3U: Joint Design of UAV-USV-UUV Networks for Cooperative Target Hunting

Wei Wei, *Student Member, IEEE,* Jingjing Wang, *Senior Member, IEEE,* Zhengru Fang, *Student Member, IEEE,* Jianrui Chen, *Student Member, IEEE,* Yong Ren, *Senior Member, IEEE,* and Yuhan Dong, *Member, IEEE*

***Abstract*—A swarm of unmanned vehicles can provide fine-grained spatial-temporal information acquisition and monitoring in comparison to a single agent, which is beneficial in terms of environment mapping, terrain exploration, and target hunting. However, the cooperation of single type of unmanned vehicles may be not qualified for fulfilling complex underwater tasks considering the motion constraints. In this paper, a joint design of the unmanned aerial/surface/underwater vehicle (UAV-USV-UUV) network, also referred to as 3U network, is proposed for cooperative underwater target hunting. We first introduce the advantages of this 3U heterogeneous system in multi-task cooperation and portray its system model. Moreover, we propose an energy-oriented target hunting model by jointly optimizing the UAV's position, the UUV's trajectory as well as their inter-connectivity. Finally, DQN algorithms are conceived to solve the proposed target hunting problem. Simulation results show the proposed scheme is suitable for underwater target hunting with a high success rate considering a trade-off between the system energy consumption and inter-connectivity.**

***Index Terms*—Unmanned vehicles, swarm intelligence, cooperative target hunting, deep Q-learning.**

I. Introduction

UNMANNED underwater vehicles (UUVs) have been widely used in dynamic and harsh deep sea for substituting human to fulfill observation, exploration, monitoring applications, acting as the underwater mobile sentry [1]. Equipped with kinds of sensors, they are capable of supporting fine-grained target hunting task and can quickly approach to targets given their high maneuverability [2]. Although equipped with onboard navigation systems, UUVs have to move to the sea surface periodically to update their positions so that tracking errors may be alleviated over time. Thus, only relying on UUVs for underwater target hunting leads to a low success rate of target hunting because of frequent self-position update.

It becomes popular for cooperative target hunting task based on multiple different unmanned vehicles against a single one where the success rate of target hunting can be substantially increased by fulfilling cooperation. To elaborate a little more, as the popular member of unmanned vehicle family, unmanned aerial vehicles (UAVs) play an important role in large-area surveillance and target hunting for deep sea marine targets due to their relatively low cost, extremely fast deployment as well as fully controllable mobility. Since UAVs are equipped with diverse payloads to collect and process information, they can be viewed as the 'eyes in the sky' with a large coverage area [3]. The UAV-UUV union can improve the accuracy of the positioning for underwater targets in comparison to a single type of UUVs [4]. However, UAVs and UUVs can only exchange information when UUVs rise to the surface, while electromagnetic (EM) signals used in UAV communications suffer severe attenuation imposed by the water. Hence, the communication between UAVs and UUVs is inefficient.

Fortunately, unmanned surface vehicles (USVs), as a type of vehicles on the sea surface characterized by flexibility, concealment and low cost, can be used as a 'bridge' between the air and underwater. The USV-UUV system is capable of establishing the surface-underwater target hunting network based on acoustic transmission, and improves the navigation and control performance avoiding periodical round-trip of UUVs from the surface to underwater [5]. However, compared with the high mobility and large view of UAVs, USVs and UUVs have a relatively limited target surveillance coverage.

A. Contributions

Hence, it is necessary to jointly design a UAV-USV-UUV (3U) heterogeneous system for cooperative underwater target hunting. To the best of our knowledge, few works focused on such a heterogeneous system in underwater target hunting. Specifically, Y. Wu *et al.* in [6] investigated the cooperation mechanism among UAVs, USVs and UUVs for the search-and-track mission, while the analysis for aforementioned system focused little on the effect of acoustic communication channel. Nevertheless, the unfavorable characteristics of the underwater acoustic channel, e.g., limited bandwidth and fading, have an impact on the decision-making of UUV. Therefore, it is necessary to quantify the influence of propagation channels and information flow.

This work of Jingjing Wang was supported in part by the National Natural Science Foundation of China under Grant No. 62071268 and No. 62222101, in part by the Young Elite Scientist Sponsorship Program of the China Association foe Science and Technology under Grant No. 2020QNRC001, in part by the Fundamental Research Funds for the Central Universities. This work of Yong Ren was also supported in part by the National Key R&D Program of China (No. 2020YFD0901000), and in part by 'The Verification Platform of Multi-tier Coverage Communication Network for Oceans (LZC0020)' of Peng Cheng Laboratory. (Corresponding author: Jingjing Wang.)

W. Wei and Y. Dong are with Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055. E-mail: weiw20@mails.tsinghua.edu.cn, dongyuhan@sz.tsinghua.edu.cn

J. Wang is with the School of Cyber Science and Technology, Beihang University, Beijing, 100191, China. E-mail: chinaeephd@gmail.com.

Z. Fang and J. Chen are with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China. E-mail: fangzhengru@gmail.com, cjr20@mails.tsinghua.edu.cn.

Y. Ren is with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China, and also with Peng Cheng Laboratory, Shenzhen 518055, China. E-mail: reny@tsinghua.edu.cn.

In this paper, we propose a joint 3U system aiming for constructing a three-layer network including aerial surveillance, sea surface relay as well as underwater hunting. The main contributions of this paper is summarized as follows:

- A joint 3U heterogeneous system[1] in the context of underwater target hunting is proposed with the consideration of both the system connectivity as well as cross-tier resource allocation.
- We modeld the motion-constrained problem as the Markov decision process and further achieved via DQN methods, relying on delicate unmanned vehicles' motion states definition and reward function design. Our methods guide underwater target hunting task by jointly optimizing the UAV's position, the UUV's trajectory as well as the inter-connectivity of the 3U system.
- Simulation results show that the proposed scheme has a high success rate for underwater target hunting striking a trade-off between the system's resource allocation and inter-connectivity.

B. Related Work

The problem of cooperative underwater target hunting has been widely studied. Some traditional algorithms such as the ant colony optimization (ACO) [7], the particle swarm optimization (PSO) [8] and the genetic algorithm (GA) [9] can pinpoint a static target without much effort. But when the target starts to escape with an unknown strategy, it is easy to avoid being hunted by UUVs. Thus, some researchers used reinforcement learning to model the underwater target hunting problem, but these methods were not practical when the state space was too large [10]. By contrast, deep reinforcement learning (DRL) can assist UUVs to catch the dynamically moving target by learning actions and states of themselves as well as the target. Specially, Y. Wang *et al.* in [11] modeled target tracking control as the Markov decision process and further achieved via DRL. Although the method was validated with dynamically moving targets, only a single UUV was considered. In [12], X. Cao *et al.* guided UUVs to the predicted target position, while DRL was not used to model the dynamic adversarial relationship between UUVs and the target. Meanwhile, as mentioned earlier, in many situations such as when conducting USV to transmit the control and target messages between UAV and UUVs, the connectivity of communication is a critical constraint. Consequently, it desires to have full autonomy to maintain connectivity and hunt the moving target with an intelligent DRL method. In our work, as a widely adopted DRL variant, a deep Q-learning (DQN) algorithm employing neural networks attempts to solve the cooperative target hunting in the underwater environment.

The remainder of this paper is outlined as follows. The system model and problem formulation are detailed to elaborate a joint 3U system for cooperative target hunting in Section II. In Section III, the implementation of the aforementioned united 3U system as well as a DQN based algorithm is presented. In Section IV, simulation results are provided for characterizing the proposed resource allocation and inter-connectivity model for 3U system, followed by conclusions in Section V.

[1] 3U system shows a three-layer network, where the UAV is able to adjust the search scope by adjusting the search radius. Operating as a wireless communication relay, USV can transmit the control and target messages between UAV and UUVs. Specifically, each UUV is utilized as a target hunter, which has to guarantee the constraints of connectivity and energy.

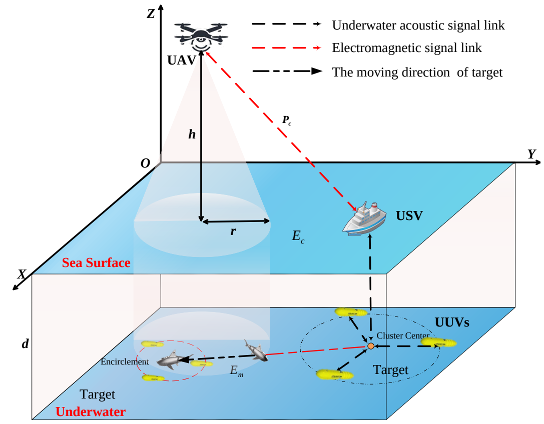


Fig. 1. 3U network for cooperative target hunting.

II. System Model and Problem Formulation

A. System Model

As shown in Fig. 1, we consider a three-tier heterogeneous system for supporting underwater target hunting, which consists of a UAV, a USV and several UUVs. For providing roughly hunting scope, one UAV serves as air-based monitor with the fly height h and the search radius r. M UUVs are adopted to execute underwater tasks with the operational depth d. Moreover, USV serves as a communication relay between UUVs and UAV. Then, we define the coordinates of the UAV and USV as $\boldsymbol{U} = (u_x, u_y, u_z)$ and $\boldsymbol{S} = (s_x, s_y, s_z)$, respectively. Without loss of generality, due to the short distance between each UUV, $\boldsymbol{G} = (g_x, g_y, g_z)$ denotes the location of each UUV without considering the formation of UUVs, which also represents the underwater search center.

The hunting missions of 3U fails when the target escapes from the search area. Moreover, the connection between UAV and USV is based on EM channel, and acoustic channel is used for the signal transmission between UUVs and USV. All UUVs team up with each other for cooperative target hunting under the constraints of connectivity and energy. During each time slot of target hunting, when UUVs receive the approximate location of the target $\boldsymbol{W} = (w_x, w_y, w_z)$ gathered by the UAV, they quickly form a hunting group in pursuit of the target. In order to facilitate the energy-efficiency of 3U, the current energy states of UUVs are transmitted back to the UAV at each time slot, and the UAV dynamically adjusts the target searching area. The analyses of 3U are discussed as follows:

1) UAV: In each time slot, the UAV is able to adjust the search scope by adjusting r under the constraints of path loss. Since the UAV is always visible to USV, the communication link between UAV and USV is reasonably modeled as a line-of-sight (LoS) path [13]. According to [14], the relationship between the altitude h and radius r can be expressed as:

$$\frac{\pi h}{9 \ln(10) r} + \frac{\eta_{\mathrm{LoS}} \exp[-\arctan(h/r) - 1]}{[\exp(-\arctan(h/r)]^2} = 0, \qquad (1)$$

where $\eta_{\rm LoS}$ is the environment-dependent loss corresponding to the LoS link.

2) USV: Operating as a wireless communication relay, USV runs randomly on the sea surface to transmit control and target messages between the UAV and UUVs. Moreover, the probability $P_{\rm c}$ of successful transmission represents the connectivity between the UAV and USV, which is given by:

$$P_{\rm c} = P[R > T_{\rm a}], \tag{2}$$

where $R = (p_{\rm a}\lambda\|\boldsymbol{U}-\boldsymbol{S}\|^{-a})/(\sigma^2+I_t)$ denotes the Signal to Interference plus Noise Ratio (SINR), and a denotes path loss exponent. $T_{\rm a}$ is the threshold of R. λ follows an exponential distribution with the mean of μ. Besides, $p_{\rm a}$ is the transmitting power. σ^2 represents noise power, while $I_{\rm t}$ denotes signal interference. The expectation $E[I_{\rm t}] = \mu\sum I_{\rm t}$ is defined in [15]. Thus, $P_{\rm c}$ can be reformulated as:

$$P_{\rm c} = e^{-(T_{\rm a}\|\boldsymbol{U}-\boldsymbol{S}\|^{a}\sigma^2+E[I_{\rm t}])/(\mu p_{\rm a})}. \tag{3}$$

As to the underwater connectivity, we assume that each UUV can be connected to the USV and other UUVs and let $\Psi = (\phi_{ij} \setminus l)_{\kappa\times\kappa}$ represent the connection between them, where $\kappa = M+1$ represents the total number of UUVs and the USV, while $\phi_{ij} = 1$ means there exists a communication link between the i-th vehicle and the j-th vehicle, otherwise $\phi_{ij} = 0$. l is the distance between transmitters and receivers equipped on UUVs and USV. Let δ_i represent the i-th eigenvalue of Ψ. According to [16], the connectivity $\bar{\delta}$ of the underwater vehicles is introduced as:

$$\bar{\delta} = \ln\Big(\frac{1}{M+1}\sum_{i=1}^{M+1} e^{\delta_i}\Big). \tag{4}$$

3) UUV: In the scenario, each UUV is utilized as a target hunter, which has to guarantee the constraints of connectivity and energy. Firstly, underwater acoustic channel is governed by the attenuation of underwater acoustic propagation [17]. Let $\Gamma(l,f) = l^{\zeta}\gamma(f)^l$ be the acoustic path loss versus carrier frequency f and distance l, where ζ is the spreading factor describing the geometry of propagation. Let $\gamma(f)$ represent the absorption coefficient, which can be given by:

$$10\log\gamma = \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4100+f^2} + 2.75\times10^{-4}f^2 + 0.003. \tag{5}$$

The total energy consumption $E_{\rm UUV}$ can be expressed as $E_{\rm UUV} = E_{\rm m} + E_{\rm c}$. Furthermore, $E_{\rm m} = t_{\rm h}\cdot E_{\rm p}$ represents the motion energy consumption, where t_h denotes the hunting time of UUVs. The energy consumption of UUVs' movement during each time slot can be expressed as $E_{\rm p} = \varepsilon\cdot F_{\rm d}\cdot\|\boldsymbol{V}_G\|$, where ε is the conversion efficiency of electricity, and $F_{\rm d}$ represents the drag force of UUVs. $\boldsymbol{V}_G$ is defined as the relative velocity between UUVs and currents at position $\boldsymbol{G}$. Thus, the total voyage distance $\boldsymbol{L}$ can be expressed as $\boldsymbol{L} = \int \boldsymbol{V}_G dt$.

Furthermore, $E_{\rm c} = E_{\rm t}(k,l) + E_{\rm r}(k,l)$ represents the communication energy consumption, where $E_{\rm t}(k,l)$ is the energy consumption of UUV when transmitting k bits over distance l, which can be given by $E_{\rm t}(k,l) = kE_{\rm u} + {\rm q}kT_{\rm b}dle^{\gamma(f)l}$. $E_{\rm r}(k,l)$ is defined as the energy consumption of receiving messages, which can be given by $E_{\rm r}(k,l) = kE_{\rm u}$. Let $E_{\rm u}$ denote the unit energy consumption to process one bit of message, while $T_{\rm b}$ denote the duration of bits to send messages. Moreover, we use a constant q to adjust the loss degree of acoustic channel.

4) Target: When the target enters the search area, the hunting task begins. We assume the target can sense the approach of UUVs and flee away at a speed of $\boldsymbol{V}_{\rm t} = V_{\rm t}\hat{\mathbf{e}}_{GW}$, where $V_{\rm t}$ is the random initial velocity of a target, and $\hat{\mathbf{e}}_{GW}$ denotes the direction from UUVs' cluster center to the target. The target's safe region with the radius r_2 is defined as $\mathcal{G} = \{\boldsymbol{J} \in O-xyz \mid ||\boldsymbol{J}-\boldsymbol{W}|| < r_2\}$. If there exists a time instant t when UUVs enter into $\mathcal{G}$, i.e., $||\boldsymbol{G}(t)-\boldsymbol{W}(t)|| < r_2$, the target is considered to be caught by UUVs.

B. Problem Formulation

1) Energy Minimum Formulation: In the 3U system, the overarching aim is to minimize the total consumption energy $E_{\rm UUV}$. Therefore, the optimization problem of 3U energy consumption can be defined as:

$$\begin{aligned}
&\min_{\{r,P_{\rm c},\bar{\delta},\|\boldsymbol{L}\|\}} \quad E_{\rm UUV} \\
s.t. \quad &\text{(6a)} \quad h_{\min} \le h \le h_{\max}, \\
&\text{(6b)} \quad \frac{p_{\rm a}\lambda\|\boldsymbol{U}-\boldsymbol{S}\|^{-a}}{\sigma^2+I_{\rm t}} \ge T_{\rm a}, \\
&\text{(6c)} \quad \ln\Big(\frac{1}{M+1}\sum_{i=1}^{M+1} e^{\delta_i}\Big) \ge {\rm C}_1, \\
&\text{(6d)} \quad \|\boldsymbol{W}-\boldsymbol{G}\| \le r, \\
&\text{(6e)} \quad \sum_{i=1}^{M}\Big\|E_i - \frac{1}{M}\sum_{i=1}^{M}E_i\Big\| \le {\rm C}_2.
\end{aligned} \tag{6}$$

2) Constraints: To elaborate further, the constraints are explained as follows.

- UAV hovering altitude constraint (6a): Considering air traffic control, the hovering altitude of UAV is limited within a permitted range of $[h_{\min}, h_{\max}]$.
- UAV-USV connectivity constraint (6b): Due to the high mobility of the UAV and the interference in EM channel, the connectivity between the UAV and the USV yields $R > T_{\rm a}$, which guarantees the cooperation of 3U system.
- USV-UUV connectivity constraint (6c): UUVs need to upload energy information and obtain the target information through the underwater acoustic channel. Considering the complexity of the underwater environment, the connectivity $\bar{\delta}$ needs to ensure reliabilities and quantities of the underwater acoustic links. The constant ${\rm C}_1$ depends on the system requirement of connectivity.
- Encircling constraint (6d): In order to avoid the target escaping from the search area, the distance between the target and the underwater search center $H = \|\boldsymbol{W}-\boldsymbol{G}\|$ should be less than the search radius r.
- Energy-balanced constraint (6e): The increasing disparity of residual energy within UUVs leads to the failure of target hunting. Therefore, the energy balance within the hunting group should be considered, which yields $\sum_{i=1}^{M}\Big\|E_i - \frac{1}{M}\sum_{i=1}^{M}E_i\Big\| \le {\rm C}_2$, where the constant ${\rm C}_2$ depends on the system requirement of energy balance.

III. DQN Based Efficient Solution

DQN can explore the environment, try several actions with different states, and eventually learn the best policy through experience [18]. Thus, this paper uses DQN method to solve the energy optimization problem in section II, while guiding UUVs to hunt the target. As shown below, the DQN model applied to UUVs consists of state, action, reward, and Q-value:

- State: We denote the state in the time slot t by $s(t) = \{U(t), S(t), G(t), W(t), \hat{\mathbf{e}}_{GW}, \boldsymbol{L}\}$, which consists of the location information of all vehicles, the direction of target escaping, and the total voyage distance of UUVs.
- Action: The action space is defined as a set of moving directions $A = \{a_i | i \in (0, 7)\}$, which divides 2π plane into eight discrete directions.
- Reward: After taking action $a(t)$, the transition from state $s(t)$ to state $s(t+1)$ produces reward $r(t)$. The reward $r(t)$ can be defined as conditional functions, which allows vehicles to learn generalized policy behavior to navigate automatically under certain conditions:

$$r\left(t\right) = \begin{cases} R_1, & \text{if } ||W(t) - G(t)|| < r, \quad (7a) \\ R_2, & \text{if } ||W(t) - G(t)|| \\ & \quad < ||W(t-1) - G(t-1)||, (7b) \\ R_3, & \text{others}, \quad (7c) \end{cases} \tag{7}$$

  where R_1, R_2 and R_3 are reward values under different system conditions. The reward R_1 corresponds to the encircling constraint (6d) when the target is within the UUV's search scope, and the reward R_2 keeps UUVs approach to the target step by step. If the system violates constraints in the optimization problem (6), it will get a negative reward R_3. Furthermore, the action is more likely to be chosen in the next time slot when it has a positive effect on $r(t)$ in the current time slot.
- Q-value: will be iteratively updated when UUVs taking action $a(t)$ at state s_t, which can be expressed as:

$$Q^*\left(s, a\right) = \mathrm{E}_{s' \sim s}[r + \chi \max_{a'} Q^*\left(s', a'\right) | s, a]. \tag{8}$$

  where $0 \leq \chi \leq 1$ is a discounting factor to decrease the weight of future rewards, while s' and a' are the state and action in the next time step.

IV. Simulation Results and Discussion

A. Experiment Settings

In computer simulations, three kinds of unmanned vehicles are located in a 400× 400 m^2 square region. The UAV and the USV are randomly distributed in the area, while UUVs' center is located in $(200,\ 200,\ -120)$ initially with M=3 UUVs. The hovering altitude h spans from 50 m to 120 m. The depth of the underwater target searching d is -120 m. As for the speed of each platform, the UAV keeps still at $(200,\ 200,\ h)$. The speed of the USV (V_S) is set to 3.9 knot [2] and V_G ranges from 3.9 to 27.3 knot. The conversion efficiency of electricity ε can be set as 80 % with the drag force F_{d} of 2000 N. Besides, V_t is set to 1 knot. The distance H between the target and the underwater search center is initialized with 100 m. Specifically, we use Table I to show the parameters of the system, the ACO algorithm and the DQN algorithm.

[2] 1 knot=1.852 km/h

TABLE I
PARAMETERS OF SYSTEM AND ALGORITHMS

	Parameters	Values
System parameters	UAV position ($\boldsymbol{U}$)	(200, 200, 120) m
	USV position ($\boldsymbol{S}$)	(200, 200, 0) m
	UUVs' initial position ($\boldsymbol{G}$)	(200, 200, -120) m
	Number of UUVs (M)	3
	Hovering altitude (h)	50-120 m
	Safe region radius (r_2)	15 m
	Depth (d)	120 m
	Speed of the USV	3.9 knot
	Speed of UUVs (V_G)	3.9-27.3 knot
	Speed of the target (V_t)	1 knot
	Conversion efficiency (ε)	80 %
	Drag force (F_{d})	2000 N
	Initial distance (H)	100 m
	Constant (C_1, C_2)	0.00007, 40
ACO parameters	Populations	100
	Iterations	100
	Pheromone polatility	0.2
DQN parameters	Learning rate (ξ)	0.01/0.001
	Iterations	10000
	Discounting factor (χ)	0.95
	Batch size	128
	Memory capacity	10000
	ϵ-greedy (ϵ)	0.9
	Terminal reward (R_1)	10
	Positive reward (R_2)	0.1
	Negative reward (R_3)	-1

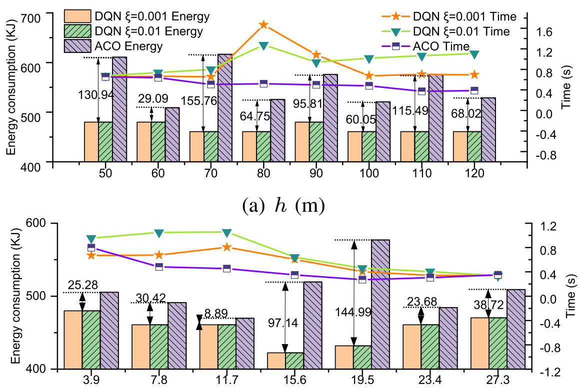
B. Simulation Results

We implement the DQN method relying on Pytorch, where the structure of DQN is established with a fully connected neural network including two hidden layers. Moreover, the DQN based hunting strategies are utilized by 3U system, while the number of iterations is 10^4. When DQN updates, a batch of 128 samples are randomly selected from the replay memory to compute the loss function with the memory capacity of 10^4. Furthermore, the exploration probability ϵ is 0.9. Besides, the terminal reward R_1, the positive reward R_2 and the negative reward R_3 are set to 10, 0.1 and -1, respectively. Furthermore, we discuss the learning rates $\xi = 0.01$ and $\xi = 0.001$ with a discounting factor $\chi = 0.95$. In this simulation, the outer performance of DQN based method comparing with the ACO algorithm [19] is verified. In order to match the 10^4 iterations of DQN, we employ 100 populations and 100 iterations in ACO algorithm. Furthermore, the communication energy E_{c} is far less than the motion energy E_{m} in reality, so we use E_{m} to represent E_{UUV} in the evaluation part.

The ACO algorithm models the behavior of UUVs for establishing the shortest path from their initial position to the target. By contrast, the DQN algorithm learns the path following the 3U target hunting task and chooses correct actions to cope with different states. During the target hunting process, DQN chooses an associated corresponding actions related to UUV motion and environment states. Then, UUVs accomplish the target hunting task by executing the actions. The results of 3U system under different UAV height (h) and UUVs' speed (V_G) are presented in Fig. 2 and Fig. 3, respectively. For each (h, V_G) pair, the results of ACO, DQN ($\xi = 0.001$) and DQN ($\xi = 0.01$) are recorded. For each algorithm, the minimum values of E_{UUV} during 10^4 iterations are given to compare the energy consumption in Fig. 2. Meanwhile, Fig. 3 shows the average minimum values

TABLE II
PERFORMANCE OF 3U SYSTEM AND TARGET HUNTING RELYING ON BOTH ACO AND DQN ALGORITHMS.

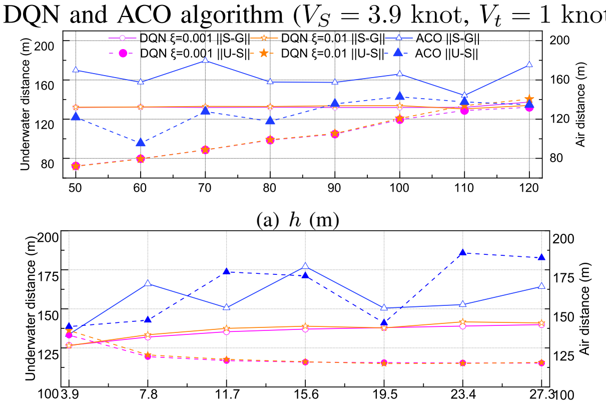
Parameters & Indicators		ACO algorithm*				DQN algorithm				Double DQN algorithm				Dueling DQN algorithm			
Initial settings	H	50	75	100	125	50	75	100	125	50	75	100	125	50	75	100	125
	ξ	—	—	—	—	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	V_t	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	V_S	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9
	V_G	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8
Perfor-mances	E_{UUV}	217.7	366.5	481.6	587.4	192	326.4	460.8	633.6	192	326.4	326.4	614.4	192	326.4	480	595.2
	$\|\boldsymbol{L}\|$	136.1	229.1	301.1	367.1	120	204	288	396	120	204	288	384	120	204	300	372
	$\|\boldsymbol{U}-\boldsymbol{S}\|$	110.1	133.2	142.8	151.7	116.2	117.7	119.5	118.7	115.9	117.8	119.8	118.7	115.5	117.4	119.7	118.9
	$\|\boldsymbol{S}-\boldsymbol{G}\|$	129.1	130.4	166.1	150.7	127.9	129.4	131.9	135.1	127.7	129.3	131.6	135.0	127.6	129.2	131.9	134.9
	Success rate	86.2%	49.3%	43.9%	58.6%	98.4%	97.9%	93.5%	66.2%	99.4%	98.8%	95.5%	67.9%	99.6%	99.5%	79.1%	64%
	Training time (h)	—	—	—	—	1	1.5	2.2	2	4	6	8.5	9	2.1	3.2	6	4.5



(a) h (m)

(b) V_G (knot)

Fig. 2. UUV's energy consumption versus h and V_G relying on DQN and ACO algorithm ($V_S = 3.9$ knot, $V_t = 1$ knot).



(a) h (m)

(b) V_G (knot)

Fig. 3. Distance versus h and V_G in different platforms ($V_S = 3.9$ knot, $V_t = 1$ knot).

of P_c and $\bar{\delta}$ during 10^4 iterations to compare the average path connectivity when 3U system conducting hunting tasks.

The UAV height h is a parameter for E_{UUV} because of its impact on searching radius r. Fig. 2(a) reveals the minimum E_{UUV} found by different algorithms versus h with V_G of 7.8 knot. What can be clearly seen is the energy consumption in DQN ($\xi = 0.001$) and DQN ($\xi = 0.01$) are always lower than that in ACO. It proves the effectiveness of DQN with the fact that DQN can always find an optimized path to quickly hunt the target with the minimum energy and maintain the connectivity when h changes. Moreover, Fig. 2(a) also shows the average time to find the optimal path for each iteration as h increases. When the search scope becomes larger, the target hunting task can be more complex. Thus, there has been a steady increase in the average search time of DQN ($\xi = 0.01$) since the convergence of DQN algorithm with higher learning rate becomes slower as the complexity of the hunting task increases. For ACO, the average search time is relatively small because of the parallel searching mechanism among populations in each iteration. V_G is another parameter affecting E_{UUV}. Similarly, Fig. 2(b) shows that DQN always finds an optimized path with the minimum energy under different V_G and ACO performs worse than DQNs (h is set to 100 m). Moreover, there exists a steady decline in the average time in both DQNs and ACO resulting from the increase in speed.

We further investigate the impact of h and V_G on path connectivity. Since the path connectivity of EM channel P_c and underwater acoustic channel $\bar{\delta}$ are negatively correlated with the distance between heterogeneous vehicles, we measure P_c and $\bar{\delta}$ with average $\|\boldsymbol{U}-\boldsymbol{S}\|$ and average $\|\boldsymbol{S}-\boldsymbol{G}\|$ during 10^4 iterations, respectively. It can be clearly seen in Fig. 3 that DQN algorithms with ξ have little difference in influence on path connectivity, and in contrast to ACO algorithm, both DQN ($\xi = 0.001$) and DQN ($\xi = 0.01$) can achieve a higher P_c and $\bar{\delta}$. Fig. 3(a) shows average $\|\boldsymbol{U}-\boldsymbol{S}\|$ of DQNs climbs as h increases, which can leads to a decrease in P_c. On the other hand, average $\|\boldsymbol{S}-\boldsymbol{G}\|$ of DQNs in Fig. 3(a) is smooth, resulting from the fact that h has little influence on the cooperation between the USV and UUVs. With the growth of V_G, Fig. 3(b) reveals that there has been a gradual decline in average $\|\boldsymbol{U}-\boldsymbol{S}\|$. This is mainly because the USV travels shorter distance when conducting underwater target hunting task as V_G increases, thus the distance between the UAV and USV decreases and P_c can be improved steadily. Meanwhile, we can see average $\|\boldsymbol{S}-\boldsymbol{G}\|$ goes up as V_G increases. It is obvious because an increase in V_G can directly affect the distance between UUVs and USV, thus the connectivity $\bar{\delta}$ of the underwater channel gets worse.

TABLE II presents the experimental data on different initial distance H between UUVs and target. It shows the minimum energy consumption optimized by the improved DQN algorithm compared with the ACO algorithm. The energy consumption increases with the growth of H, and DQN algorithm can always find a better minimum energy consumption scheme than ACO algorithm. Moreover, the total voyage distance $\|\boldsymbol{L}\|$ of UUVs is optimized and increases with the growth of H, and r doesn't change with H in TABLE II because of a fixed h in initial settings. Furthermore, the optimization variables P_c and $\bar{\delta}$ are negatively correlated with the distance between heterogeneous vehicles $\|\boldsymbol{U}-\boldsymbol{S}\|$ and $\|\boldsymbol{S}-\boldsymbol{G}\|$. Then, we can draw a conclusion that P_c and $\bar{\bar{\delta}}$ get worse as H increases. This is mainly because both UUV and USV have to travel longer distances as the distance of target appearing away from UUVs rises, and the connectivity between heterogeneous systems decreases. Finally, we define the success rate of target hunting as the proportion of UUVs entering the region $\mathcal{G}$ during 10^4 experiments. It can be observed that DQN method has a success rate of more than 95% when H equals to 50, 75 and 100. By contrast, ACO always has a relatively lower success rate. This is mainly because DQN gradually learns the constraints of 3U system and the target hunting task through interaction with the environment at each step, and also gains the behavior of the target. Although ACO performs well in path planning, it cannot always maintain the connectivity of 3U system as well as other constraints, and cannot judge the escape direction of the target. Thus, it causes the failure of the target hunting task. To further verify the effectiveness of the proposed target hunting system, we conduct experiments with more state-of-the-art DRL algorithms. Simulating results show Double DQN algorithm [20] and Dueling DQN algorithm [21] can improve the success rate of underwater target hunting task at the cost of time complexity.

V. Conclusion

In this paper, we proposed a joint design of 3U system and energy optimization problem considering the trajectory scheduling and path connectivity in the context of a multi-platform aided heterogeneous network for cooperative target hunting task. Then, based on the modified DQN algorithm, we addressed the issue of maintaining the connectivity and obtaining a near-optimal UUV trajectory. Simulations were conducted and results showed that the DQN-based energy optimization scheme achieved a higher energy efficiency than the ACO algorithm. The connectivity between heterogeneous vehicles of 3U system can be guaranteed when conducting cooperative target hunting task. Furthermore, as the DQN algorithm can assist 3U system to learn the cooperation within the heterogeneous vehicles and the pursuit-evasion game with the target, we can get a success rate of target hunting more than 95% in most cases with our DQN-based solution to 3U system. The future work will focus on applications of the proposed 3U system to cooperating target hunting task in underwater environment: in particular, we will further enrich the UUV strategy in a multi-target environment. The dynamics of UUVs will be adopted to simulate the real-underwater environment, while targets will develop more intelligent escape strategies.

References

[1] S. Guan, J. Wang, C. Jiang, R. Duan, Y. Ren, and T. Q. S. Quek, "Magicnet: The maritime giant cellular network," *IEEE Communications Magazine*, vol. 59, no. 3, pp. 117–123, Mar. 2021.

[2] 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, pp. 1775–1789, Feb. 2022.

[3] R. S. de Moraes and E. P. de Freitas, "Multi-UAV based crowd monitoring system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 2, pp. 1332–1345, Apr. 2020.

[4] P. B. Sujit and S. Saripalli, "An empirical evaluation of co-ordination strategies for an AUV and UAV," *Journal of Intelligent and Robotic Systems*, vol. 70, no. 1-4, pp. 373–384, Apr. 2013.

[5] C. Hu, L. Fu, and Y. Yang, "Cooperative navigation and control for surface-underwater autonomous marine vehicles," in *IEEE Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chengdu, China, Dec., 2017, pp. 589–592.

[6] Y. Wu, K. H. Low, and C. Lv, "Cooperative path planning for heterogeneous unmanned vehicles in a search-and-track mission aiming at an underwater target," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6782–6787, Jun. 2020.

[7] X. Yu, W.-N. Chen, X.-M. Hu, T. Gu, H. Yuan, Y. Zhou, and J. Zhang, "Path planning in multiple-AUV systems for difficult target traveling missions: A hybrid metaheuristic approach," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 3, pp. 561–574, Sep. 2020.

[8] Z. Yan, J. Li, Y. Wu, and Z. Yang, "A novel path planning for AUV based on objects' motion parameters predication," *IEEE Access*, vol. 6, pp. 69 304–69 320, Nov. 2018.

[9] 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. 14 559 – 14 571, Oct. 2021.

[10] W. Shi, S. Song, C. Wu, and C. L. P. Chen, "Multi pseudo Q-learning-based deterministic policy gradient for tracking control of autonomous underwater vehicles," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 12, pp. 3534–3546, Dec. 2019.

[11] Y. Wang, C. Tang, S. Wang, L. Cheng, R. Wang, M. Tan, and Z. Hou, "Target tracking control of a biomimetic underwater vehicle through deep reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems, (DOI: 10.1109/TNNLS.2021.3054402)*, Feb. 2021.

[12] X. Cao and X. Xu, "Hunting algorithm for multi-AUV based on dynamic prediction of target trajectory in 3D underwater environment," *IEEE Access*, vol. 8, pp. 138 529–138 538, Jul. 2020.

[13] R. Ma, R. Wang, G. Liu, W. Meng, and X. Liu, "UAV-aided cooperative data collection scheme for ocean monitoring networks," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 222 – 13 236, Sep. 2021.

[14] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal lap altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, Dec. 2014.

[15] Q. Wang, H. Dai, Q. Wang, M. K. Shukla, W. Zhang, and C. G. Soares, "On connectivity of UAV-assisted data acquisition for underwater Internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5371–5385, Jun. 2020.

[16] Q. Liang, T. Sun, and D. Wang, "Reliability indexes for multi-AUV cooperative systems," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 179–186, Feb. 2017.

[17] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *Acm Sigmobile Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[19] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 26, no. 1, pp. 29–41, Feb. 1996.

[20] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *AAAI Conference on Artificial Intelligence (AAAI)*, Phoenix, Arizona, USA, Feb., 2016, pp. 2094–2100.

[21] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *International Conference on Machine Learning*, New York, New York, USA, Jun., 2016, pp. 1995–2003.