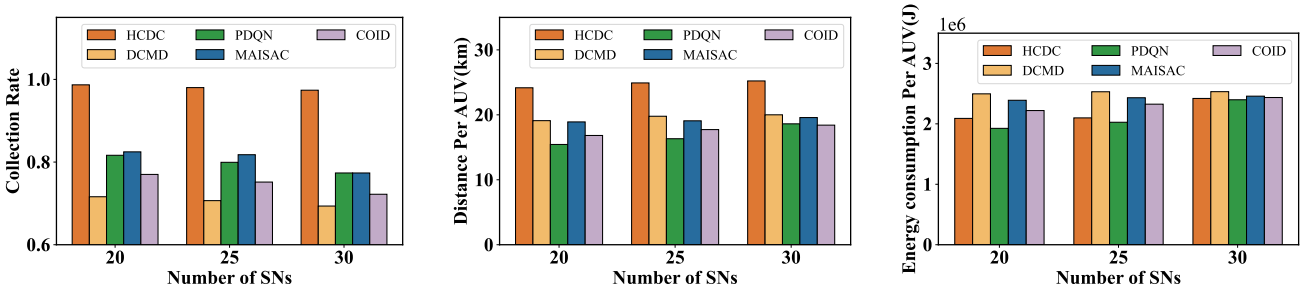
MASIAC design auxiliary rewards to guide AUVs toward their target SNs and complete the collection tasks, which effectively alleviates the problem of collection blind spots. However, due to the limitations of the auxiliary rewards, they still exhibit relatively large variations in collection rates across SNs. Although COID also optimizes the AUVs' trajectories toward target SNs, PDQN employs a parameterized deep Q-network with a hybrid action space to achieve more refined control of AUV actions. As a result, PDQN achieves a higher collection rate than COID. The proposed HCDC adopts a hierarchical DRL framework, in which the high-level MA-GTS module determines the collection targets, while the low-level DDPG-LTP module obtains intrinsic rewards based on the selected targets. This design effectively mitigating collection blind spots. Therefore, HCDC achieves a more balanced collection performance across SNs compared to other methods. Moreover, HCDC also achieves a higher collection rate than the other methods.

Table III shows that all methods exhibit smaller standard deviations in small deployment area, indicating a reduction in collection blind spots. This is because a smaller area requires fewer exploration steps to reach each SN, thereby reducing the likelihood of blind spots. Moreover, according to the collection rate of each SN under these methods, we can also validate our assumption that SNs located farther from the AUVs' starting location and requiring more exploration steps to be collected



(a) Collection rate. (b) Movement distance per AUV. (c) Energy consumption per AUV.

Fig. 10: Collection rate, movement distance, and energy consumption per AUV of HCDC and other methods with varying numbers of AUVs.



(a) Collection rate. (b) Movement distance per AUV. (c) Energy consumption per AUV.

Fig. 11: Collection rate, movement distance, and energy consumption per AUV of HCDC and other methods with varying numbers of SNs.

are more likely to become collection blind spots. Specifically, SNs (1–5) are deployed in the region farthest from the AUVs' starting location during testing, and they exhibit the lowest collection rates. In contrast, SNs (11–15), which are deployed closest to the starting location, achieve the highest collection rates. This is because SNs located near the AUVs' starting location require fewer exploration steps to be collected and are therefore less likely to become collection blind spots.

E. Performance Comparison With Other Methods (RQ3)

In this set of experiments, we compare HCDC with DCMD, PDQN, COID, and MAISAC to evaluate its performance in terms of collection rate, movement distance, and energy consumption per AUV. Fig. 10 and Fig. 11 show the average results of JCTD and other methods with varying numbers of AUVs and SNs. In Fig. 10, we set the number of SNs to 25 and the number of AUVs to 2, 3, 4. In Fig. 11, we set the number of AUVs to 3, and the number of SNs to 20, 25, 30. The total test episode number is set to 10,000.

1) Collection rate: In Fig. 10(a) and Fig. 11(a), we show the average collection rate of HCDC and other methods with different numbers of AUVs and SNs. We can observe that HCDC performs significantly better on collection rate compared to other methods. This improvement is attributed to HCDC's ability to eliminate collection blind spots, enabling more SNs to be collected, as well as its collision avoidance mechanism, which enhances the reliability of data collection. These two aspects work in concert to achieve the superior performance of HCDC. The relative performance and underlying reasons for the other methods are consistent with the analysis presented in Section V-D.

It is also noteworthy that the collection rate of HCDC decreases as the number of AUVs increases, whereas the collection rates of other methods tend to increase with more AUVs. This is because HCDC is capable of eliminating collection blind spots and collecting data from all SNs, making its performance primarily constrained by the increased likelihood of collection collisions caused by simultaneous actions of multiple AUVs. Therefore, an excessive number of AUVs increases the probability of collisions, ultimately leading to a reduction in the collection rate. In contrast, other methods are not able to collect data from all SNs. Increasing the number of AUVs helps expand their exploration area, which in turn facilitates the collection of more SNs. However, it is noteworthy that these methods are also affected by collection collisions caused by simultaneous collection from multiple AUVs. Although the collection rates of other methods tend to increase with AUVs, the growth becomes significantly slower or even stagnates, as the additional collisions introduced at higher AUV densities outweigh the potential benefits. Therefore, employing more AUVs may not completely eliminate blind spots but can increase the probability of collisions, which can ultimately reduce the collection rate. The collection rate of both HCDC and other methods decreases as the number

of SNs increases, because a higher number of SNs leads to a greater probability of collection collisions.

2) Movement Distance and Energy Consumption: Fig. 10(c) and Fig. 10(c) shows the movement distance and energy consumption per AUV of different methods with varying numbers of AUVs. When the number of AUVs is relatively small, HCDC exhibits higher movement distance and energy consumption per AUV compared to other methods. This is because HCDC collects data from more SNs, especially those located farther from the starting location. In terms of energy consumption, HCDC's disadvantage is slightly less pronounced than in movement distance. This is because AUVs may reduce their cruising velocity to avoid collection collision, which helps reducing energy consumption. As the number of AUVs increases, the movement distance and energy consumption for HCDC decrease since each AUV is responsible for fewer SNs. In contrast, the movement distance and energy consumption for other methods remain nearly constant, as they fail to collect data from all nodes and continue to explore in an attempt to cover the remaining SNs. Consequently, with more AUVs, HCDC's movement distance and energy consumption become comparable to or even lower than other methods.

In Fig. 11(b) and Fig. 11(c), we compare the movement distance and energy consumption per AUV of JCTD with other methods with varying numbers of SNs. Overall, the movement distance and energy consumption for all methods increase slightly as the number of SNs increases. This is because the AUVs' movement distance and energy consumption are primarily influenced by the size of the deployment area rather than the number of SNs. The relative performance among different algorithms is mainly affected by the number of AUVs, as analyzed in the previous paragraph.

VI. Conclusion

In this paper, we have proposed the HCDC to address the collection blind spots and the collection collision problem in multi-AUV-assisted data collection, which decomposes the MADC problem into high-level GTS and low-level LTP sub-problems. For GTS, we have designed the MA-GTS algorithm to assign target SN to each AUV, which employs global and local rewards to collaboratively optimize the overall energy consumption while avoiding individual penalty. Based on the target SN, the DDPG-LTP algorithm have been proposed to conduct AUV trajectory planning, utilizing intrinsic rewards to enhance learning efficiency and eliminate collection blind spots. Furthermore, we have analyzed the conditions for collision-free data collection and have proposed the ABS algorithm to schedule AUV collection slots. By incorporating ABS, the DDPG-LTP dynamically adjusts AUV velocities, reducing energy consumption while ensuring collision-free collection. Extensive simulation results demonstrate that HCDC achieves better collection rate and energy efficiency than the state-of-the-art schemes.

In the future, our proposed framework will be applied to real underwater environments. The most significant challenge in this process lies in filling the simulation-to-implementation gap, since the simulation environment cannot perfectly replicate complex and dynamic underwater conditions. To address this challenge, we plan to investigate reinforcement learning techniques that enhance out-of-distribution generalization, thereby improving the robustness and adaptability of our approach in practical deployments.

References

[1] S. Song, J. Liu, J. Guo, C. Zhang, T. Yang, and J. Cui, "Efficient velocity estimation and location prediction in underwater acoustic sensor networks," *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2984–2998, 2022.

[2] G. Han, W. Lai, H. Wang, and S. Zhu, "Hybrid-algorithm-based full coverage search approach with multiple auvs to unknown environments in internet of underwater things," *IEEE Internet of Things Journal*, vol. 11, no. 6, pp. 11058–11072, 2024.

[3] J. Guo, S. Song, J. Liu, H. Chen, J.-H. Cui, and G. Han, "A hybrid noma-based mac protocol for underwater acoustic networks," *IEEE/ACM Transactions on Networking*, vol. 32, no. 2, pp. 1187–1200, 2024.

[4] J. Zhang, G. Yang, G. Han, L. Liu, J. Liu, and Y. Qian, "Space/frequency-division-based full-duplex data transmission method for multihop underwater acoustic communication networks," *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 1654–1665, 2023.

[5] C. Xu, S. Song, J. Liu, M. Pan, G. Xu, and J.-H. Cui, "Joint power control and multipath routing for internet of underwater things in varying environments," *IEEE Internet of Things Journal*, vol. 12, no. 11, pp. 15197–15210, 2025.

[6] S. Yoon and C. Qiao, "Cooperative search and survey using autonomous underwater vehicles (auvs)," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 3, pp. 364–379, 2011.

[7] M. Cheng, Q. Guan, Q. Wang, F. Ji, and T. Q. S. Quek, "Fer-restricted auv-relaying data collection in underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9131–9142, 2023.

[8] S. Sun, B. Song, P. Wang, H. Dong, and X. Chen, "Real-time mission-motion planner for multi-uuvs cooperative work using tri-level programing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1260–1273, 2022.

[9] M. Cheng, Q. Guan, F. Ji, J. Cheng, and Y. Chen, "Dynamic-detection-based trajectory planning for autonomous underwater vehicle to collect data from underwater sensors," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13168–13178, 2022.

[10] S. Liu, H. Yan, L. Ma, Y. Liu, and X. Han, "Uacc-gan: A stochastic channel simulator for underwater acoustic communication," *IEEE Journal of Oceanic Engineering*, vol. 49, no. 4, pp. 1605–1621, 2024.

[11] V. Niazmand and Q. Ye, "Joint task offloading, dnn pruning, and computing resource allocation for fault detection with dynamic constraints in industrial iot," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 5, pp. 3486–3501, 2025.

[12] Q. Ye, W. Shi, K. Qu, H. He, W. Zhuang, and X. Shen, "Joint ran slicing and computation offloading for autonomous vehicular networks: A learning-assisted hierarchical approach," *IEEE Open Journal of Vehicular Technology*, vol. 2, pp. 272–288, 2021.

[13] G. Han, Z. Feng, H. Wang, Y. Hou, and F. Zhang, "Underwater multi-target node path planning in hybrid action space: A deep reinforcement learning approach," *IEEE Transactions on Mobile Computing*, vol. 23, no. 12, pp. 13033–13047, 2024.

[14] Z. Zhang, J. Xu, G. Xie, J. Wang, Z. Han, and Y. Ren, "Environment- and energy-aware auv-assisted data collection for the internet of underwater things," *IEEE Internet of Things Journal*, vol. 11, no. 15, pp. 26406–26418, 2024.

[15] T. Zhang, Y. Gou, J. Liu, S. Song, T. Yang, and J.-H. Cui, "Joint link scheduling and power allocation in imperfect and energy-constrained underwater wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 23, no. 10, pp. 9863–9880, 2024.

[16] Z. Huang and S. Wang, "Multilink and auv-assisted energy-efficient underwater emergency communications," *IEEE Internet of Things Journal*, vol. 10, no. 9, pp. 8068–8082, 2023.

[17] J. Guo, S. Song, J. Liu, H. Chen, Y. Xu, and J.-H. Cui, "Exploring applicable scenarios and boundary of mac protocols: A mac performance analysis framework for underwater acoustic networks," *IEEE Transactions on Mobile Computing*, vol. 23, no. 12, pp. 12717–12730, 2024.

[18] G. Han, Z. Zhou, Y. Zhang, M. Martínez-García, Y. Peng, and L. Xie, "Sleep-scheduling-based hierarchical data collection algorithm for gliders in underwater acoustic sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9466–9479, 2021.

[19] J. Guo, S. Song, J. Liu, H. Chen, B. Lin, and J.-H. Cui, "An efficient geo-routing-aware mac protocol based on ofdm for underwater acoustic networks," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9809–9822, 2023.

[20] 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 Transactions on Mobile Computing*, vol. 19, no. 5, pp. 1010–1022, 2020.

[21] M. T. R. Khan, S. H. Ahmed, and D. Kim, "Auv-aided energy-efficient clustering in the internet of underwater things," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 4, pp. 1132–1141, 2019.

[22] D. Wei, C. Huang, X. Li, B. Lin, M. Shu, J. Wang, and M. Pan, "Power-efficient data collection scheme for auv-assisted magnetic induction and acoustic hybrid internet of underwater things," *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 11675–11684, 2022.

[23] P. Gjanci, C. Petrioli, S. Basagni, C. A. Phillips, L. Bölöni, and D. Turgut, "Path finding for maximum value of information in multi-modal underwater wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, pp. 404–418, 2018.

[24] Z. Liu, Z. Liang, Y. Yuan, K. Y. Chan, and X. Guan, "Energy-efficient data collection scheme based on value of information in underwater acoustic sensor networks," *IEEE Internet of Things Journal*, vol. 11, no. 10, pp. 18255–18265, 2024.

[25] J. Yang, J. Huo, M. Xi, J. He, Z. Li, and H. H. Song, "A time-saving path planning scheme for autonomous underwater vehicles with complex underwater conditions," *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 1001–1013, 2023.

[26] 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 Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 108–120, 2023.

[27] S. Song, J. Liu, J. Guo, B. Lin, Q. Ye, and J. Cui, "Efficient data collection scheme for multi-modal underwater sensor networks based on deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6558–6570, 2023.

[28] B. Jiang, J. Du, C. Jiang, Z. Han, and M. Debbah, "Underwater searching and multiround data collection via auv swarms: An energy-efficient aoi-aware mappo approach," *IEEE Internet of Things Journal*, vol. 11, no. 7, pp. 12768–12782, 2024.

[29] 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, 2021.

[30] Z. Fang, J. Wang, J. Du, X. Hou, Y. Ren, and Z. Han, "Stochastic optimization-aided energy-efficient information collection in internet of underwater things networks," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 1775–1789, 2022.

[31] X. Hou, J. Wang, T. Bai, Y. Deng, Y. Ren, and L. Hanzo, "Environment-aware auv trajectory design and resource management for multi-tier underwater computing," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 2, pp. 474–490, 2023.

[32] X. Zhuo, W. Wu, L. Tang, F. Qu, and X. Shen, "Value of information-based packet scheduling scheme for auv-assisted uasns," *IEEE Transactions on Wireless Communications*, vol. 23, no. 7, pp. 7172–7185, 2024.

[33] 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*, vol. 11, no. 5, pp. 9086–9106, 2024.

[34] L. Zhang, T. Liu, and M. Motani, "Optimal multicasting strategies in underwater acoustic networks," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 678–690, 2021.

[35] E. P. M. C. Júnior, L. F. M. Vieira, and M. A. M. Vieira, "Uw-seedex: A pseudorandom-based mac protocol for underwater acoustic networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 9, pp. 3402–3413, 2022.

[36] K. Son, D. Kim, W. J. Kang, D. E. Hostallero, and Y. Yi, "Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning," *ArXiv*, vol. abs/1905.05408, 2019.

[37] S. Song, B. Huangfu, J. Guo, J. Liu, J. Cui, and X. Shen, "A digital twin-based intelligent network architecture for underwater acoustic sensor networks," *IEEE Transactions on Mobile Computing*, vol. 24, no. 9, pp. 8196–8213, 2025.

[38] J. Zhu, X. Pan, J.-H. Cui, and T. Yang, "uw-wifi: An underwater wireless sensor network for data collection and network control in real environments," in *Proceedings of the 15th International Conference on Underwater Networks & Systems*, WUWNet '21, (New York, NY, USA), Association for Computing Machinery, 2022.

[39] M. Xi, J. Yang, J. Wen, H. Liu, Y. Li, and H. H. Song, "Comprehensive ocean information-enabled auv path planning via reinforcement learning," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17440–17451, 2022.

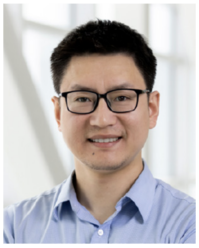
Hao Chen received the B.S. degree in computer science and technology from Jilin University, Changchun, China, in 2022, where he is currently pursuing the Ph.D. degree with the College of Computer Science and Technology. His major research interests include underwater data collection.

Jiani Guo received the BS degree (2016) in computer science and technology from Beijing Jiaotong University, Beijing, China, received PhD degree (2024) in computer science and technology from Jilin University, Changchun, China. She is currently a Post-Doctoral Researcher with the Department of Computer Science and Technology, Jilin University, Changchun, China. Her current research interests include protocol design, performance analysis, data collection, and simulation design for underwater acoustic networks.

Bowen Zhang received the B.S. degree in computer science and technology from Shandong Normal University, jinan, China, in 2023. He is currently pursuing the M.S. degree with the College of Computer Science and Technology, Jilin University, Changchun, China. His major research interests include underwater data collection.

Shanshan Song Shanshan Song received the B.S. and M.S. degrees in computer science and technology and the Ph.D. degree in management science and engineering from Jilin University, Changchun, China, in 2011, 2014, and 2018, respectively.

She was a Post-Doctoral Researcher with the department of computer science and technology, Jilin University, where she is currently a Associate Professor with the department of computer science and technology. Her major research focuses on underwater data collection, localization and navigation, and machine learning. She serves as Reviewer for IEEE/ACM Transactions on Networking, IEEE Sensors Journal, Future Generation Computer Systems. She also serves as TPC member for the WUWNet'21 Conference, and session chair for the ICCC'22 Conference.



Qiang (John) 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 2023, he has been an Assistant Professor with the Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary (UCalgary), Calgary, AB, Canada. Before joining UCalgary, he worked as an Assistant Professor with the Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada, from September 2021 to August 2023, and with the Department of Electrical and Computer Engineering and Technology, Minnesota State University, Mankato, MN, USA, from September 2019 to August 2021. He was with the Departmentof Electrical and Computer Engineering, University of Waterloo, as a Postdoctoral Fellow and then a Research Associate, from December 2016 to September 2019. He has published around 80 research papers in top-ranked journals and conference proceedings.

Dr. Ye has been selected as an IEEE ComSoc Distinguished Lecturer for the class of 2025 and 2026. He received the Best Paper Award in the IEEE/CIC International Conference on Communications in China (ICCC) in 2024 and the IEEE Transactions on Cognitive Communications and Networking Exemplary Editor Award in 2023. He is/was a General, Publication, Publicity, TPC, or a Symposium Co-Chair for different reputable international conferences and workshops, such as INFOCOM, GLOBECOM, VTC, ICCC, and ICCT. He also serves/served as the IEEE Vehicular Technology Society (VTS) Region 7 Chapter Coordinator in 2024, the IEEE Communications Society (ComSoc) Southern Alberta Chapter Vice Chair from 2024, and the VTS Regions 1–7 Chapters Coordinator from 2022 to 2023. He is also the Leading Chair of a special interest group in the IEEE ComSoc—Internet of Things, Ad Hoc and Sensor Networks Technical Committee. He serves as an Associate Editor for prestigious IEEE journals, such as IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.

Miao Pan (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Dalian University of Technology, Dalian, China, in 2004, the MASc. degree in electrical and computer engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2007, and the Ph.D.degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2012. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA. His research interests include wireless/AI for AI/wireless, deep learning privacy, cybersecurity, and underwater communications and networking. He was the recipient of the NSF CAREER Award in 2014, IEEE TCGCC (Technical Committee on Green Communications and Computing) Best Conference Paper Awards 2019, and Best Paper Awards in ICC 2019, VTC 2018, Globecom 2017 and Globecom 2015, respectively. Dr. Pan is the Editor of IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY, an Associate Editor for ACM Computing Surveys and IEEE INTERNET OF THINGS Journal (Area 5: Artificial Intelligence for IoT), and was an Associate Editor for IEEE INTERNET OF THINGS Journal (Area 4: Services, Applications, and Other Topics for IoT) from 2015 to 2018. He is also a Technical Organizing Committee for several conferences such as TPC Co-Chair for Mobiquitous 2019 and ACM WUWNet 2019.