Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 2524764, 19 pages
https://doi.org/10.1155/2022/2524764

Research Article

Reinforcement Learning Based Mobile Underwater Localization for Silent UUV in Underwater Acoustic Sensor Networks

Ruiheng Liao, Wei Su, Xiurong Wu, and En Cheng

Information and Communication Engineering, Xiamen University, Xiamen, China

Correspondence should be addressed to Wei Su; suweixiamen@xmu.edu.cn

Received 24 June 2022; Revised 8 September 2022; Accepted 16 September 2022; Published 7 October 2022

Academic Editor: Xin Ning

Unmanned underwater vehicles (UUVs) that are widely utilized for underwater cooperative combat, underwater environment detection and underwater resource exploration have to be localized by underwater acoustic sensor networks (UASNs). However, the localization accuracy is hard to guarantee due to the limited bandwidths, long propagation latency, and limited energy resources of the UASNs. In this paper, we propose a reinforcement learning (RL) and neural network based mobile underwater localization scheme to optimize the anchor nodes selection in the UASNs to localize the target precisely. More specifically, this scheme applies SqueezeNet to select the line-of-sight (LOS) anchor nodes based on the received signals. In addition, an RL-based approach is further proposed to make further selection from the LOS anchor nodes without knowing the underwater environment model. The Dyna architecture is applied to reduce the convergence time of the anchor nodes selection. Simulation results based on a nonisovelocity geometry-based underwater acoustic channel model show that the proposed schemes significantly improve the localization accuracy and reduce energy consumption of the UASN to achieve trajectory correction.

1. Introduction

The location service enables unmanned underwater vehicles (UUVs) to arrive in the target area in time, collect effective information, and return safely in applications such as underwater communication, underwater exploration, and underwater environment detection. [1, 2] The equipped strapdown inertial navigation system (SINS) navigates the UUVs from the starting points to the destinations. However, the cumulative error of the SINS increases over time.

Underwater acoustic sensor networks (UASNs) that consist of a variable number of sensors the UUVs assist the UUVs to localization [3]. UASNs that eliminate the need for cables and do not interfere with shipping activities are envisioned to enable applications for environment monitoring of physical, chemical, and biological indicators, tactical surveillance, disaster prevention, assisted navigation, and undersea exploration [4]. The Communication Signal Propagation Loss Localization Scheme (CSPLLS) proposed in [5] uses the communication signal strength information to calculate the distance from fix number of anchor nodes to assist localizing target, which is a typical transmission loss-distance based cooperative passive localization scheme. An efficient packet transmission scheduling algorithm proposed in [6] for underwater acoustic communications overcomes the difficulty of the long propagation delay in UASNs. Node cooperation (NC) based on the fact that underwater nodes can overhear the transmission of the others proposed in [7] can increase the data collection efficiency for the surface node in UASNs. A node selection algorithm for UASN based on particle swarm optimization proposed in [8] improves the energy utilization of nodes, balances positioning performance as well as energy use efficiency, and optimizes the positioning result of UASN. Consequently, it is foreseeable that an underwater acoustic sensor network which covers key sea areas will be established in the near future. However, the underwater localization through UASNs is a challenge compared with the terrestrial localization due to limited

coverage area, the time-varying of the complex underwater environment, and the depth-dependent sound speed profile [9]. Especially, the non-line-of-sight (NLOS) acoustic signals receiving from UASNs lead to the low-precision localization results and high energy consumption. The specific summary is as follows.

(1) The propagation speed of underwater acoustic signal is approximately 1500 m/s, which is 5 orders of magnitude lower than that of the radio signal, causing higher latency and longer end-to-end time [10]

(2) The coverage area of a single underwater acoustic location anchor is no more than a few dozen square kilometers, and thus, the UASNs with limited energy resources can only cover some critical areas. Hence, only parts of the UUV voyage are localized by the UASNs in most instances

(3) The non-line-of-sight (NLOS) signals affect the receiving delay of the acoustic signal and cause the distance measurement error between the anchor node and the target, which degrade the location accuracy in dynamic underwater environments

In this paper, a UUV mobile underwater localization scheme based on reinforcement learning (RL) and neural network techniques is proposed to improve the localization accuracy with less UASN energy consumption. To be specific, firstly, the signal processing chips are installed on the UUV and directly process the received localization signals without transmitting the signals to the land monitoring center, which ensures real-time performance of data processing, reduces the communication overhead, and improves the concealment of UUV. Then, UUV classifies the received signals from anchor nodes with the lightweight convolutional neural network (CNN) to determine the type of the anchor node in this localization cycle, i.e., the LOS or NLOS anchor node. UUV selects the combination of the LOS anchor nodes to determine the location, which is more accurate than that determined by the NLOS anchor nodes in this localization cycle. The process of anchor nodes selection can be formulated as a Markov decision process (MDP) and the underwater channel model is hard to obtain because of its the nonisovelocity property and multiple reflections on the sea surface and bottom [11], in which the RL technique can be applied to determine the optimal selection policy based on the observed state. The state consists of the selected anchor nodes, energy consumption, and localization error and is selected via trail-and-error. Consequently, the optimal selection policy is determined without relying on the underwater channel model. Moreover, the proposed scheme uses the Dyna architecture to generate anchor nodes selection simulated experiences and thus reduces the convergence time in the framework of RL.

The main contributions of this paper are outlined as follows:

(1) We investigate the silent UUV localization problem in UASN. Meanwhile, we have designed the entire underwater motion positioning framework, including UUV motion tracking model, UASN energy consumption model, underwater channel, and signal receiving and transmitting model, in which we have located the UUV and calculated the energy consumption of UASN. In UUV motion tracking model, UUV tracks the optimal path obtained by the path planning algorithm to reach the destination. In the UASN energy consumption model, we calculate the energy consumption of each anchor node in the UASN. In the underwater channel, we adopt a new nonisovelocity geometry-based underwater acoustic channel, in which acoustic signals sent from anchor nodes reach the target through multiple reflections on the sea surface and bottom as well as refraction between different sound velocity layers

(2) We apply the lightweight neural network to distinguish the type of the anchor nodes based on the received anchor node signals, which reduces the collection of bad signals and improves the localization accuracy. The lightweight neural network can be trained faster because of less communication and is feasible for deployment on memory limited hardware, which are the advantages of applying to UUV. A reinforcement learning based mobile underwater localization scheme is proposed to select the optimal anchor nodes from LOS anchor nodes, which further improves the localization performance and reduces the energy consumption of the UASN system. Dyna architecture is applied to reduce the convergence time of the learning process

(3) An underwater trajectory correction framework is proposed, which introduces the acoustic signals in the background of UASN. Meanwhile, the signal process module is placed on the UUV to improve the real-time performance. We apply the location of UUV obtained from RL-based mobile underwater localization scheme to the underwater trajectory correction framework to change the motion state of UUV, which reduces the error between the actual path and the ideal path

(4) We theoretically derive the Cramer-Rao lower bound (CRLB) of the proposed scheme. Simulations are performed to evaluate the performance of the proposed scheme in terms of the localization accuracy, the energy consumption, the utility, and the CRLB which are compared with the benchmarks

The structure of this paper is shown as follows. First, we review related work in Section 2. Then, the system model is presented in Section 3 and the underwater path planning algorithm and reinforcement learning based underwater mobile localization algorithm are presented in Section 4. The CRLB of the proposed scheme is derived in Section 5. Finally, we provide the simulation results in Section 6 and conclude the work in Section 7.

2. Related Works

Up to now, there are many researches on underwater moving target localization, underwater moving target trajectory tracking and correction, resolution of underwater LOS signals and NLOS signals, and the underwater optimal path planning. For instance, an inertial trajectory prediction system proposed in [12] applies inertial sensors to predict the trajectory of the autonomous underwater vehicle (AUV) and uses the Kalman filter method to reduce the accumulation of errors. An error-based adaptive model predictive control and a proportional derivative controller designed in [13] combine a real-time acoustic localization system to guide UUV towards sensor nodes installed on surface ships, and a hybrid acoustic-optical underwater communication scheme is proposed, in which the acoustic link is used for NLOS transmission and the optical link is used for LOS transmission. By coordinating these two complementary technologies, they can overcome their respective weaknesses to achieve precise localization tracking and high-speed underwater data transmission. An integrated navigation algorithm based on deep learning model as proposed in [14] deals with Doppler velocity measurement (DVL) failure to improve the SINS/DVL integrated navigation system when DVL is polluted by outliers and interrupted. The effectiveness of this proposed algorithm is verified through comparison with related work. A navigation strategy based on D∗Lite search algorithm as proposed in [15] chooses the optimal path to the destination that avoids the obstacles and reduces the travel time. A tracking algorithm based on second-order time difference of arrival (TDOA) combined with particle filters proposed in [16] eliminates the unknown signal period and overcomes the traditional limitations of the TDOA-based method. A mobile beacon-based iterative location (MBIL) mechanism proposed in [17] obtains a higher localization rate in a shorter time, which effectively reduces the localization error and extends the service life of UASN. A two-step classifier based on signal strength and propagation delay range measurements proposed in [18] can accurately distinguish between LOS and NLOS links.

Meanwhile, RL has been applied in the underwater communication and localization. For example, an RL-based energy-efficient underwater localization algorithm proposed in [19] applies Dyna-Q to reduce the localization error and the energy consumption. An unsupervised wireless localization method proposed in [20] applies deep RL to reduce localization error. An RL-based localization algorithm as proposed in [21] obtains the positions of the UUV, active sensor nodes, and passive sensor nodes by performing an online value iteration process as well as applies ray compensation strategy and the mobility compensation strategy to improve the localization accuracy. An underwater multimodal communication scheme based on reinforcement learning is proposed in [22] to improve the reliability of the underwater network and reduce the delay of underwater applications via the relay selection. In addition, reinforcement learning can also be applied in navigation. Although localization and navigation are different concepts, there is a strong correlation between them. Navigation based on reinforcement learning is investigated in plenty of works. For example, a massive MIMO UAV navigation scheme proposed in [23] applies deep RL to select the optimal strategy based on the received signal strength to improve the navigation performance. An end-to-end navigation strategy based on deep RL proposed in [24] converts the results of laser ranging into motion actions and achieves map-free navigation in a complex indoor environment. A hybrid and hierarchical reinforcement learning method proposed in [25] optimizes the learning effect through different learning methods, different types of status information, and reward distribution system to achieve robot online guidance and navigation tasks. A navigation based on supervised learning and fuzzy reinforcement learning proposed in [26] applies the best action of fuzzy rules to achieve robot navigation. A new incremental learning algorithm proposed in [27] merges new information into the exiting environment and weakens the conflicts between them in advance to greatly improve the convergence rate of reinforcement learning in a dynamic environment. A tracking algorithm based on partial reinforcement learning neural network proposed in [28] is introduced into the wheeled mobile robotic system to track the trajectory by controlling the time-varying advance angle.

Neural networks have been also applied in different applications including signal recognition and classification. For instance, an efficient convolutional neural network (CNN) is used to classify the acoustic signals of reinforced concrete (RC), which outperforms typical feature extraction and traditional machine learning based methods [29]. A deep belief network based modulation recognition scheme for wireless signals as proposed in [30] reaches 92.12% recognition rate under high signal-to-noise. A CNN-based satellite link interference signal classification proposed in [31] classifies 5 types of interference signals with strong robustness, including audio interference, narrowband interference, pulse interference, sweep interference, and spread spectrum interference. A deep neural network framework combined with multitask learning proposed in [32] improves the learning efficiency of modulation and wireless signal classification accuracy.

Inspired by above related works, we focus our research on the mobile underwater localization, underwater navigation, and signal recognition.

3. System Model

3.1. Application Scenarios. The starting point of the UUV is $J_0 = [x_0, y_0, z_0]$ and the destination is $J_d = [x_d, y_d, z_d]$. The position of the i-th anchor node is $U_i = [x_i, y_i, z_i]$. The error of the SINS will gradually accumulate over time without the UASNs assistance. Thus, the route of the silent UUV is prearranged to cross the UASNs and we use the established UASN in the sea area to send acoustic signals to UUV. Meanwhile, due to the influence of underwater terrain and ships on the sea surface, many communication links in real life between anchor node and UUV are NLOS links, which will greatly affect the localization accuracy. In order to solve this problem, we classify the received signals according to

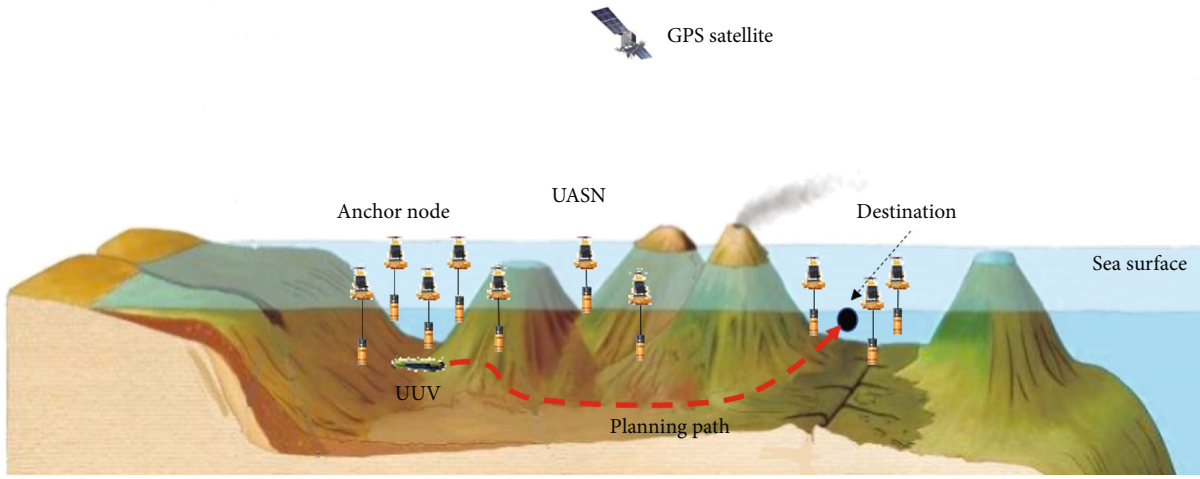


Figure 1: Presentation of the application scenario.

the neural network mounted on the UUV and select the optimal anchor nodes based on reinforcement learning. In the range of UASN-assisted localization, the silent UUV can receive signals to localize itself and correct the trajectory so that it gradually approaches the ideal path. The whole system model is shown in Figure 1 and the framework of the whole system is shown in Figure 2.

3.2. UUV Motion Tracking Model. UUV has autonomous navigation, which is composed of Doppler velocity log (DVL), gyroscope, depth gauge, and so on. Taking into account the uncertainty of ocean currents and acoustic speed, we denote the entire state space vector of UUV as $St_i = [x_i, y_i, z_i, \theta_i, v_i, u_{1i}, u_{2i}]$ where the state variables u_{1i} and u_{2i} are the measurement noises on the UUV velocity and yaw, respectively. It is assumed that the measurement noises of velocity and yaw are both zero-mean Gaussian noises with the variances ϑ_1^2 and ϑ_2^2 [33] We assume that ocean currents only occur in the x and y directions which is discovered in [34, 35]. The self-propelled velocity v_i is on the xOy plane, also called thrust velocity, which is directly measured by the DVL. The yaw θ_i is the heading angle of the UUV on the horizontal plane, which is directly measured by the on-board compass. The UUV coordinate (x_i, y_i) at time i is obtained through the inertial navigation system and the depth z_i is directly obtained through the UUV's own depth gauge. Then, the entire UUV motion model is given by

$$\begin{aligned} x_{i+1} &= x_i + (v_i + u_{1i}) \cdot \cos(\theta_i + u_{2i}) \cdot t, \\ y_{i+1} &= y_i + (v_i + u_{1i}) \cdot \sin(\theta_i + u_{2i}) \cdot t, \end{aligned} \tag{1}$$

where t is the time interval.

UUV usually needs to reach the destination from the departure when performing tasks. In order to save the energy consumption of the UUV, an optimal path should be found through a path planning algorithm. Then, UUV tracks the optimal path to reach the destination with the least energy. After UUV obtains the trajectory point of the ideal path, it will track the trajectory point according to its own model and adjust the forward-looking distance through $l_{di} = k \cdot v_i$ where v_i is UUV thrust velocity at time i and l_{di} is the forward-looking distance, and k is the forward-looking distance coefficient. The forward-looking distance l_{di} is constrained by the coefficient k. The larger forward-looking distance means the smoother the tracking trajectory. The smaller the forward-looking distance will make the tracking more accurate, but it will also bring control shocks. If (x_0, y_0) is the next track point to be tracked by UUV, yaw is updated which is given by

$$\theta_{i+1} = \theta_i + \frac{v_i}{L_p} \cdot \frac{2 \cdot L \cdot \sin(\arctan((y_0 - y_i)/(x_0 - x_i)) - \theta_i)}{L_{di}} \cdot t, \tag{2}$$

where L_p is the preview distance of UUV and L is the length of UUV.

3.3. UASN Energy Consumption Model. Energy consumption directly affects the life cycle and cost of the entire localization system, which is often reflected in the sensor nodes communication reception and transmission, perception data processing, and movement adjustment. Meanwhile, the transmission power of underwater acoustic communication is much higher than that of radio wave communication. Consequently, in order to extend the life of UASN and improve the tracking accuracy, we have to reduce the energy consumption of the underwater localization. When UUV enters the UASN, each anchor node starts to send an acoustic signal to UUV. Then, the UUV gives a feedback signal in time. The energy consumption in the entire UASN is the sum of the energy consumed by all anchor nodes which send signals. We apply a commonly used underwater communication energy model to the entire system model. [33, 36] Thereby, the energy consumption of a single anchor can be described as

$$E_t = bE_0 + 4.2 \cdot bT_b Z e^{g(f)d} \cdot 10^{-9.5}, \tag{3}$$

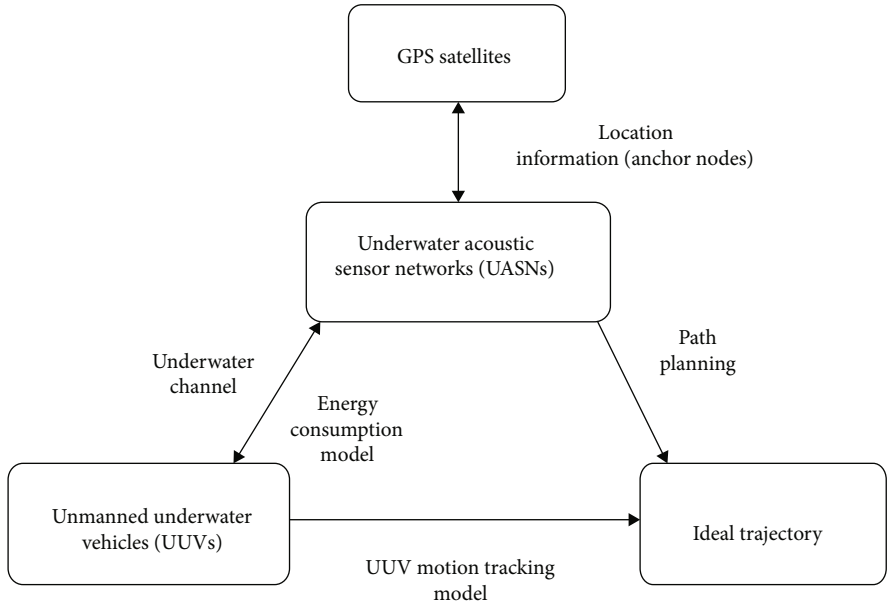


Figure 2: Illustration of the UASN, where the UUV receives the message from anchor nodes to position itself for trajectory correction.

where d denotes the transmission distance from the anchor node to UUV, b denotes the bit length of the data packet, and T_b denotes bit duration. Specifically, E_0 represents the unit energy consumption for processing 1 bit message. Z represents the depth and $g(f)$ defines the absorption coefficient in [36]. In addition, f is the center frequency of the transmission channel. Meanwhile, within a localization period, the total energy consumption of the entire network can be described as $E_{\text{all}}=N\cdot E_t$ where N is the number of anchor nodes that send acoustic signals in this localization period.

3.4. Underwater Channel and Signal Receiving and Transmitting Model. Since the underwater isovelocity assumption does not hold in many real-world scenarios, we adopt a new nonisovelocity geometry-based underwater acoustic channel signal transmission model. Underwater acoustic speed changes with the depth. [37–39] Consequently, the geometry-based stochastic underwater acoustic (UWA) channel modeling method has to consider the non-uniform velocity characteristics generated by ocean layers with different sound velocity characteristics and acoustic signal sent from anchor node reach the target through multiple reflections on the sea surface and bottom. [11]

In this paper, we have expanded the geometric model in [40] regarding the propagation conditions of non-equal sound velocity and then proceeded from the geometric model, referring to [11], we further simulate the underwater channel model which is bounded by the sea surface and bottom. These natural boundaries can be regarded as reflectors of sound waves, thus taking into account the specular reflections on the sea surface and bottom. Meanwhile, the simulated sound velocity varies piecewise linearly with depth, thus taking into account the refraction between different sound velocity layers. The entire underwater channel model is shown in Figure 3. There are 3 paths for the transmission of acoustic signals from the transmitter to the receiver. The first is the LOS direct path, the second is the downward arriving (DA) path to the target, and the last is the upward arriving (UA) path to reach the target.

In this paper, we assume that sound speed changes piecewise linearly with the depth of the water. The one-dimensional geometric sound velocity model with water depth h is divided into K different equal-width layers, and the width of each-width layer is given by $\Delta z = h/K$. The sound speed profile is modeled as

$$v_k = v_s + k \cdot g \cdot \Delta z, \tag{4}$$

where v_k is the sound velocity in the k layer, v_s is the initial sound velocity, g is the sound velocity gradient, and k =1,2, ⋯, K.

When the acoustic signal passes through different equal-width layers, it will be refracted. According to Snell's law, we can obtain the angle between the propagation path in the j-th layer and the k-th layer, which is given by

$$R_k = \arcsin\left(\frac{v_k}{v_j}\cdot \sin\left(R_j\right)\right), \tag{5}$$

where R_k is the angle between the propagation and each equal-width layer, k =1,2, ⋯, K, and $k \neq j$. The propagation distance of the acoustic signal can be denoted as

$$d_k = \frac{\Delta z}{\cos\left(R_k\right)}, \tag{6}$$

where d_k is the propagation distance in the k-th equal-with layer.

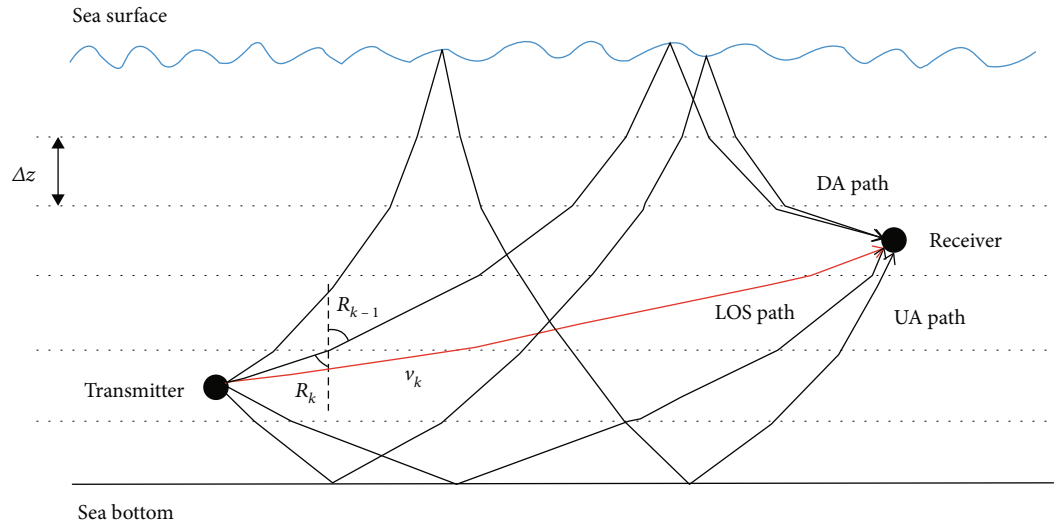


FIGURE 3: Illustration of entire underwater channel model, in which there are 3 paths for the transmission of acoustic signals from the transmitter to the receiver.

According to [41], we can know that the acoustic signal can only be received when it is in the monitoring range of UUV. In order to allow UUV with unknown coordinate to receive the signal transmitted from anchor nodes, according to [42], we equip the signal transmitter with an omnidirectional hydrophone which can send signals every certain angle ω. If the transmitter is in the i-th equal-width layer and the receiver is in the j-th equal-width layer, the horizontal propagation path of the acoustic signal can be described as

$$s_m = \begin{cases} z_{\text{up}} \cdot \tan(R_0) + \sum_{k=i-1}^{j} d_k \cdot \sin(R_k) \\ z_{\text{down}} \cdot \tan(R_0) + \sum_{k=i+1}^{j} d_k \cdot \sin(R_k) \end{cases}, \quad (7)$$

where z_{up} is the distance between the transmitter and the upper surface in the same layer and z_{down} is the distance between the transmitter and the lower surface in the same layer. The vertical propagation path of the acoustic signal z_m can be described as

$$z_m = \begin{cases} z_t - z_{\text{up}} - \Delta z \cdot (i - j - 1) \\ z_t + z_{\text{down}} + \Delta z \cdot (j - i - 1) \end{cases}, \quad (8)$$

where z_t is the depth of the transmitter. The distance l between the signal and the transmitter can be given by

$$l = \sqrt{(s_m - s_{t-r})^2 + (z_m - z_r)^2}, \quad (9)$$

where s_{t-r} is the horizontal distance between the transmitter and the receiver and z_r is the depth of the receiver. The entire design flow is summarized as follows. First of all, we divide the water depth h into k equal-width layers and the sound velocity of each layer is v_k. Then, the signal transmitter transmits signals every certain angle ω. In this underwater channel, the acoustic signal is refracted between different equal-width layers and reflected on the sea bottom and sea surface. Finally, the distance l to the receiver is judged according to the propagation distance of the acoustic signal. The maximum monitoring range of the receiver is $l_{\max}$. If $l \le l_{\max}$, the receiver can receive the signal; if $l > l_{\max}$, the receiver can not receive the signal.

According to [42], the time-variant channel impulse response (TVCIR) of the underwater channel model can be denoted as

$$h(t) = h_{\text{LOS}}(t) + h_{\text{DA}}(t) + h_{\text{UA}}(t), \quad (10)$$

where $h_{\text{LOS}}(t)$ describes the LOS component; $h_{\text{DA}}(t)$ describes the DA component, and $h_{\text{UA}}(t)$ describes the UA component. The propagation loss coefficient of the signal in the underwater acoustic channel can be simplified as [42].

$$c_0 = (c_{\text{up}})^{n_1} \cdot (c_{\text{down}})^{n_2} \cdot \frac{k}{d}, \quad (11)$$

where c_{up} is attenuation coefficient of the sea surface; c_{down} is attenuation coefficient of the sea bottom; n_1 is the number of reflections on the sea surface; n_2 is the number of reflections on the sea bottom; k is the attenuation constant of the underwater acoustic channel and d is the acoustic signal propagation distance. In the positions of the transmitter and receiver change, there will be no LOS path.

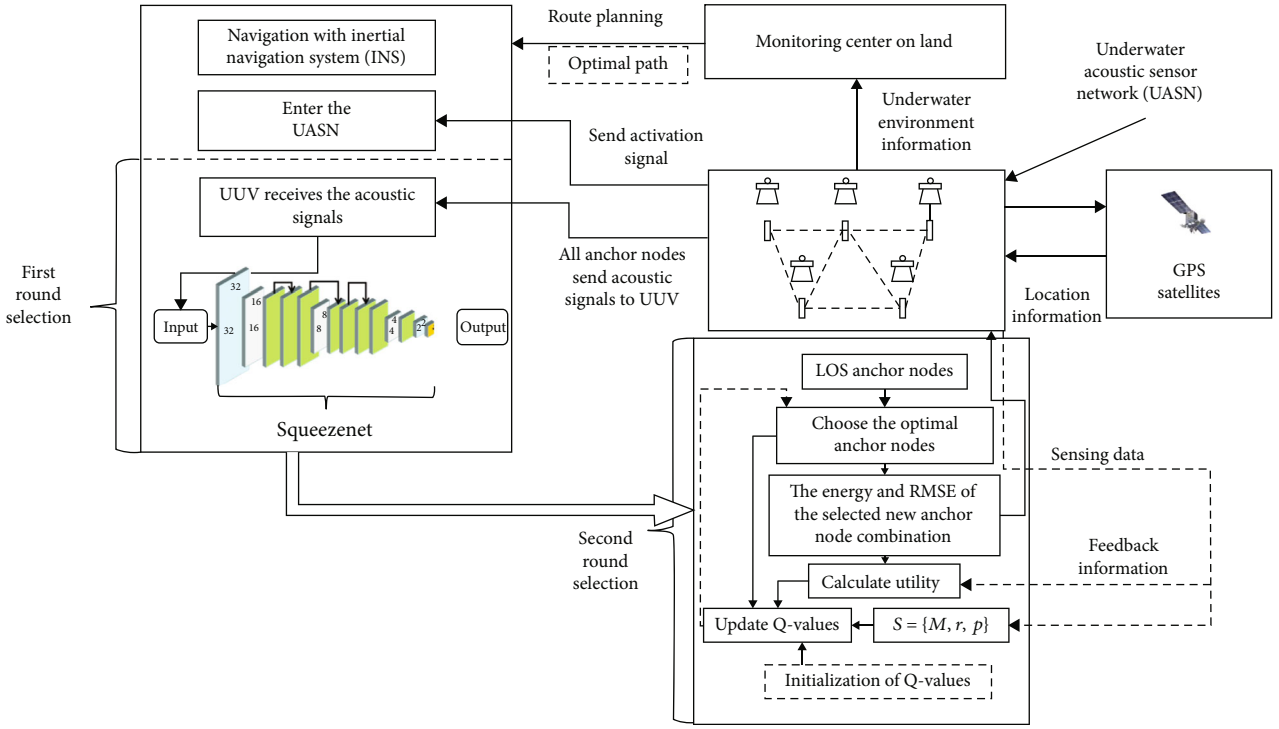


Figure 4: Illustration of underwater trajectory correction process.

4. Reinforcement Learning and Lightweight Underwater NLOS Signal Recognition Neural Network Based Energy-Efficient Mobile Underwater Localization Algorithm

We propose a reinforcement learning and neural network (SqueezeNet [43]) based energy-efficient mobile underwater localization scheme in UASN that optimizes the anchor node selection policy and selects the anchor nodes in two rounds according to the signals transmitted from anchor nodes to UUV. Then, the optimal anchor nodes are used to locate UUV so as to balance the localization accuracy and the energy consumption of UASN. To be specific, in order to minimize the energy consumption of UUV from departure to destination, the path has to be shortest. However, due to the complex underwater environment, there are many underwater obstacles between the departure and the destination, which makes it impossible for UUV to reach the destination directly in a straight line. At this time, all anchor nodes in the UASN are required to conduct a rough monitoring of the underwater terrain, which rasterizes the entire underwater map. After getting the entire two-dimensional matrix, it is sent to UUV. Then, the ideal path is generated through path planning algorithm. Finally, UUV tracks the ideal path through pure pursuit algorithm. When UUV enters the UASN, it will send signals to activate all anchor nodes. After activating all anchor nodes, clock synchronization is performed between each anchor node and the position of each anchor node is obtained through GPS. Then, all anchor nodes send acoustic signals to UUV at the same time. When the UUV receives the signals sent by all anchor nodes, it uses SqueezeNet to classify the received LOS and NLOS signals and then selects the LOS anchor nodes and discards the NLOS anchor nodes. Thus, the first round of anchor node selection through SqueezeNet is completed. After first round selection, the optimal anchor nodes are selected through reinforcement learning from the obtained LOS anchor nodes for second round selection. In second round selection, the current decision of UUV is only dependent on the latest state, so the anchor nodes selection process can be formulated as a Markov decision process (MDP), where the RL technique can be applied to determine the optimal transmission policy based on the observed state via trail-and-error. More specifically, at time slot k, the target obtains the current state x^k which includes the previous selected anchor nodes, the previous localization error, and the previous energy consumption. Meanwhile, the anchor nodes are selected according to the current state and Q-function which is updated according to the Bellman equation iteratively. [44] When the optimal anchor nodes are obtained, UUV will locate itself by the least square method. Then, the motion state of UUV is adjusted by purepursuit algorithm according to its own coordinates, which makes it close to the ideal path. The whole process is shown in Figure 4.

4.1. Signal Classification Neural Network. When UUV enters the UASN, it sends out a command signal to activate all anchor nodes. Due to the influence of underwater terrain and ships, UUV receives LOS signals and NLOS signals. However, NLOS signal will greatly affect the localization accuracy, which affects the UUV trajectory correction. In

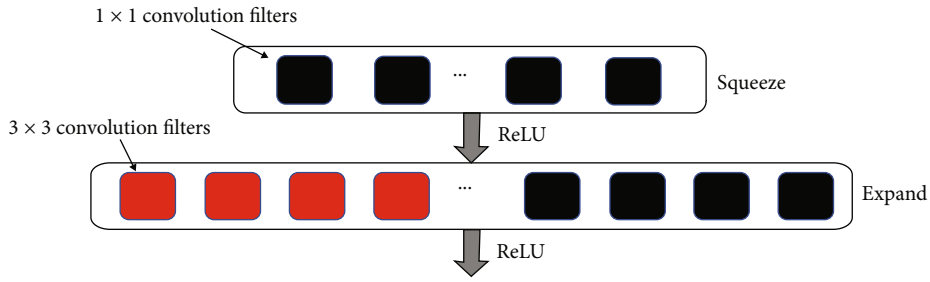


Figure 5: Fire module.

this paper, SqueezeNet is applied to identify the received acoustic signal, so as to make full use of LOS signal and eliminate NLOS signal.

In recent years, many researches about deep convolutional neural networks have focused on improving the classification accuracy. It is not difficult to find multiple CNNs that can reach a certain level of accuracy. With the same level of accuracy, a smaller CNN model can facilitate us with three advantages. First, smaller CNNs require less cross-server communication when conducting distributed training and can receive training faster because of less communication, which have great advantages for the classification of underwater LOS/NLOS signals. Second, smaller CNNs can simplify the process of exporting new models from the cloud to UUV which makes it easier for UUV to import new training models, which is very important for complex underwater environments. Last but not least, smaller CNN model can be deployed on hardware with limited memory. When the CNN model is too large, it cannot be deployed on UUV. Considering all these advantages, we choose SqueezeNet to classify underwater LOS/NLOS signals whose model size is only 0.5 MB. [29]

SqueezeNet is composed of several Fire module combined with convolution layers, downsampling layers, and fully connected layers; the developers of which mainly adopted three strategies to obtain fewer parameters:

(1) The first strategy for designing SqueezeNet is to replace 3×3 filters with 1×1 filters. Most filters are 1×1, which makes the parameters of the model 9 times less

(2) The second strategy adopted to build the SqueezeNet is to reduce the number of input channels to 3×3 filters

(3) The last strategy adopted is to down sample late in the network in order to ensure that SqueezeNet has fewer parameters, which can obtain a convolutional layer with a large activation map that can lead to higher classification accuracy. For down sampling, strides are set to greater than one in some convolutional and pooling layers

In short, the first two strategies are related to the reduction of the number of parameters in CNN and the last strategy is about maximization accuracy under a limited budget of parameters.

As shown in Figure 5 below, fire module is the most important part of SqueezeNet which consists of squeeze layer and expand layer. The squeeze layer is composed of a set of continuous 1×1 and 3×3 convolution filters. In fire module, the number of 1×1 convolution filters in the squeeze layer is recorded as $s_{1\times1}$, the number of 1×1 convolution filters in the expand layer is recorded as $e_{1\times1}$, and the number of 3×3 convolution filters in the expand layer is recorded as $e_{3\times3}$. Meanwhile, in the fire module, $s_{1\times1} < e_{1\times1}+s_{3\times3}$, which helps to keep the number of input channels limited to 3×3 filters, as discussed for the second strategy adopted for SqueezeNet.

Figure 6 represents the SqueezeNet structure with simple bypass used in this paper. It starts with a convolution layer, which is named conv1 in Figure 6. After conv1, there are 8 fire modules, where the number of filters in each fire module is gradually increasing. After fire module 4 and fire module 8, max pooling is performed. Finally, it ended with a convolution layer after which max pooling is performed. Meanwhile, the input is a matrix signal of dimension 64×64. Initial learning is 0.001 and input batch size for training is 32. In addition, the optimizer of SqueezeNet is "Adam" and the output is the precision and the classified signal. Moreover, the loss function Loss is the cross-entropy loss function, and the expression is

$$\text{Loss} = -\frac{1}{S}\sum_{j}\sum_{d=1}^{G} y_{jd} \log\left(p_{jd}\right), \tag{12}$$

where S is the number of the samples and G is the number of the label categories. y_{jd} is a symbolic function. When the true category of sample j is equal to d, $y_{jd}=1$; otherwise, $y_{jd}=0$. Moreover, p_{jd} is the probability value for each prediction result by softmax, where we choose 0.5 as the threshold due to the binary classification. The architecture parameters of the SqueezeNet are shown in Table 1.

4.2. Path Planning Algorithm. According to [45, 46], path planning algorithms can be divided into grid map method, roadmap method, and artificial potential field method. All anchor nodes in the UASN conduct a rough monitoring of the underwater terrain and rasterize the entire underwater map to form a two-dimensional grid. Meanwhile, according to whether there are obstacles in the grid, we can divide each grid into two states, where the barrier-free grid is called the

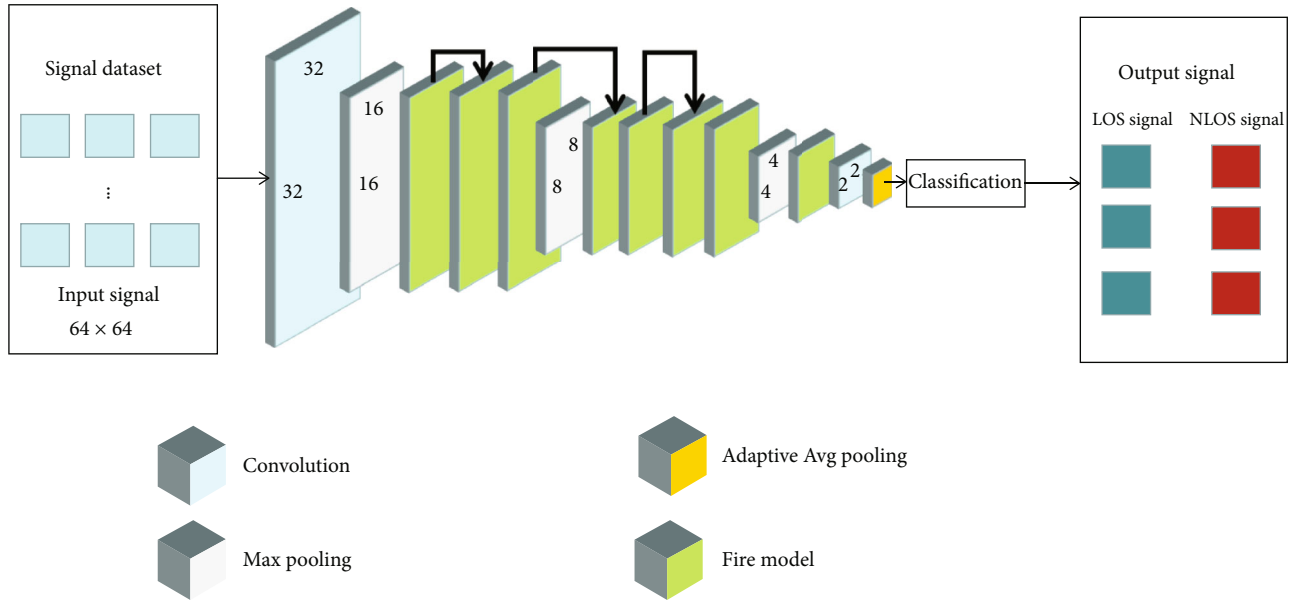


Figure 6: The structure of SqueezeNet.

Table 1: The architecture parameters of the SqueezeNet.

Layer		Filter size/number	Output size
Input signal			64 × 64/3
Convolution		3 × 3/96	32 × 32/96
Maxpool		2 × 2/96	16 × 16/96
Fire2	squeeze2	1 × 1/16	16 × 16/128
	expand2	1 × 1/64 3 × 3/64	
Fire3	squeeze3	1 × 1/16	16 × 16/128
	expand3	1 × 1/64 3 × 3/64	
Fire4	squeeze4	1 × 1/32	16 × 16/256
	expand4	1 × 1/128 3 × 3/128	
Maxpool		3 × 3/256	8 × 8/256
Fire5	squeeze5	1 × 1/32	8 × 8/256
	expand5	1 × 1/128 3 × 3/128	
Fire6	squeeze6	1 × 1/48	8 × 8/384
	expand6	1 × 1/192 3 × 3/192	
Fire7	squeeze7	1 × 1/48	8 × 8/384
	expand7	1 × 1/192 3 × 3/192	
Fire8	squeeze8	1 × 1/64	8 × 8/512
	expand8	1 × 1/256 3 × 3/256	
Maxpool		3 × 3/512	4 × 4/256
Fire9	squeeze9	1 × 1/64	4 × 4/512
	expand9	1 × 1/256 3 × 3/256	

free grid; the obstacle grid is called the obstacle grid. The UUV path planning problem is actually to find the shortest path from the starting grid to the target grid by bypassing the obstacle grid. Since the A∗ algorithm can handle fixed threats and sudden threats and can find the optimal path in a short time [45], it can achieve online real-rime path planning. Meanwhile, it is an efficient heuristic searching algorithm, which can improve the search efficiency and ensure the optimal cost of the voyage. At the same time, the simulated annealing (SA) algorithm is a general probability algorithm which is also widely applied in path optimization. In order to find the optimal path quickly and accurately, we compare the A∗ algorithm and the SA algorithm under the 200 × 200 grid map. Meanwhile, in order to compare the robustness of the algorithm, we compose different underwater environments numbered $E_1 \sim E_5$ by changing the position of obstacles in the grid map. The simulation result is shown in the following Table 2. According to the simulation results, the paths obtained by the A∗ algorithm are better than the SA algorithm in different underwater environments. Consequently, we choose the A∗ algorithm to plan the optimal path.

Table 2: The distance of the optimal path of the A∗ algorithm and SA algorithm under different underwater environments $E_1 \sim E_5$.

Algorithm		A∗	SA
Distance (m)	E_1	144.84	178.98
	E_2	98.28	104.14
	E_3	144.84	150.70
	E_4	132.42	138.28
	E_5	146.56	212.42

4.3. RL-Based Mobile Underwater Localization Algorithm. The pseudo-code of RL-based mobile underwater localization algorithm is summarized in Algorithm 1. The number of anchor nodes in UASN is N. After UUV receives the signals transmitted from all anchor nodes, it uses the trained neural network to judge the received signals, which is the

1: Initialize learning rate α, discount rate β, the constant of the utility C and μ, probability constant ε, initial Q-table $Q(s_k, a_k) = 0$ and initial state s_0.
2: **for** k =1, 2, 3$\cdots$ **do**
3: Observe the state s_k=$[M_{k-1}, r_{k-1}, p_{k-1}]$
4: Choose a_k via ε-greedy
5: **for** each selected anchor node a **do**
6: Send U_a,h_a, and t_{ra} and store in M_k
7: **end for**
8: Calculate $v_{\text{ave}(k)}$ via (13)
9: Calculate $\mathbf{u}_k$ via (14)
10: Calculate r_k via (15)
11: Calculate p_k via (3)
12: Evaluate δ_k via (16)
13: $Q(s_k, a_k) \longleftarrow (1-\alpha)Q(s_k, a_k) + \alpha(\delta_k + \beta \cdot \max Q(s_{k+1}, a))$
14: $\Theta'(s_k, a_k, s_{k+1}) \longleftarrow \Theta'(s_k, a_k, s_{k+1}) + 1$
15: $\Theta(s_k, a_k) \longleftarrow \sum_{s_{k+1} \in \Lambda} \Theta'(s_k, a_k, s_{k+1})$
16: Update $R(s_k, a_k)$ via (20)
17: Update $\Xi(s_k, a_k, s_{k+1})$ via (21)
18: **for** n=1,2,3$\cdots$ **do**
19: Randomly select $(\dot{s}_n, \dot{a}_n)$
20: Calculate $R(\dot{s}_n, \dot{a}_n)$ via (20)
21: Obtain $\dot{s}_{n+1}$ via (21)
22: Update Q-function $Q(\dot{s}_n, \dot{a}_n)$ via Bellman equation
23: **end for**
24: **end for**

Algorithm 1: RL-based mobile underwater localization algorithm(RMUL).

first selection in order to obtain LOS anchor nodes N_{los}. When N_{los} are obtained, the target uses Algorithm 1 to select multiple optimal anchor nodes from N_{los} to localize itself. The selected anchor nodes information M_{k-1}, localization error r_{k-1}, and energy consumption p_{k-1} are obtained by the UUV in order to formulate the state s_k, which is given by $S_k = [M_{k-1}, r_{k-1}, p_{k-1}]$. Then referring to the current state, UUV uses trial-and-error to select anchor nodes. UUV needs at least 3 anchor nodes to localize itself. Consequently, the number of selected anchor nodes localization combinations is $\sum_{i=3}^{N_{LOS}} C_{N_{LOS}}^{i}$. To be specific, the index of selected anchor nodes I_k is the N_{los}-bit binary number, where the a -th binary bit takes the value 0 or 1 to indicate whether the anchor node a is selected and the selected anchor node is stored in M. Then according to the selected the anchor nodes information, UUV calculates its own localization and energy consumption and the unselected anchor nodes are not included in the calculation and keep silent in order to reduce energy consumption. Meanwhile, UUV applies the ε-greedy method to select to avoid falling into local optimum. More specifically, the optimal anchor nodes with maximum Q-value are selected with a high probability 1-ε and UUV selects anchor nodes randomly with a small probability ε [47].

After receiving anchor nodes information including anchor node coordinate U, depth h, and reception time t_r, in order to simplify UUV operation when performing tasks, we apply an isogradient depth-dependent acoustic speed profile and the assumption of a straight-line propagation [9], where the acoustic speed decreases linearly with depth according to the formula $v = b - az$, where a is a constant depending on the environment, b indicates the sound speed at the surface, and z denotes the underwater depth. In real scene, since we do not know the underwater channel model accurately, UUV uses pressure sensors to estimate its depth and calculates the average velocity v_{ave} of acoustic signal between itself and anchor node via

$$v_{\text{ave}} = \frac{a(z-h)}{\ln(b-ah) - \ln(b-az)}. \tag{13}$$

Similar to [19], UUV estimates the distance l between itself and anchor node based on signal reception time and average speed v_{ave} obtained above. Then according to the received anchor node coordinates, UUV calculates its own position $\mathbf{u}_k$ which is given by

$$\mathbf{u} = \left(A^T A\right)^{-1} A^T b, \tag{14}$$

where $A = [U_i - U_m]_{1 \le i \le m-1}^T$, $b = [\|U_i\|^2 - \|U_m\|^2 + \|l_m\|^2 - \|l_i\|^2]_{1 \le i \le m-1}^T$, and m is the number of selected anchor nodes. After obtaining $\mathbf{u}_k$, in real life, since we cannot know the real location of UUV. Consequently, UUV estimates the localization error r_k via [19].

$$r = \frac{1}{m} \sum_{i \in m} \left[l_i^2 - \|U_i - \mathbf{u}\|^2 \right]. \tag{15}$$

Then, using r_k and p_k, UUV obtains its utility δ_k which is calculated by

$$\delta_k = C - \mu r_k - \sum_{j \epsilon m} p_{(j)k}, \tag{16}$$

where C and μ are the constants to ensure that r_k and $\sum_{j \epsilon m} p_{(j)k}$ are in a same scale and also determine the weight between the localization error and energy consumption. [47] Moreover, the smaller localization error and energy consumption, the better its utility. At the same time, the Q-function is updated each time slot according to the Bellman equation iteratively [44] with the learning rate α and the discount rate β. In the whole reinforcement learning framework, the Q-function is applied to learn the optimal anchor node selection strategy to find the optimal anchor node, which reduces the localization error and energy consumption and optimizes the utility of the entire UASN.

We use Dyna architecture to reduce the convergence time of the reinforcement learning. More specifically, UUV records each state-action pair based on historically selected actions to generate a virtual environment and accelerates the learning process according to this virtual environment. After real learning, the current state, action, next state, and reward are recorded to obtain each new exploration experience. Then, UUV updates count vector via

$$\Theta'(s_k, a_k, s_{k+1}) = \Theta'(s_k, a_k, s_{k+1}) + 1. \tag{17}$$

From the combination of actions and states that have occurred, a total state-action counter vector $\Theta(s_k, a_k)$ that consists of a vector $\Theta'(s_k, a_k, s_{k+1})$ of all possible next state counts under the current state-action pair has been constructed, which is given by

$$\Theta(s_k, a_k) = \sum_{s_{k+1} \epsilon \Lambda} \Theta'(s_k, a_k, s_{k+1}). \tag{18}$$

After each real experience obtained, the corresponding model rewards denoted by $R'(s_k, a_k, \Theta(s_k, a_k))$ can be recorded by UUV via

$$R'(s_k, a_k, \Theta(s_k, a_k)) \longleftarrow \delta_k(s_k, a_k). \tag{19}$$

Meanwhile, based on $R'(s_k, a_k, \Theta(s_k, a_k))$, the reward function denoted by $R(s_k, a_k))$ can be updated via

$$R(s_k, a_k) \longleftarrow \frac{1}{\Theta(s_k, a_k)} \sum_{n=1}^{\Theta(s_k, a_k)} R'(s_k, a_k, n). \tag{20}$$

Based on $\Theta'(s_k, a_k, s_{k+1})$ and $\Theta(s_k, a_k)$, a transition probability from the current state to the predictive next state can be constructed, which is given by

$$\Xi(s_k, a_k, s_{k+1}) = \frac{\Theta'(s_k, a_k, s_{k+1})}{\Theta(s_k, a_k)}. \tag{21}$$

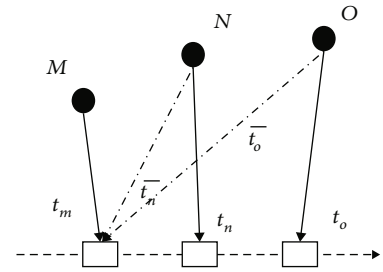


Figure 7: The model of moving target receiving acoustic signals, in which M, N, and O are selected anchor nodes, t_m, t_n, and t_o are real reception time, and $\bar{t}_n$ and $\bar{t}_o$ are real reception time with compensation time.

Table 3: The recognition rate of SqueezeNet for LOS/NLOS signals at different SNR.

Parameter	Value
Area	5000×5000 m^2
Center frequency	20 kHz
Bandwidth	20 kHz
Modulation	4FSK
Forward-looking distance coefficient k	0.7
UUV speed v	5 m/s
Bit length of data packet b	2
Energy of data packet E_0	0.5
Attenuation coefficient of the sea surface c_{up}	0.9
Attenuation coefficient of the sea bottom c_{down}	0.5
Learning rate α	0.85
Discount rate β	0.95
Constant C	20

Table 4: The recognition rate of SqueezeNet for LOS/NLOS signals at different SNR.

SNR(dB)	−6	−4	−2	0	2	4
Precision	0.672	0.776	0.906	0.964	0.984	0.996

In model learning, the UUV randomly selects an action-state pair from the experiences recorded in the virtual environment at each time slot. According to (20) and (21), the UUV predicts the next state and gets a reward. Then, the Q-function is updated based on the state-action pair, next state, and the model reward according to the Bellman equation, which iterates multiple times. [47] Thereby, hypothetical experience in the model is obtained to speed up the convergence. In addition, in order to reflect the role of virtual experience, we do not add Dyna structure in this algorithm at the same time, which is called RMUL-Q.

To sum up, during a trajectory correction cycle, UUV first filters out the LOS anchor nodes through SqueezeNet and then selects the optimal anchor nodes through RMUL-

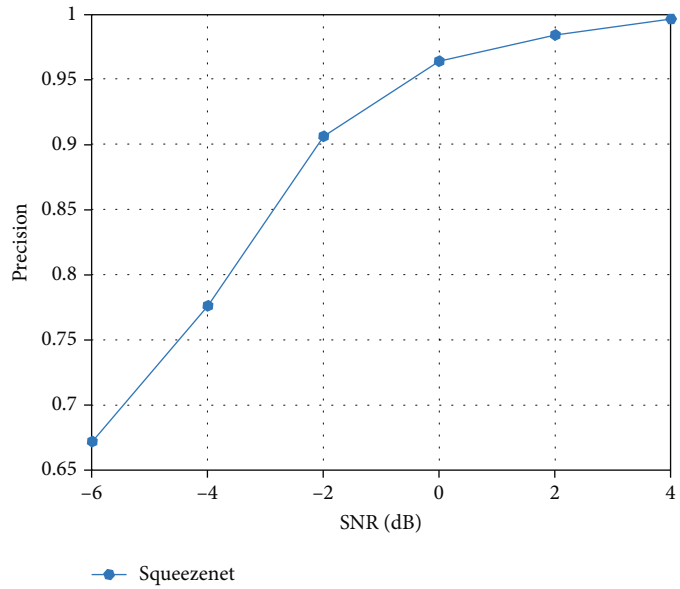


Figure 8: The performance of SqueezeNet for LOS/NLOS signals at different SNR.

Dyna-Q. In the remaining of the trajectory correction period, no acoustic signal is sent from non-optimal anchor nodes in order to save energy. Meanwhile, during this trajectory correction period, UUV continuously locates itself by receiving the signals sent by the optimal anchor nodes. After obtaining its own calculated location, UUV approaches the ideal trajectory according to the pure pursuit algorithm, so as to achieve trajectory correction. When this trajectory correction cycle ends, the next trajectory correction cycle is performed immediately. Then, the above operations are repeated.

5. CRLB

As a good indicator for the uncertainty in the parameter estimation, the Cramer-Rao Lower Bound (CRLB) expresses a lower bound on the variance of any unbiased estimator of a deterministic parameter. In order to examine the performance limit of the localization problem, we derive a CRLB without considering the target movement first. Then, we derive a CRLB by considering the movement of the target and optimal anchor nodes.

Theorem 1. *The CRLB for the localization without considering the target movement is given by*

$$CRLB(\vartheta) = Tr\left[\left(\widehat{F}(\vartheta)\right)^{-1}\right] = \frac{\widehat{F}_{\vartheta\,1,1} + \widehat{F}_{\vartheta\,2,2}}{\widehat{F}_{\vartheta\,1,1}\widehat{F}_{\vartheta\,2,2} - \widehat{F}_{\vartheta\,1,2}\widehat{F}_{\vartheta\,2,1}}. \quad (22)$$

The Fisher information matrix (FIM) [48] for $\widehat{F}(\vartheta)$ is given by

$$\widehat{F}(\vartheta) = \begin{bmatrix} \widehat{F}_{\vartheta\,1,1} & \widehat{F}_{\vartheta\,1,2} \\ \widehat{F}_{\vartheta\,2,1} & \widehat{F}_{\vartheta\,2,2} \end{bmatrix}, \quad (23)$$

in the formula

$$\widehat{F}_{\vartheta\,1,1} = \sum_{i=1}^{M} \frac{(x - x_i)^2}{\sigma_i^2 (\vartheta - U_i)^2 \bar{V}_i^2}, \quad (24)$$

$$\widehat{F}_{\vartheta\,1,2} = \widehat{F}_{\vartheta\,2,1} = \sum_{i=1}^{M} \frac{(x - x_i)(y - y_i)}{\sigma_i^2 (\vartheta - U_i)^2 \bar{V}_i^2}, \quad (25)$$

$$\widehat{F}_{\vartheta\,2,2} = \sum_{i=1}^{M} \frac{(y - y_i)^2}{\sigma_i^2 (\vartheta - U_i)^2 \bar{V}_i^2}. \quad (26)$$

Proof. Given a vector $\vartheta = [x, y]^T$, the M measurements on the reception time are as follows

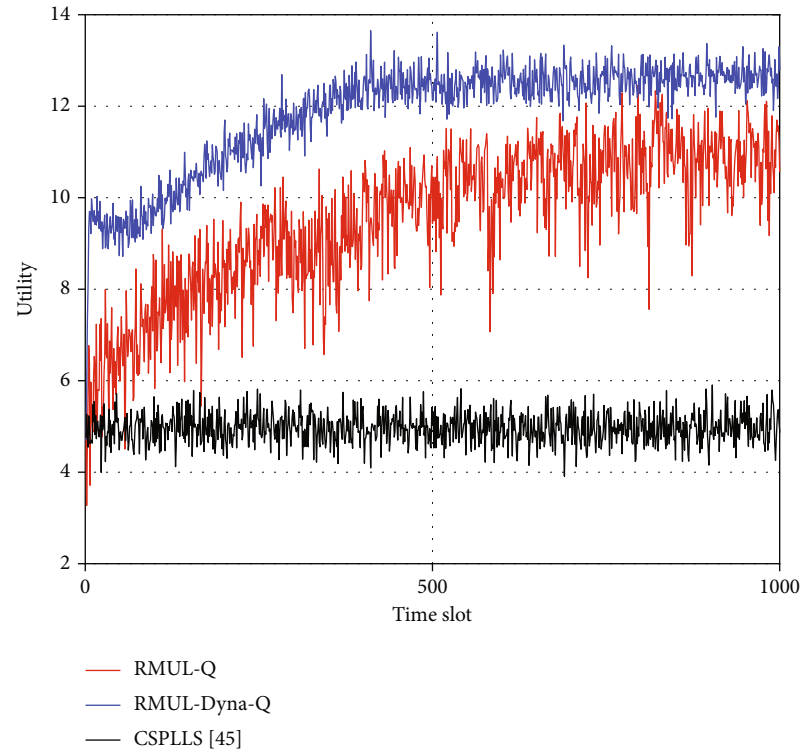
$$t_i = \frac{\|\vartheta - U_i\|}{\bar{V}_i} + n_i, i = 1, \cdots, M, \quad (27)$$

where $n_i \sim N(0, \sigma_i^2)$ is the measurement error of the reception time between the target and anchor node i. Consequently, the log-likelihood function denoted as $\bar{L}(\vartheta)$ is given by

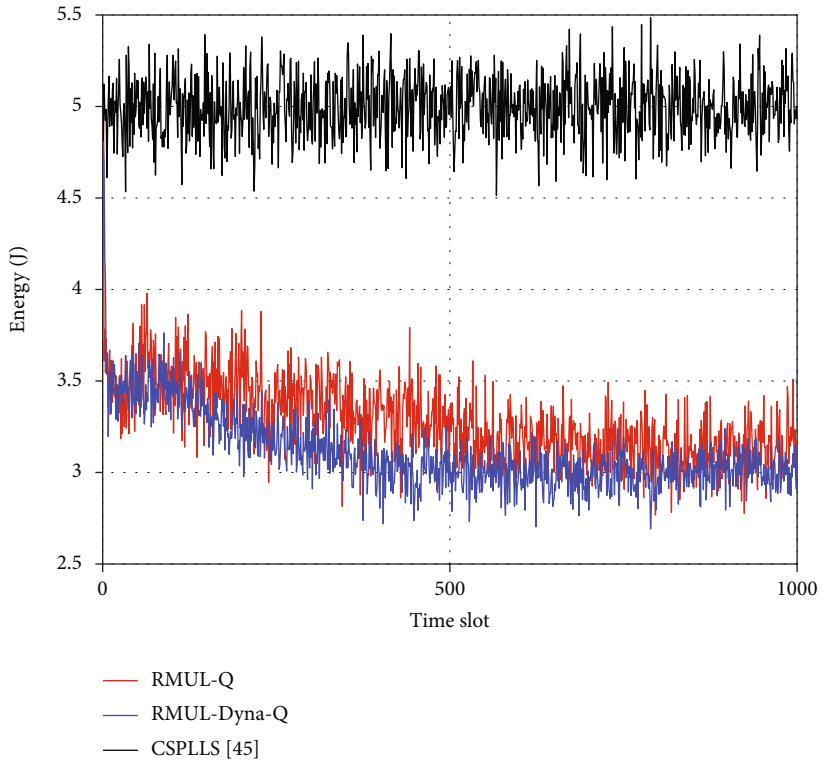
$$\bar{L}(\vartheta) = -\frac{1}{2} \ln\left(2\pi \prod_{i=1}^{M} \sigma_i^2\right) - \sum_{i=1}^{M} \frac{1}{2\sigma_i^2} \left(\frac{\|\vartheta - U_i\|}{\bar{V}_i} - t_i\right)^2. \quad (28)$$

Then, the FIM is given by

$$\widehat{F}(\vartheta) = E\left[\frac{\partial^2 \bar{L}(\vartheta)}{\partial \vartheta^2}\right]. \quad (29)$$

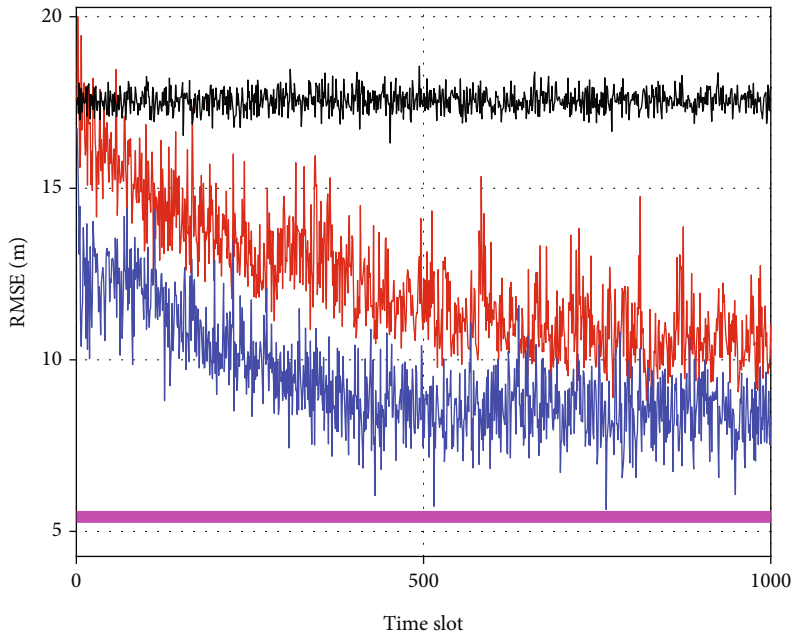


(a) Utility



(b) Energy consumption

Figure 9: Continued.



(c) RMSE

Figure 9: Performance in the underwater environment.

Based on (28) and (29), the FIM for $\widehat{F}(\vartheta)$ is derived as (23); thus, CRLB for the localization without considering the target movement is derived. □

Theorem 2. *The CRLB for considering the movement of the target and optimal anchor nodes is given by*

$$CRLB(\zeta) = Tr\left[\left(\widehat{F}(\zeta)\right)^{-1}\right] = \frac{\widehat{F_{\zeta}}_{1,1} + \widehat{F_{\zeta}}_{2,2}}{\widehat{F_{\zeta}}_{1,1}\widehat{F_{\zeta}}_{2,2} - \widehat{F_{\zeta}}_{1,2}\widehat{F_{\zeta}}_{2,1}}. \quad (30)$$

The Fisher information matrix (FIM) [48] for $\widehat{F}(\vartheta)$ is given by

$$\widehat{F}(\zeta) = \begin{bmatrix} \widehat{F_{\zeta}}_{1,1} & \widehat{F_{\zeta}}_{1,2} \\ \widehat{F_{\zeta}}_{2,1} & \widehat{F_{\zeta}}_{2,2} \end{bmatrix}, \quad (31)$$

in the formula

$$\widehat{F_{\zeta}}_{1,1} = \sum_{i=1}^{N} \frac{(x - x_i)^2}{\left(\sigma_i^2 + \sigma_{si}^2\right)(\zeta - U_i)^2 \bar{V}_i^2}, \quad (32)$$

$$\widehat{F_{\zeta}}_{1,2} = \widehat{F_{\zeta}}_{2,1} = \sum_{i=1}^{N} \frac{(x - x_i)(y - y_i)}{\left(\sigma_i^2 + \sigma_{si}^2\right)(\zeta - U_i)^2 \bar{V}_i^2}, \quad (33)$$

$$\widehat{F_{\zeta}}_{2,2} = \sum_{i=1}^{N} \frac{(y - y_i)^2}{\left(\sigma_i^2 + \sigma_{si}^2\right)(\zeta - U_i)^2 \bar{V}_i^2}, \quad (34)$$

where N is the number of selected anchor nodes.

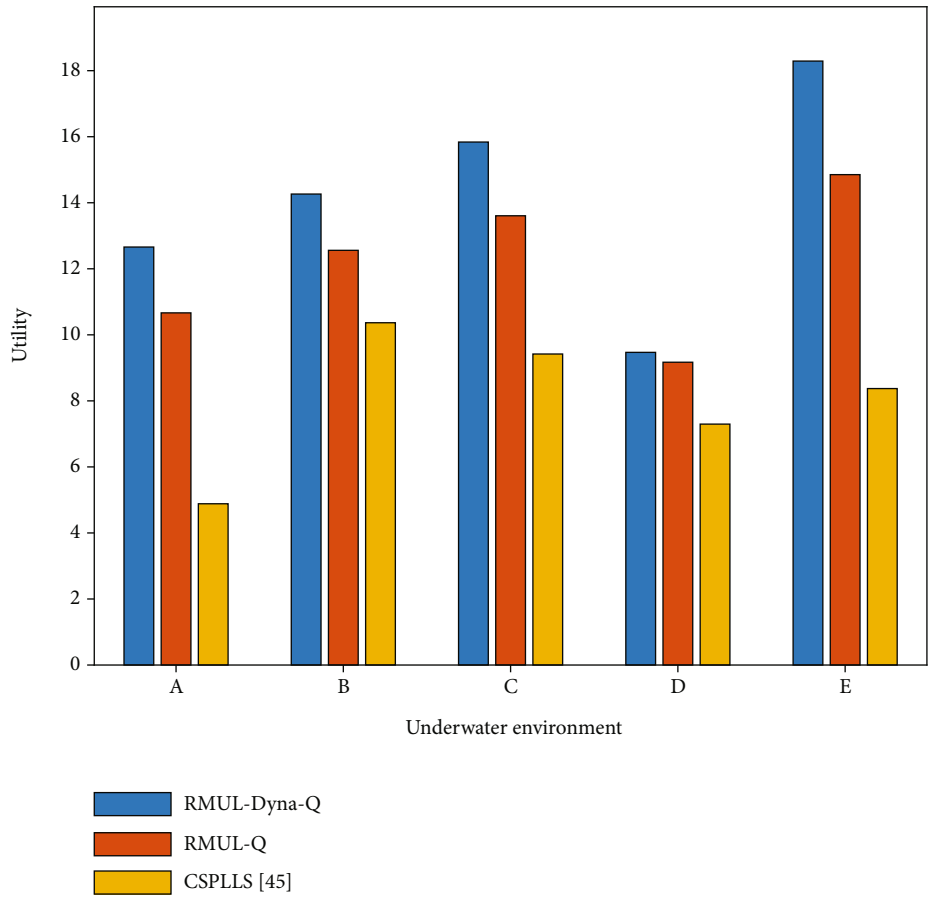
Proof. Given a vector $\zeta = [x, y]^T$, the N measurements on the reception time are as follows

$$\bar{t}_i = \frac{\|\vartheta - U_i\|}{\bar{V}_i} + n_i + n_{si}, i = 1, \cdots, N, \quad (35)$$

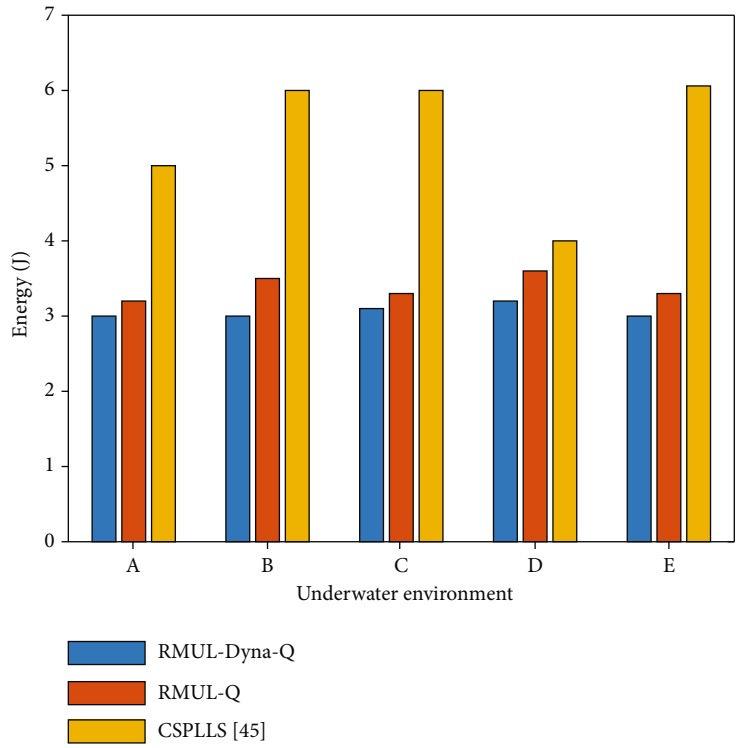
where $n_i \sim N(0, \sigma_i^2)$ is the measurement error of the reception time between the target and anchor node i when the target is stationary and $n_{si} \sim N(0, \sigma_{si}^2)$ is the compensation error of the reception time between the target and anchor node i when the target is in motion as shown in Figure 7. Since n_i and n_{si} are independent normal distributions, their sum is $n_{ti} \sim N(0, \sigma_i^2 + \sigma_{si}^2)$.

Based on (28), (29), and (35), the FIM for $\widehat{F}(\zeta)$ is derived as (31); thus, CRLB for the localization by considering the movement of the target and optimal anchor nodes is derived. □

Remark 3. The UUV applies the SqueezeNet and the RL-based mobile underwater localization algorithm to optimal the anchor nodes selection policy without knowing the underwater acoustic channel in dynamic localization process. If the UUV is stationary in the underwater environment, the CRLB only considering the impact of reception

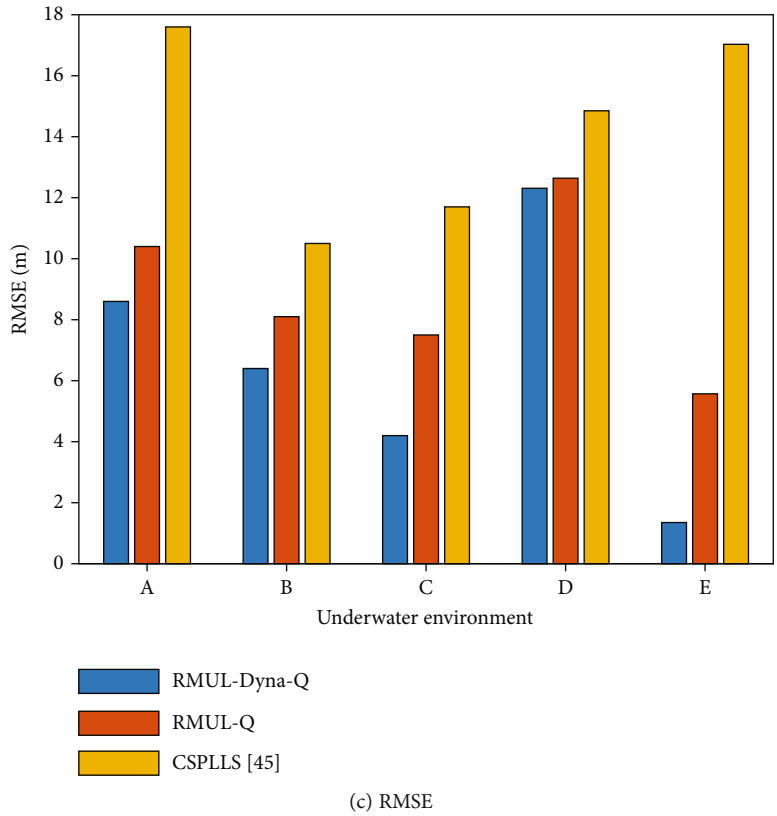


(a) Utility



(b) Energy consumption

Figure 10: Continued.



(c) RMSE

Figure 10: Performance in the different underwater environment.

time measurement error is derived as (22). Consequently, the CRLB can be obtained by substituting the coordinates (x, y) of the UUV, the coordinates (x_i, y_i) of the anchor node i, the underwater sound speed $\bar{V}_i$, and the variance of the measurement error σ_i^2 into the formula (22)–(26). Moreover, if the UUV performs a task, the UUV movement will affect the reception time measurement error. In this case, we derive the CRLB as shown in (30). Consequently, the CRLB can be obtained by substituting the coordinates (x, y) of the UUV, the coordinates (x_i, y_i) of the anchor node i, the underwater sound speed $\bar{V}_i$, the variance of the measurement error σ_i^2, and the variance of the compensation error σ_{si}^2 into the formula (30)-(34).

6. Simulation Results

In order to evaluate the performance of the entire trajectory correction algorithm, we have performed multiple simulations on MATLAB. The entire range of UUV motion is 5000×5000 m^2, in which 20 fixed anchor nodes are randomly located at an area of 1000×1000 m^2 within the depth of 500 m. In these simulations, in order to improve the authenticity of the simulation, we choose the underwater channel designed in Chapter 3 as the underwater channel between target and anchor node. The center frequency and the bandwidth of the underwater acoustic signal are set as 20 kHz. The transmission range of the UUV and anchor nodes are 1000 m and the modulation and the communication rate are 4FSK and 2 kbps, respectively. In pure pursuit algorithm, the relevant parameters are as follows, where the forward-looking distance coefficient is k =0.7 and the velocity of UUV is v =5 m/s; in underwater energy consumption, refer to [33], the relevant parameters are as follows, where bit length of data packet is b =2 and unit energy of data packet is E_0 =0.5; in underwater channel, the relevant parameters are as follows, where attenuation coefficient of the sea surface is c_{up} =0.9 and attenuation coefficient of the sea bottom is c_{down} =0.5; in RL-based mobile underwater localization algorithm, the relevant parameters are as follows, where learning rate α is 0.85, discount rate β is 0.95, and the constant C is 20. The Communication Signal Propagation Loss Localization Scheme (CSPLLS) proposed in [5] and RMUL-Q are evaluated as the benchmarks in simulations. Meanwhile, the CRLB is taken into comparison as a baseline in localization accuracy. More specifically, the parameter table is shown in Table 3.

In the first round of anchor node selection, we apply SqueezeNet to identify the received signals. Simulation shows the recognition rate of SqueezeNet for LOS/NLOS signals at different signal-to-noise (SNR) ratios. The performance of SqueezeNet is counted in Table 4 and is shown in Figure 8.

In the second round of anchor node selection, simulation results of the performance of CSPLLS, RMUL-Q, and RMUL-Dyna-Q schemes versus 1000 time slots are plotted in Figure 9. As shown in Figure 9, the proposed RMUL-

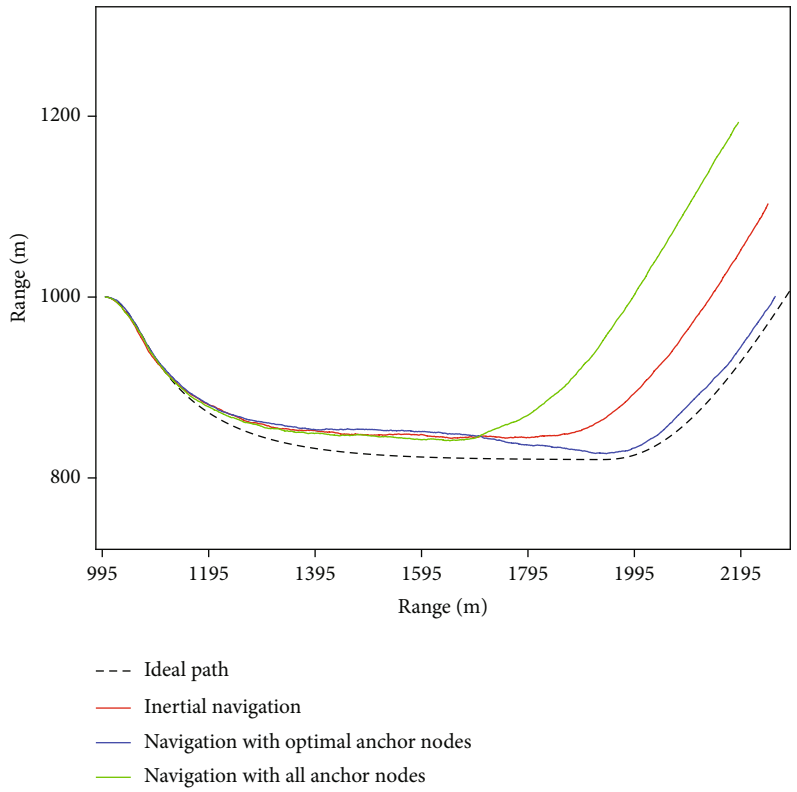


Figure 11: Navigation performance.

DynaQ and RMUL-Q schemes decrease the RMSE and energy consumption and increase utility in 1000 time slots. However, the benchmark basically keeps RMSE, energy consumption, and utility within a stable range. To be specific, the RMUL-Q scheme reduces the RMSE from 17.5 m to 10.8 m and decreases the energy consumption from 5.0 J to 3.2 J in 1000 time slots. At the same time, the RMUL-Dyna-Q scheme reduces the RMSE from 16.7 m to 8.8 m and decreases the energy consumption from 5.0 J to 3.0 J in 500 time slots. From Figure 9, we can infer that the performance of the RMUL-Dyna-Q outperforms that of the benchmarks. More specifically, compared with RMUL-Q and CSPLLS, the RMUL-Dyna-Q has the lowest RMSE, lowest energy consumption, and highest utility. As shown in Figure 9(a), the RMUL-Dyna-Q achieves 50.2% and 15.0% higher utility compared with CSPLLS and RMUL-Q relatively. Meanwhile, as shown in Figure 9(b), the RMUL-Dyna-Q achieves 40.0% and 6.2% lower energy consumption compared with CSPLLS and RMUL-Q relatively. As shown in Figure 9(c), the RMUL-DynaQ achieves 49.7% and 18.5% lower RMSE compared with CSPLLS and RMUL-Q relatively. Moreover, RMUL-Dyna-Q is closer to CRLB compared with CSPLLS and RMUL-Q.

When UUV performs underwater missions, it often needs to experience different underwater environments. Meanwhile, when the underwater environment is different, the position of anchor nodes, the topology of UASN, and the number of NLOS anchor nodes in the UASN will change. In order to obtain simulation results in different underwater environments, we have randomly changed the position of underwater obstacles, the starting point, and destination of the UUV. Correspondingly, these underwater environments are called A, B, C, D, and E. We then evaluate the performance of the RMUL-Dyna-Q, RMUL-Q, and CSPLLS in different underwater environments. As shown in Figure 10, the RMUL-Dyna-Q scheme has the lowest RMSE, lowest energy consumption, and highest utility from 1-1000 time slots in different underwater environments. As can be seen from Figure 10, by simulating in different underwater environments, we can conclude that RMUL-Dyna-Q can find the optimal anchor nodes in a short time in different underwater environments.

As shown in Figure 11, after accurately positioning through optimal anchor nodes the UUV, the UUV can correct its trajectory through the pure pursuit algorithm mentioned in Chapter 3, which achieves close to the ideal path to reach the destination. However, the trajectory of UUV after positioning through all anchor nodes is more deviated from the ideal path than the trajectory of only INS. The reason for this phenomenon is that there are many NLOS transmissions of signals because of the underwater terrain environment, which greatly affects the positioning accuracy and the trajectory of the UUV. Consequently, it is necessary to make multiple selections of anchor nodes.

7. Conclusion

In this paper, we have proposed an UUV underwater trajectory correction scheme based on reinforcement learning (RL) and neural network techniques to address the problems

of the existing methods and reduce energy consumption of the UASN. Meanwhile, we designed a nonisovelocity geometry-based underwater acoustic channel signal transmission model and signal receiving and transmitting model. We provided the CRLB of the proposed scheme. Simulation results showed that the proposed scheme outperforms the benchmarks in localization accuracy and energy consumption in different underwater environments. For instance, compared with CSPLLS and RMUL-Q, the RMUL-Dyna-Q achieves 39.0% and 10.5% higher utility, 40.0% and 6.3% lower energy consumption, and 51.1% and 17.3% lower RMSE, respectively.

As a result, we can come to the conclusion that the proposed method enables UUVs to achieve trajectory correction so as to accurately arrive at the destination to perform tasks and save energy in complex underwater environments. However, there are still some shortcomings in the proposed method, such as low recognition rate under low SNR and slow convergence speed of reinforcement learning. In the future, the proposed method will be extended to the more complex underwater acoustic communication environment. In addition to this, we will validate our method in underwater experiments. Meanwhile, how to further reduce the convergence time is also our future work.

Data Availability

The data used to support the findings of this study were supplied by Ruiheng Liao under license and so cannot be made freely available. Request for access to these data should be made to Ruiheng Liao (23320201154000@stu.xmu.edu.cn)

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Science and Technology on Underwater Information and Control Laboratory (2021-JCJQ-LB-030-10) and supported by National Natural Science Foundation of China (62071400,61871336)

References

[1] K. Wang, H. Gao, X. Xu, J. Jiang, and D. Yue, "An energy-efficient reliable data transmission scheme for complex environmental monitoring in underwater acoustic sensor networks," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4051–4062, 2016.

[2] E. Felemban, F. K. Shaikh, U. M. Qureshi, A. A. Sheikh, and S. B. Qaisar, "Underwater sensor network applications: a comprehensive survey," *International Journal of Distributed Sensor Networks*, vol. 11, no. 11, Article ID 896832, 2015.

[3] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, 2005.

[4] X. Cheng, H. Shu, Q. Liang, and D. H. Du, "Silent positioning in underwater acoustic sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1756–1766, 2008.

[5] G. Qiao, C. Zhao, F. Zhou, and N. Ahmed, "Distributed localization based on signal propagation loss for underwater sensor networks," *IEEE Access*, vol. 7, pp. 112985–112995, 2019.

[6] S. Park, "An efficient transmission scheme for underwater sensor networks," in *OCEANS 2009*, pp. 1–3, Europe, 2009.

[7] Y. Chen, X. Jin, and X. Xu, "Mobile data collection paths for node cooperative underwater acoustic sensor networks," in *OCEANS 2016*, pp. 1–5, Shanghai, 2016.

[8] E. Cheng, L. Wu, F. Yuan, C. Gao, and J. Yi, "Node selection algorithm for underwater acoustic sensor network based on particle swarm optimization," *IEEE Access*, vol. 7, pp. 164429–164443, 2019.

[9] H. Ramezani, H. Jamali-Rad, and G. Leus, "Target localization and tracking for an isogradient sound speed profile," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1434–1446, 2013.

[10] J.-H. Cui, J. Kong, M. Gerla, and S. Zhou, "The challenges of building mobile underwater wireless networks for aquatic applications," *IEEE Network*, vol. 20, pp. 12–18, 2006.

[11] M. Naderi, A. G. Zaji, and M. Pätzold, "A nonisovelocity geometry-based underwater acoustic channel model," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 2864–2879, 2018.

[12] Y. Li, K. Cai, Y. Zhang, Z. Tang, and T. Jiang, "Localization and tracking for AUVs in marine information networks: research directions, recent advances, and challenges," *IEEE Network*, vol. 33, no. 6, pp. 78–85, 2019.

[13] D. Zhang, I. N'Doye, T. Ballal, T. Y. Al-Naffouri, M.-S. Alouini, and T.-M. Laleg-Kirati, "Localization and tracking control using hybrid acoustic–optical communication for autonomous underwater vehicles," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10048–10060, 2020.

[14] D. Li, J. Xu, H. He, and M. Wu, "An underwater integrated navigation algorithm to deal with DVL malfunctions based on deep learning," *IEEE Access*, vol. 9, pp. 82010–82020, 2021.

[15] A. Tendeng, S. Guo, R. An, and C. Li, "D∗ lite-based navigation algorithm for multiple spherical underwater robots collaboration," in *2021 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 611–616, Takamatsu, Japan, 2021.

[16] S. Sun, X. Zhang, C. Zheng, J. Fu, and C. Zhao, "Underwater acoustical localization of the black box utilizing single autonomous underwater vehicle based on the second-order time difference of arrival," *IEEE Journal of Oceanic Engineering*, vol. 45, no. 4, pp. 1268–1279, 2020.

[17] Y. Su, L. Guo, Z. Jin, and X. Fu, "A mobile-beacon-based iterative localization mechanism in large-scale underwater acoustic sensor networks," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3653–3664, 2021.

[18] R. Diamant, H.-P. Tan, and L. Lampe, "LOS and NLOS classification for underwater acoustic localization," *IEEE Transactions on Mobile Computing*, vol. 13, no. 2, pp. 311–323, 2014.

[19] X. You, Z. Lv, Y. Ding, W. Su, and L. Xiao, "Reinforcement learning based energy efficient underwater localization," in *2020 International Conference on Wireless Communications and Signal Processing (WCSP)*, Nanjing, China, 2020.

[20] Y. Li, X. Hu, Y. Zhuang, Z. Gao, P. Zhang, and N. El-Sheimy, "Deep reinforcement learning (DRL): another perspective for unsupervised wireless localization," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6279–6287, 2020.

[21] J. Yan, Y. Gong, C. Chen, X. Luo, and X. Guan, "AUV-aided localization for internet of underwater things: a reinforcement-learning-based method," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9728–9746, 2020.

[22] S. Basagni, V. Valerio, and P. Gjanci, "Finding MARLIN: exploiting multi-modal communications for reliable and low-latency underwater networking," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, 2017.

[23] H. Huang, Y. Yang, H. Wang, Z. Ding, H. Sari, and F. Adachi, "Deep reinforcement learning for UAV navigation through massive MIMO technique," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1117–1121, 2020.

[24] H. Shi, L. Shi, M. Xu et al., "End-to-end navigation strategy with deep reinforcement learning for mobile robots," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2393–2402, 2020.

[25] Y. Zhou, E. Kampen, and Q. Chu, "Hybrid hierarchical reinforcement learning for online guidance and navigation with partial observability," *Neurocomputing*, vol. 331, no. 28, pp. 443–457, 2019.

[26] F. Fathhinezhad, V. Derhami, and M. Rezaeian, "Supervised fuzzy reinforcement learning for robot navigation," *Applied Soft Computing*, vol. 40, no. 1, pp. 33–41, 2016.

[27] Z. Wang, C. Chen, H. Li, D. Dong, and T. Tarn, "Incremental reinforcement learning with prioritized sweeping for dynamic environments," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 2, pp. 621–632, 2019.

[28] L. Ding, S. Li, H. Gao, C. Chen, and Z. Deng, "Adaptive partial reinforcement learning neural network-based tracking control for wheeled mobile robotic systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 7, pp. 2512–2523, 2020.

[29] M. A. Habib, M. J. Hasan, and J.-M. Kim, "A lightweight deep learning-based approach for concrete crack characterization using acoustic emission signals," *IEEE Access*, vol. 9, pp. 104029–104050, 2021.

[30] Y. Zhang, T. Liu, L. Zhang, and K. Wang, "A deep learning approach for modulation recognition," in *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pp. 1–5, Shanghai, 2018.

[31] Z. Wu, Y. Zhao, Z. Yin, and H. Luo, "Jamming signals classification using convolutional neural network," in *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 62–67, Bilbao, Spain, 2017.

[32] J. Krzyston, R. Bhattacharjea, and A. Stark, "Modulation pattern detection using complex convolutions in deep learning," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 2233–2239, Milan, Italy, 2021.

[33] J. Yan, H. Zhao, B. Pu, X. Luo, C. Chen, and X. Guan, "Energy-efficient target tracking with UASNs: a consensus-based Bayesian approach," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 3, pp. 1361–1375, 2020.

[34] A. Novikov and A. C. Bagtzoglou, "Hydrodynamic model of the lower Hudson river estuarine system and its application for water quality management," *Water Resources Management*, vol. 20, no. 2, pp. 257–276, 2006.

[35] A. C. Bagtzoglou and A. Novikov, "Chaotic behavior and pollution dispersion characteristics in engineered tidal embayments: a numerical investigation1," *JAWRA Journal of the American Water Resources Association*, vol. 43, no. 1, pp. 207–219, 2007.

[36] Z. Liang and Q. Liang, "Optimum cluster size for underwater acoustic sensor networks," in *MILCOM 2006*, pp. 1–5, Washington, DC, USA, 2007.

[37] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, Springer-Verlag, New York, NY, USA, 3rd edition, 2002.

[38] T. Jenserud and S. Ivansson, "Measurements and modeling of effects of out-of-plane reverberation on the power delay profile for underwater acoustic channels," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 4, pp. 807–821, 2015.

[39] W. Carey, J. Doutt, and L. M. Dillman, "Shallow-water transmission measurements taken on the New Jersey continental shelf," *The Journal of the Acoustical Society of America*, vol. 89, no. 4B, p. 1981, 1991.

[40] A. G. Zajic, "Statistical modeling of MIMO mobile-to-mobile underwater channels," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1337–1351, 2011.

[41] B. Zhang, Y. Wang, H. Wang, X. Guan, and Z. Zhuang, "Tracking a duty-cycled autonomous underwater vehicle by underwater wireless sensor networks," *IEEE Access*, vol. 5, pp. 18016–18032, 2017.

[42] M. Naderi, M. Pätzold, R. Hicheri, and N. Youssef, "A geometry-based underwater acoustic channel model allowing for sloped ocean bottom conditions," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2394–2408, 2017.

[43] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2017, https://arxiv.org/abs/1602.07360.

[44] T. Hu and Y. Fei, "QELAR: a machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 6, pp. 796–809, 2010.

[45] Z. He and L. Zhao, "The comparison of four UAV path planning algorithms based on geometry search algorithm," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pp. 33–36, Hangzhou, China, 2017.

[46] Q. Tang, X. Zhang, and L. Zuo, "Initial study on the path plannings algorithms for unmanned aerial vehicles," *AeronauticaI Computer Technique*, vol. 33, pp. 125–128, 2003.

[47] R. Liao, X. You, W. Su, K. Chen, L. Xiao, and E. Cheng, "Reinforcement learning based energy efficient underwater passive localization of hidden mobile node," in *2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, Xian, China, 2021.

[48] S. M. Kay, *Fundamentals Statistical Signal Processing*, Prentice Hall PTR, 1993.