Journal of
*Marine Science
and Engineering*

MDPI

Article

USVs Path Planning for Maritime Search and Rescue Based on POS-DQN: Probability of Success-Deep Q-Network

Lu Liu [1], Qihe Shan [1,*] and Qi Xu [2]

[1] Navigation College, Dalian Maritime University, Dalian 116026, China; ll05290630@163.com
[2] Research Institute of Intelligent Networks, Zhejiang Lab, Hangzhou 311121, China; xuqi@zhejianglab.com
* Correspondence: shanqihe@163.com

Abstract: Efficient maritime search and rescue (SAR) is crucial for responding to maritime emergencies. In traditional SAR, fixed search path planning is inefficient and cannot prioritize high-probability regions, which has significant limitations. To solve the above problems, this paper proposes unmanned surface vehicles (USVs) path planning for maritime SAR based on POS-DQN so that USVs can perform SAR tasks reasonably and efficiently. Firstly, the search region is allocated as a whole using an improved task allocation algorithm so that the task region of each USV has priority and no duplication. Secondly, this paper considers the probability of success (POS) of the search environment and proposes a POS-DQN algorithm based on deep reinforcement learning. This algorithm can adapt to the complex and changing environment of SAR. It designs a probability weight reward function and trains USV agents to obtain the optimal search path. Finally, based on the simulation results, by considering the complete coverage of obstacle avoidance and collision avoidance, the search path using this algorithm can prioritize high-probability regions and improve the efficiency of SAR.

Keywords: SAR; task allocation; deep reinforcement learning

check for updates

Citation: Liu, L.; Shan, Q.; Xu, Q. USVs Path Planning for Maritime Search and Rescue Based on POS-DQN: Probability of Success-Deep Q-Network. *J. Mar. Sci. Eng.* **2024**, *12*, 1158. https://doi.org/10.3390/jmse12071158

Academic Editor: Sergei Chernyi

Received: 22 June 2024
Revised: 5 July 2024
Accepted: 8 July 2024
Published: 10 July 2024

1. Introduction

As a core component of global logistics, international ocean cargo transportation has a huge impact on world trade [1]. Due to the complex and changeable maritime environment, maritime accidents caused by natural and human factors often occur during transport. According to the data statistics of the China Maritime Search and Rescue Centre (CMSRC), last year, an average of 133 accidents occurred each month, and the success rate of SAR was 97 percent. During the SAR process, most people who fell into the water were missing or died because of the long search time [2]. Therefore, improving the timeliness and efficiency of maritime SAR has important research significance.

Maritime SAR is a necessary means to protect the lives and property of people and the maritime ecological environment. It is also an important part of the maritime emergency response system. SAR includes two parts: search and rescue. Search is the premise and key to all rescue work. With the development of machine technology and electronic equipment, unmanned SAR has a faster response than manual vessel SAR. Unmanned SAR is capable of performing SAR tasks in high-risk regions and harsh weather conditions. Therefore, unmanned SAR has gradually become an efficient maritime SAR program [3]. As vessels with the outstanding characteristics of being unmanned and intelligent, USVs have the advantages of reliability, flexibility, and easy expansion. They are less affected by harsh environments and can maintain stability [4]. Through reasonable task allocation, USVs can efficiently perform maritime SAR tasks, so they have been widely used in unmanned SAR in recent years [5,6]. Path planning in maritime SAR tasks is the key and challenge of the entire process, and it plays a vital role in improving SAR efficiency.

USV path planning for maritime SAR is mainly divided into two parts: the task allocation of USVs in the search region and the path planning performed by each USV according to the task region. Due to the dynamic nature of the search area, IP network-based communication methods present difficulties in meeting the needs of efficient routing and data transmission [7]. To improve the efficiency of SAR, it is crucial to quickly and accurately transmit the latest information of the search region to the USV. Therefore, this paper adopts the Geo Networking communication method. It is a mobile self-organizing communication network layer protocol based on wireless technology that provides communication in a mobile environment without the need for coordination infrastructure, using geographical location to spread information and transmit data packets [8]. After determining the search region, USVs must coordinate and command in a unified manner, reasonably allocating each USV's SAR tasks to participate in maritime SAR tasks as quickly as possible. Maritime SAR task allocation is the process of reasonably allocating tasks to various SAR equipment to improve the success rate of the task [9]. The task allocation algorithm mainly focuses on two aspects: priority division and overall regional division. Priority division is the task allocation of multiple maritime SAR vehicles based on the optimal search theory [10]. It is widely used in real maritime SAR operations [11,12]. Although this method gives priority to searching high-probability regions, the generated task regions are repeated and cannot cover the entire search region. The overall division of the region calculates the area allocated to each task region based on the initial position, search capability, maximum speed, and other parameters of the search equipment [13–15]. Although the search regions allocated by this method are non-repetitive and completely covered, they ignore the priority of the tasks. An efficient maritime SAR task allocation algorithm must combine the two methods to allocate tasks in order to reasonably achieve the highest efficiency.

After dividing the search region using a task allocation algorithm, each USV performs path planning according to the allocated task region. The path planning is used to develop an optimal navigation path that avoids conflicts with other SAR equipment and obstacles. It is mainly divided into two processes: optimal path planning for SAR equipment to reach the vicinity of the search region quickly from the initial position, and search path planning within the task region [16–19]. In SAR tasks, search path planning is a key factor in improving the success rate. Search path planning is mainly divided into traditional search methods and region coverage methods. The traditional search method considers the performance of unmanned equipment, and its main search methods include a fan-shaped search [20], the parallel line scanning (PA) search [21], and the cross-shift line scanning (CA) search [22]. The region coverage method uses coverage path planning (CPP). Its core goal is to design a path that can avoid obstacles and traverse all points in a region or space [23]. According to the research of Tan [24], CPP is divided into classical algorithms and heuristic-based algorithms. Classical algorithms mainly include the random walk algorithm [25], the artificial potential field method [26], and the bug algorithm [27]. Heuristic algorithms include the A* algorithm [28], the swarm intelligence algorithm [29], human-inspired algorithms [30], etc. Human-inspired algorithms include deep reinforcement learning algorithms, which have both the decisiveness of reinforcement learning and the perception of deep learning. These are conducive to path planning in complex environments. In the case of environmental changes, deep reinforcement learning allows the agent to interact with the environment. It uses feedback from the environment to iterate and continuously optimize its path planning strategy [31], which is very important for path planning in complex SAR environments. This paper focuses on path planning in the search region, where the state space is two-dimensional and the action space is discrete. Therefore, this paper chooses the DQN algorithm as the basic algorithm. Through deep neural networks, DQN can efficiently and flexibly approximate the Q-value function of high-dimensional complex state spaces, which greatly enhances the expressiveness and generalization capabilities of the model [32]. In addition, it uses the experience replay buffer so that the agent can repeatedly learn from past experiences, which improves data utilization and helps stabilize the learning process. However, for the path planning of searching in the task region, it is necessary not only to achieve complete coverage to avoid obstacles but also to prioritize the search of high-probability regions. Therefore, it is crucial that the POS adjustment based on the DQN algorithm is solved according to the optimal search theory.

Therefore, this paper proposes a POS-DQN deep reinforcement learning algorithm to solve the path planning of USVs in maritime SAR. The main innovations can be summarized as follows:

1. This paper uses an improved task allocation algorithm to allocate the search region of USVs according to the parameters of the maritime SAR environment and SAR equipment. Finally, the task region assigned to each USV satisfies the priority of the task and is reasonably allocated without duplication.
2. This paper proposes a POS-DQN deep reinforcement learning algorithm to solve the problem of maritime SAR path planning. This algorithm designs a probability weight reward function for SAR to train each USV to avoid obstacles and collisions and prioritize high-probability regions. It can complete the SAR task safely and efficiently.

The remaining structure of this paper is as follows: the second part describes task allocation in the search region; the third part describes the POS-DQN algorithm; the fourth part provides an experimental simulation, and the fifth part provides the conclusions obtained in this paper.

2. Task Allocation in the Search Region

In maritime SAR, it is essential to determine the location of the search region first. With the increase in the number of rescue USVs, the communication volume of USV SAR scenarios based on IP networks has increased significantly, which will cause information redundancy and network congestion, seriously affecting the efficiency of SAR. Considering that the accuracy of USV SAR receiving signals is highly dependent on the timeliness of the geographical location information of the search region, this paper adopts a communication network mode based on geographic location identification that is adapted to the requirements of effectively saving maritime communication resources. The communication network mode based on geographic location identification is a network layer protocol for mobile self-organizing communication that uses geographic location to propagate information and transmit data packets [33]. It can provide communication in a mobile environment without the need for coordinated infrastructure [34]. In maritime SAR, the USV search and rescue communication network architecture is established based on the Geo Networking mode, as shown in Figure 1.

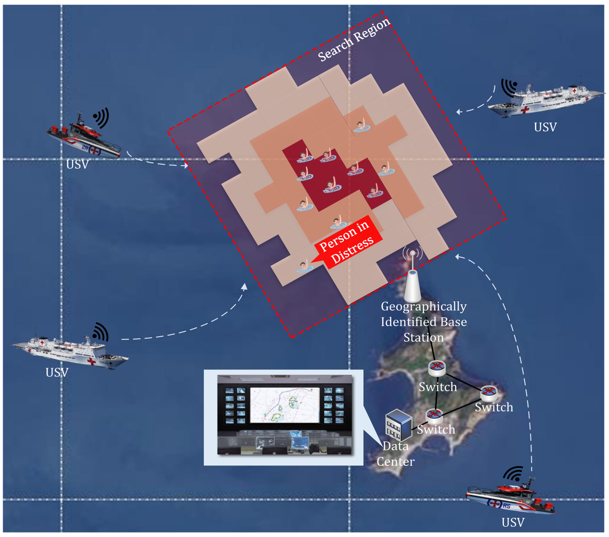


Figure 1. USVs SAR communication network architecture diagram based on Geo Networking.

In maritime SAR involving USVs, the allocation of the search region must take into account the priority of the task, and the task region must be completely covered without duplication to achieve the maximum probability of success efficiently and quickly. Therefore, this paper combines the two methods of priority division and overall division to propose an improved task allocation algorithm. This section introduces in detail the optimal search theory based on priority division and the improved task allocation algorithm. The important variables in this section are shown in Table 1.

Table 1. Important variables.

Variable	Significance
i	Grid i
m_i	Number of particles in grid i
M	Number of particles contained in the total distribution region
C	Coverage (measure of how well the search region is covered)
W	Sweep width
R	Route spacing
a	USV a
n	Total number of USVs
D_a	Initial distance of USV from the search region
$\hat{v}_a$	Maximum speed of USV a
T	Time spent on the entire task = time spent by the USVs on SAR tasks
$\widehat{T}_a$	Time for USV to reach the search region at full speed
T_a	Total search time for the search task of USV a
$\bar{T}_a$	Search time in the search region covered by USV a
b_a	Search capability of USV a—represents the ability of the equipment to cover the area within a certain region
S_a	Search region allocated to USV a
S_n	Search region allocated to USV n
l	Scan line
l_s	Initial point of the scan line
l_e	End of the scan line
m	Number of vertices in the search region and sum of the initial points of the search
W_m	Point m in the point set
k	k in the point set

2.1. Optimal Search Theory

Optimal search theory is the basis for determining the search region, planning SAR forces, and allocating search tasks [35]. Optimal search theory can be summarized in three important components: probability of successful search (POS), probability of containment (POC), and probability of detection (POD).

- POS
  POS is the probability of searching the SAR objectives successfully within the search region. POS is the main measurement indicator of SAR operations. The larger the POS, the higher the search efficiency. It is related to POC and POD. The calculation formula is as follows:

$$POS = POC \times POD, \tag{1}$$

- POC
  POC is the probability that the SAR objectives exist in the search region. The search region is expanded until all regions where particles exist are included, then the POC of the region reaches 100 percent. In actual maritime SAR tasks, the number of SAR forces is often limited, which requires them to give priority to regions with higher POC [36]. Therefore, the search region is usually divided into square grids of the same size. The probability of SAR objectives being present in the search region is quantified by calculating the POC of each grid unit. The calculation formula is as follows:

$$POC = \frac{m_i}{M}, \tag{2}$$

- POD
  POD is the probability that the SAR objectives can be detected within a particular search region. The calculation formula is as follows:

$$POD = 1 - e^{-C}, \tag{3}$$

  POD consists of two important concepts: coverage and sweep width. Coverage is a measure of how well the search equipment covers the search region during the search process. The calculation formula is as follows:

$$C = \frac{W}{R}, \tag{4}$$

  Sweep width is a key parameter that determines the efficiency of maritime SAR. It is affected by a variety of factors, including the variable maritime environment, the performance of the SAR equipment, and the characteristics of the SAR objective. IAMSAR provides correction factors for sweep width under different environmental conditions [37], as shown in Table 2. It is the effective distance to locate a search object within a particular search region.

Table 2. Weather correction factors.

Weather: Winds (km/h) or Seas (m)	Search Object	
	Person in water, raft, or vessel < 10 m	Other search objects
Winds 0–28 km/h or seas 0–1 m	1.0	1.0
Winds 28–46 km/h or seas 1–1.5 m	0.5	0.9
Winds > 46 km/h or seas > 1.5 m	0.25	0.9

2.2. Improved Task Allocation Algorithm

Maritime SAR is often a collaborative SAR process with multiple SAR vehicles. Because the coverage capability of a single USV is limited, it cannot quickly cover the entire search and rescue region. Therefore, USVs are usually required to perform SAR tasks. The region where the SAR objective is located should have a reasonable task allocation strategy. Different task regions are assigned to different USVs so that the SAR tasks can be completed quickly and efficiently. The SAR task allocation should consider the initial location, maximum speed, search capability, and POS of the search region. Therefore, this paper combines the search region task allocation algorithm proposed by Xing [15] and optimal search theory to make adjustments. The target of the task allocation is to select the best USVs to operate together. In this way, it takes the least time T to complete the search coverage of the entire region to be searched. Afterward, the positions where the USVs reach the search region are the search initial points. The search region is divided according to the initial point to complete the task allocation of USVs. The algorithm steps are as follows:

1. Because each USV's initial location, maximum speed, and search capability are different, first, the time to reach the search region at maximum speed is calculated based on the initial distance of USV a from the search region and the maximum speed, as shown in Formula (5). The total area of the search region is S, and the search time covered by USV a in the search region is $\widehat{T}_a$. The total search time formula for the USV a search task is shown in Formula (6). After that, the optimal SAR USV collaboration should minimize the search time, as shown in Formula (7). Finally, the region constraint set of the SAR unit $\{S_1, S_2, \ldots, S_a \ldots S_n\}$ is obtained according to Formula (8).

$$\widehat{T}_a = \frac{D_a}{\hat{v}_a}, \tag{5}$$

$$T_a = \bar{T}_a + \widehat{T}_a, \tag{6}$$

$$\min T = \frac{\sum_{a=1}^{n}\left(\widehat{T}_a\right)b_a + S}{\sum_{a=1}^{n} b_a}, \tag{7}$$

$$\sum_{a=1}^{n}\left(T - \widehat{T}_a\right)b_a = S_a, \tag{8}$$

2. The initial location point of the USV is connected to the grid center point with the highest POC in the search region to obtain the ideal path for the USV to reach the search region. If the USV is outside the search region, the intersection of the path and the boundary of the search region is the initial point of the search; if the USV is within the search region, the intersection of the extension line of the route and the boundary of the search region is the initial point of the search. The search initial points are sorted in counterclockwise order, and the obtained point set is $\mathrm{N} = \{N_1, N_2, N_3, \ldots, N_n\}$.
3. The search initial points and search region vertices are sorted in counterclockwise order to obtain the point set $W = \{W_1, W_2, W_3, \ldots, W_m\}$, as shown in Figure 2.

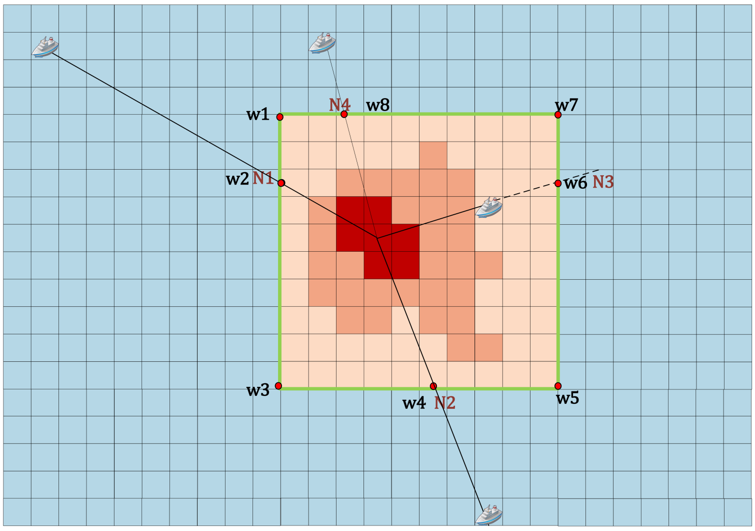


Figure 2. Schematic diagram of the point set.

4. The scan line is defined as l, which represents a straight line segment with both endpoints on the edge of the search region. The scan line $l = (l_s, l_e)$, $N_1 = W_k$, $k = 0, 1, 2 \ldots m$, $l \leftarrow (W_1, W_k)$ is initialized to obtain the polygon, and its region S is calculated.
5. If $S < S_a, l \leftarrow (W_1, W_{k+1})$, go to step (4); if $S < S_a$, move the start point of the scan line counterclockwise along the boundary of the search area until the area $S = S_a$, then go to step (6).
6. Move the l_s, l_e position while holding on the region line until $S = S_a$, which is horizontal or perpendicular to the region line. This gives the assigned region for the USV a. Repeat the above steps to obtain the search region for each USV.

3. POS-DQN Algorithm

3.1. DQN Algorithm

Maritime SAR path planning can be formulated as a Markov decision process (MDP). MDP is a mathematical framework for describing decision-making in a stochastic environment, consisting of tuples $\{S, A, P, R, \gamma\}$. To search the SAR region quickly and efficiently, this paper proposes a maritime SAR path planning model based on deep reinforcement learning. In this model, reinforcement learning plays an important role in scenarios that require interactions with the environment. In such scenarios, the agent selects a corresponding action based on its strategy according to the current state of the environment. Once the action is executed, the state of the environment will be changed to a new state, and the agent will receive a reward value associated with the action [38]. The agent will evaluate and adjust its strategy based on the reward value it receives after each action so that it can receive higher rewards in future decisions, as shown in Figure 3.

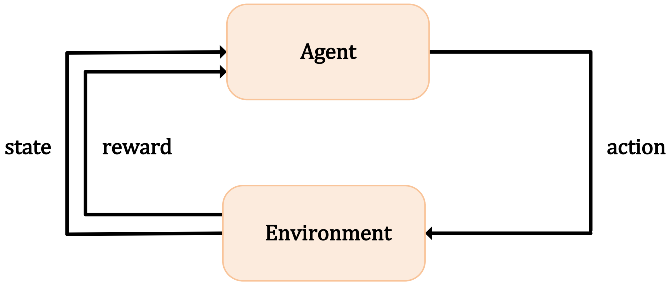


Figure 3. Reinforcement learning process.

In this paper, the deep Q network (DQN) algorithm [39] combines reinforcement learning with deep learning. It is used as the basic algorithm for training maritime SAR path planning. In the case of a two-dimensional state space and a discrete action space, DQN can perform efficient policy learning in a high-dimensional complex state space using a deep neural network. It greatly enhances the generalization ability of the model. In addition, DQN uses an experience replay buffer, so the agent can repeatedly learn from past experiences to improve data utilization. During the training process, the USV agent collects a series of observations, actions taken, and rewards received through interactions with the environment. The USV agent evaluates the expected value of the cumulative rewards that each possible action in its list of actions can bring in the future based on its current state. Based on these evaluations, the USV agent selects the action with the highest expected reward as its next action. In this way, the USV agent can learn and optimize its path planning strategy in complex maritime SAR tasks to achieve efficient and accurate SAR results. The important parameters in the algorithm are shown in Table 3.

Table 3. Important variables.

Variable	Significance
S	USV agent's state set
A	Set of all executable actions of the USV agent
P	State transfer probability
R	Reward function
γ	Discount factor between 0 and 1 to balance the importance of immediate and future rewards
s	Current state
a	Current action
θ	Parameters to evaluate network
r	Instant rewards
Q	Expectation of future earnings
Q^*	Optimal value of Q
E	Expectation
s'	New state
a'	New action
y_i	Target Q

Table 3. *Cont.*

Variable	Significance
θ^{-}	Parameters of target network
D	Experience replay pool
$\{a_1, a_2, a_3, a_4\}$	Actions in four directions: up, down, left, and right
π	Strategy of action selection
ε	Greedy probability
p	Probability
$L(\theta)$	Loss function
$\nabla_\theta L(\theta)$	Gradient
α	Learning rate
r_t	Probability weight reward function
R_m	Obstacle reward
R_O	Number of visits rewarded
R_{weight}	Weight reward

The DQN algorithm consists of two networks: the evaluate network and the target network. The evaluate network is responsible for learning the mapping of the state of the environment to the value of an action in real time. The USV agent observes the state of the environment at each time step. The evaluate network is then used to compute the predicted value Q of each possible action in that state, which is the expected future benefit of that action $Q(s, a : \theta)$. The optimal value $Q^*(s, a)$ for calculating the value Q using Bellman's equation is as follows:

$$Q^*(s,a) = E\left[r + \gamma \max_{a'} Q^*(s', a')\right], \tag{9}$$

The target network is a delayed update version of the evaluate network. Though they have the same structure, the target network θ^- is not updated in real-time, but a copy of the evaluate network θ is given to θ^- after a specific time interval. This is done to reduce the instability during the training process. The target network is used to generate the target value Q, which provides a more stable target value Q. This helps to reduce the variance during the learning process. The formula for the target value Q is as follows:

$$y_i = r + \gamma \max_{a'} Q'(s', a' : \theta^-), \tag{10}$$

The USV agent action designed in this paper can select four directions and then select the action to execute according to the ε-greedy to obtain the reward and observe the new state s. This strategy aims to balance the relationship between exploration and utilization to optimize long-term benefits in the decision-making process. ε is a parameter between 0 and 1, representing the probability of exhibiting random exploration behavior. It can quickly explore the entire state space at an early stage and prevent the algorithm from falling into local optimal solutions. The formula for selecting the action is as follows:

$$\text{random action } = \{a_1, a_2, a_3, a_4\}, \tag{11}$$

$$a = \begin{cases} \operatorname{argmax} Q(s, a; \theta) & p = 1 - \varepsilon \\ \text{random action} & p = \varepsilon \end{cases} \tag{12}$$

Then, the formula for the strategy π of action selection is as follows:

$$\pi = \begin{cases} 1 - \varepsilon + \varepsilon/4 & a = \text{random action} \\ \varepsilon/4 & a = \operatorname{argmax} Q(s, a; \theta) \end{cases} \tag{13}$$

A series of tuples (s, a, r, s') are obtained via continuous iteration and stored in the experience replay pool D. The model selects a set of quintuples from the experience replay pool at each round to calculate the loss function based on the mean square error

(MSE). The parameters of the evaluate network are updated using the MSE as the loss function. Adjustments, θ, are made to minimize the loss function and ultimately improve the expectation of future benefits, $Q(s,a;\theta)$, predicted by the evaluate network. The formula of the loss function is as follows:

$$L(\theta) = E_{(s,a,r,s')-D}\left[\left(y_i - Q(s,a;\theta)\right)^2\right], \tag{14}$$

In the parameter update phase, the gradient descent method is used to adjust the evaluate network's θ to minimize the loss function $L(\theta)$ with the gradient formula:

$$\nabla_\theta L(\theta) = E_{(s,a,r,s')-D}\left[\left(r + \gamma \max_{a'} Q(s',a';\theta^-) - Q(s,a;\theta)\right)\nabla_\theta Q(s,a;\theta)\right], \tag{15}$$

where the parameters θ of the evaluate network are updated according to the gradient formula. The updated formula is as follows:

$$\theta \leftarrow \theta - \alpha \cdot \nabla_\theta L(\theta), \tag{16}$$

3.2. POS-DQN Algorithm

This paper proposes a POS-DQN algorithm to solve SAR path planning. By designing the reward function, the USV agent obtains a search path based on algorithm learning and training in the maritime SAR environment model. The final generated path avoids obstacles, prioritizes searching for high-probability regions, and provides complete coverage of the region without duplication.

In the field of reinforcement learning, the core target of the USV agent is to maximize the cumulative reward it obtains by selecting actions. The key to this process is the design of the reward function, which acts as a feedback mechanism for the USV agent's actions and plays a crucial role in the USV agent's decision-making process. When the USV agent performs an action, the environment will give the agent an immediate reward or punishment based on the result of the action. The value of this reward or punishment represents the positive or negative impact of the action on the USV agent's long-term targets. The USV agent will use this feedback information to evaluate the quality of its actions and adjust its strategy to obtain higher rewards in future decisions.

In the USVs' maritime SAR path planning, it must be ensured that the designed reward function can enable the USV agent to maximize the benefit. At the same time, the generated path can also achieve complete coverage, avoiding obstacles and searching first in high-probability regions. According to the above requirements, this paper designs the probability weight reward function, as shown in Figure 4. The formula of the probability weight reward function is as follows:

$$r_t = R_m \times R_o \times R_{\text{weight}}, \tag{17}$$

Safely avoiding obstacles is the basis of path planning. It is essential to ensure effective searches in complex environments, so the obstacle reward formula is designed as follows:

$$R_m = \begin{cases} r_o < 0 & \text{With obstacles} \\ r_p > 0 & \text{Without obstacles} \end{cases}, \tag{18}$$

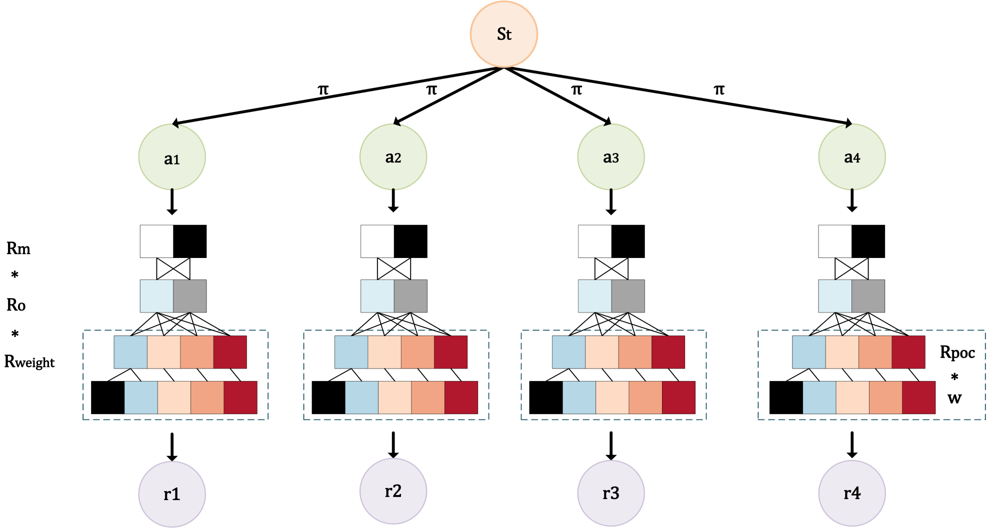


Figure 4. Probability weight reward function schematic diagram.

Using the minimum SAR time to achieve an efficient search region should reduce the repetition rate of the search, so the reward formula for the design of the number of visits is as follows:

$$R_O = \begin{cases} r_1 > 0 & \text{First visit} \\ r_2 < 0 & \text{Repeat visit} \end{cases}, \tag{19}$$

The search of high-probability regions is prioritized, and the grid with the highest POS is quickly reached. When the scan width is fixed, the grid with a large POC should be searched first, so the weight reward formula is designed as follows:

$$R_{\text{weight}} = w \times R_{POC}, \tag{20}$$

The weight reward is divided into two parts to satisfy the high probability priority search for the search region. The POC reward is designed according to the POC of the grid unit color. The larger the POC, the greater the reward of the grid. The POC reward formula for the grid unit color is as follows:

$$R_{POC} = 100 \times POC, \tag{21}$$

The weight is designed based on the POC and obstacles of the grid unit colors. The larger the POC, the greater the weight of the grid. Because avoiding obstacles safely is the most important, obstacles have the highest weight. The weight formula is as follows:

$$w = \begin{cases} w_{POC} \\ w_m \end{cases}, w_m > w_{POC}, \tag{22}$$

The POS-DQN algorithm is combined with the maritime SAR environment model to obtain the maritime SAR path planning model based on reinforcement learning in this paper. The path planning obtained after training can achieve complete coverage of avoiding obstacles and prioritizing searches in high-probability regions. The maritime SAR path planning model is shown in Figure 5.

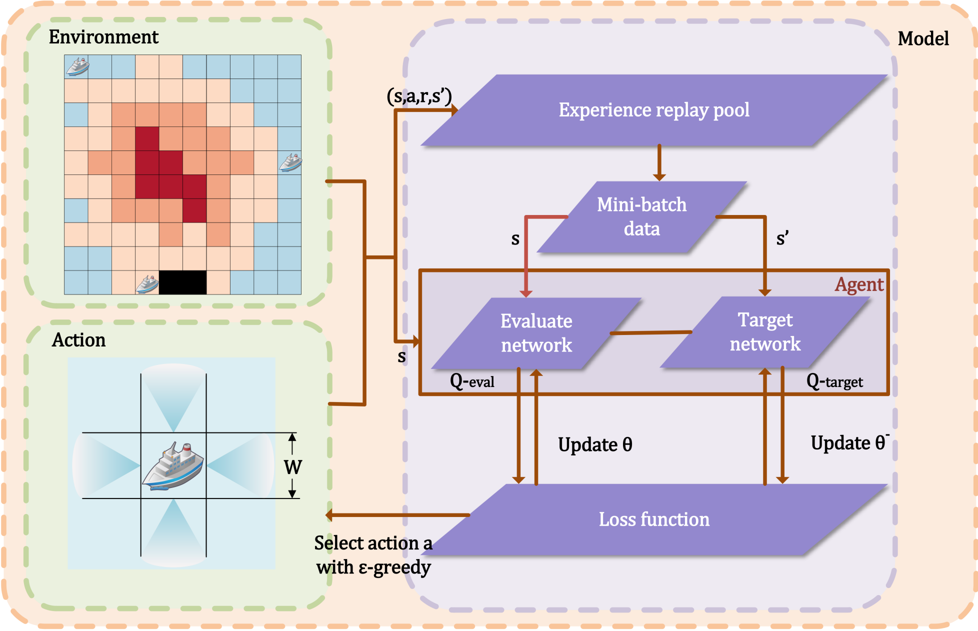


Figure 5. Maritime SAR path planning model.

Therefore, the pseudo-code for POS-DQN training used in this paper is proposed in Algorithm 1.

Algorithm 1 POS-DQN Algorithm

```
Initialize the experience pool $D$ with capacity $N$
The parameters $\theta$ of the value function $Q$ initialize the evaluate network
The parameters $\theta'$ of the Target network's target action value function $Q'$ are set
while episode = 1, $M$ do
    Initialize state $S_0$
    for t = 1, $T$ do
        Using the $\varepsilon$-greedy policy, select action $a_t$
        if explore then
            Choose a random action $a_t$
        else
            $a_t = \text{argmax}_a Q(s_t, a; \theta)$
        end if
        Execute action $a_t$, obtain reward $r_t$ and next state $s_{t+1}$ using the reward function
        $r_t = R_t \times R_m \times R_{\text{weight}}$
        Store the tuple $(s_t, a_t, r_t, s_{t+1})$ in experience pool $D$
        Randomly select a mini-batch of samples from $D$
        Calculate the target $y_i$ as follows:
        if terminal $s_{j+1}$ then
            $y_i = r_j$
        else
            $y_i = r_j + \gamma \max_{a'} Q(s_{j+1}, a'; \theta)$
        end if
        Perform gradient descent on the network parameters $\theta$ to minimize the loss function $(y_j - Q(s_j, a_j; \theta))^2$, updating the parameters of the value function $Q$
        Every $C$ steps, reset the Target network parameters $Q' = Q$
    end for
end while
```

4. Experimental Simulation

4.1. Experimental Environment Modeling

To verify the effectiveness of the maritime SAR path planning model, we conduct simulation experiments for training analysis. By selecting the vicinity of Bohai Bay, the 12 h drift simulation experiment is simulated for the person in distress whose vessel experienced a collision. There are many factors that affect maritime SAR, including wind pressure, wind currents, ocean currents, tidal currents, and the factors of the drifting objects themselves. However, the wind and current have the greatest impacts. When determining the search region, the sweep width of the USV should be considered, which affects the POC of the region. To achieve an efficient and fast search, scientific and reasonable calculation methods must be adopted to determine the objective search region. First, the drift trajectory is predicted using ocean environment data and the last known location. In the drift prediction process, SAR objectives are treated as particles. This paper uses the Lagrangian particle tracking algorithm [40] to calculate the displacement trajectory of particles under the influence of the marine environment. The calculation formula is as follows:

$$\mathrm{d}X/\mathrm{d}t = A(X_t) + B(X,t)Z_n, \tag{23}$$

where X_t is the displacement of the particle, $A(X_t)$ is the drift coefficient, $B(X,t)$ is the diffusion coefficient, and Z_n is a random number. The drift trajectory is shown in Figure 6.

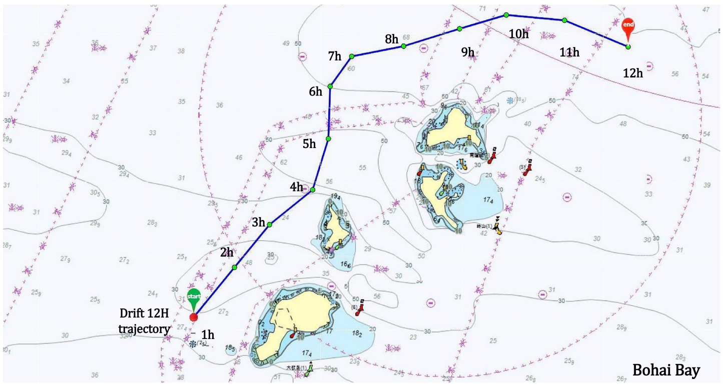


Figure 6. 12 h drift trajectory simulation in Bohai Bay.

After the drift calculation, the objective particle distribution map at different times is obtained to determine the search region. For path planning of USV navigation, it is first necessary to know the locations of obstacles. The grid-based method is usually used to process the electronic navigational chart (ENC) of the maritime environment [41,42]. The ENC of the search region is divided into regular grids according to the grid-based method, and the size of the grid unit is designed according to the sweep width of the USV. Each grid is independent. The center point coordinates of each grid are used as the location coordinates of the grid. The POC of each grid is calculated based on the particle distribution map obtained by drifting. The grid unit color is assigned according to the size of the POC. This color grid map is the POC distribution map of the search region. The POC distribution map is combined with the grid map of the search region to generate a maritime SAR environment model, as shown in Figure 7.

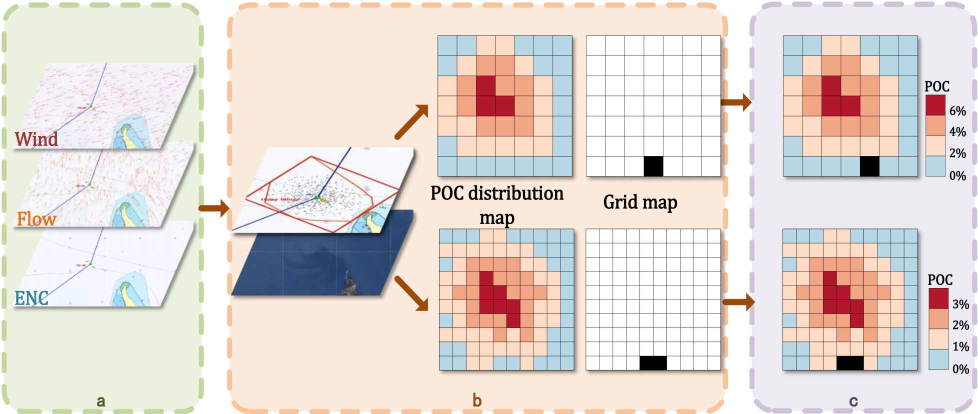


Figure 7. Maritime SAR environment model. (**a**) Wind charts, flow charts, and electronic charts at that time. (**b**) POC distribution map and grid map. (**c**) Ultimate maritime SAR environment.

4.2. Comparative Experiment

The training was based on the maritime SAR environment, with the parameters set as shown in Table 4.

Table 4. Parameter settings.

Parameter	Significant
num action = 4	Action
lr = 0.008	Learning rate
batch size = 64	Batch size
memory capacity = 2000	Memory length
gamma = 0.95	Initial action selection strategy

A single USV is trained in a maritime SAR environment, and comparative experiments are performed using the parallel line scanning algorithm (PA), the cross-shift line scanning algorithm (CA), the A* algorithm, and the DQN algorithm with the POS-DQN algorithm used in this paper. The experimental results of each algorithm are summarized in Table 6 and the final generated paths are shown in Figure 8.

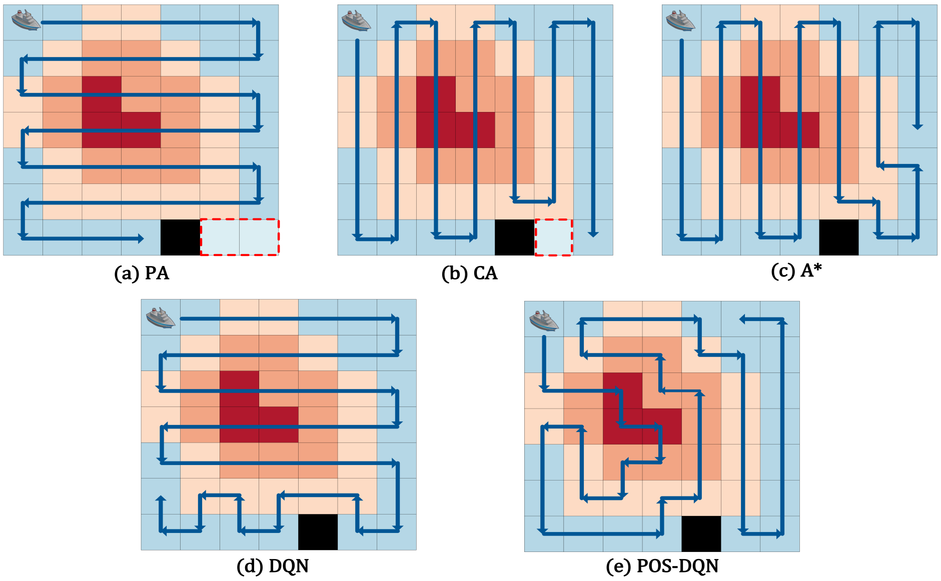

(a) PA
(b) CA
(c) A*
(d) DQN
(e) POS-DQN

Figure 8. Final single USV path diagrams generated by (**a**) the PA algorithm, (**b**) the CA method, (**c**) the A* algorithm, (**d**) the DQN algorithm, (**e**) the POS-DQN algorithm. Red highlighted regions indicate unsearched regions.

The training results of a single USV path using this method achieve complete coverage without obstacles or duplication. Although the path of this algorithm can prioritize the search of high-probability regions, the effect is not optimal because it needs to consider complete coverage. In addition, it takes a long time for a single USV to search the entire region, and the SAR is not timely enough. Therefore, USVs are usually used for SAR.

Three USVs are selected, with different initial locations within 3 nmi of the center of the search region and the highest POC at that time. Their initial locations are shown in Figure 9. Their sweep width is the same. The parameters of the USVs are given in Table 5.

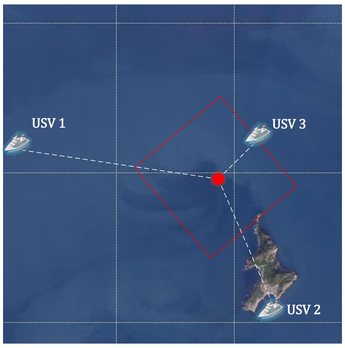


Figure 9. USVs' initial locations.

Table 5. The parameters of the USVs.

No.	Initial Distance (n mile)	Maximum Vessel Speed (n mile/h)	SAR Capability (n mile^2/h)
1	1.5	5	3
2	0.9	15	1
3	0	2	0.9

USVs allocate tasks according to the improved task allocation algorithm, and the final allocation region result diagram is shown in Figure 10.

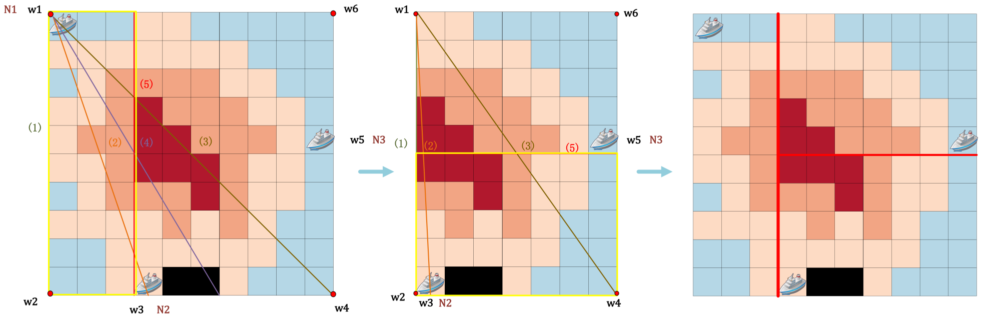


Figure 10. The final allocation region result diagram.

USVs are trained in the maritime SAR environment, and comparative experiments are performed using the PA algorithm, the CA algorithm, the A* algorithm, the DQN algorithm, and the POS-DQN algorithm. The experimental results of each algorithm are summarized in Table 6, and the final generated paths are shown in Figure 11.

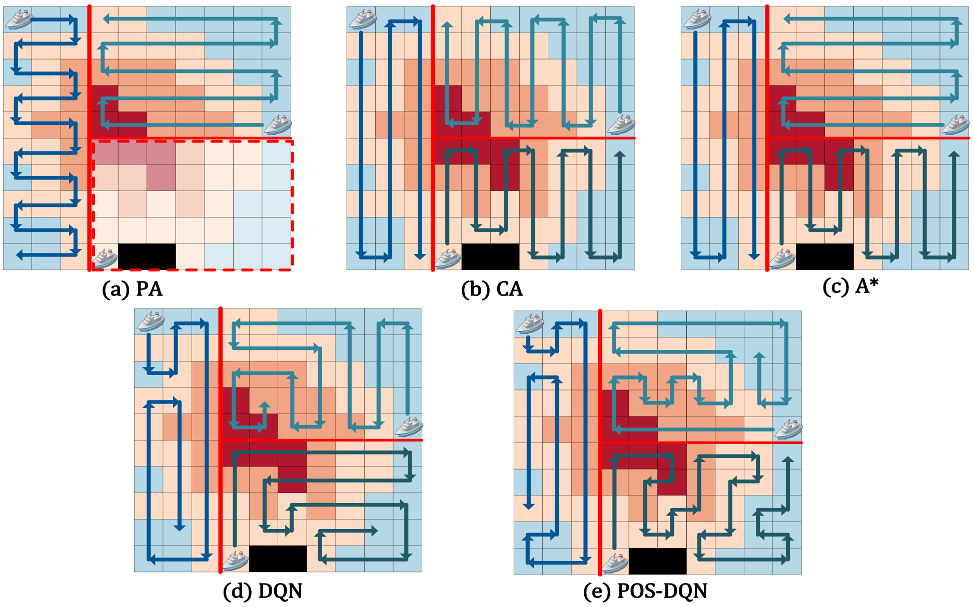


Figure 11. Final USV cluster path diagrams generated by (**a**) the PA algorithm, (**b**) the CA method, (**c**) the A* algorithm, (**d**) the DQN algorithm, (**e**) the POS-DQN algorithm. Red highlighted regions indicate unsearched regions.

The training results of the USVs paths using this method achieve complete coverage of obstacle avoidance and no repetition rate, and they preferentially search for high-probability regions. The final path in the SAR region is shown in Figure 12.

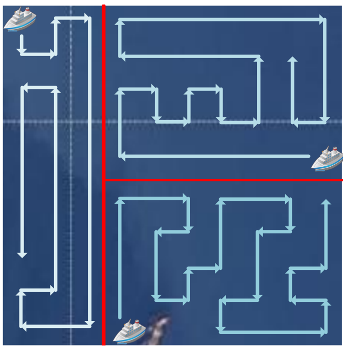
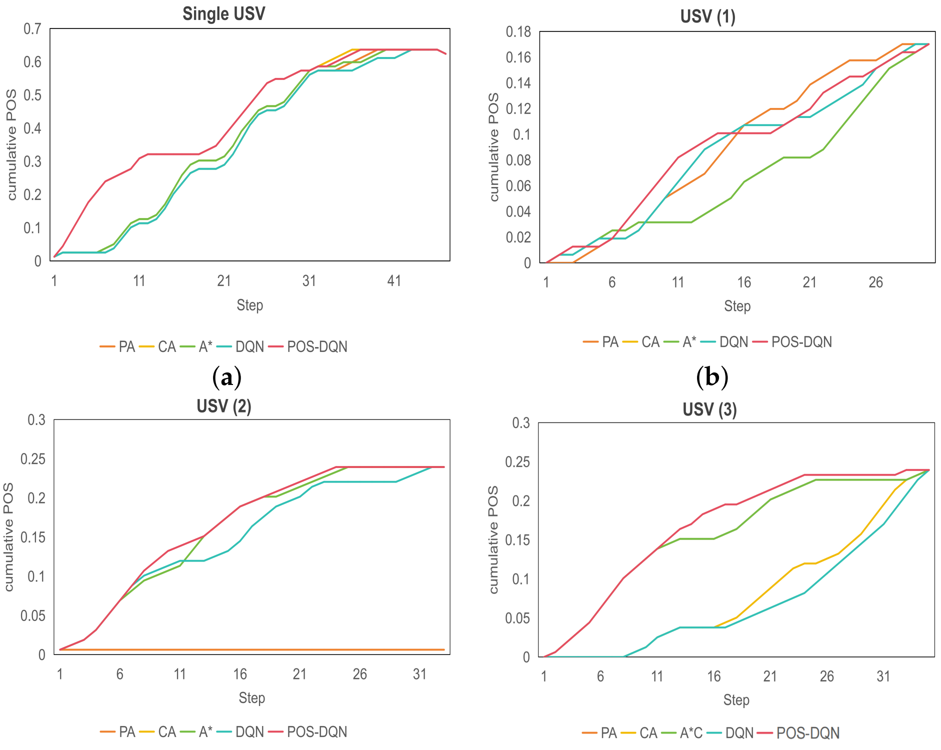


Figure 12. The final path in the SAR region.

4.3. Analysis of Results

This paper calculates and analyzes the cumulative POS at each step to evaluate the above algorithms. As shown in Figure 13, The algorithm used in this paper has the fastest change in growth rate. The cumulative POS of this algorithm is higher for the same number of steps, which indicates that high-probability regions can be searched preferentially. The final cumulative POS is the same as the total POS of the task region, which represents complete coverage of the whole region. This paper also calculates and summarizes the repeated coverage and step, as shown in Table 6, of the experimental results. Compared with other algorithms, this paper uses the POS-DQN algorithm's maritime SAR to train according to the reward function of the designed probability weight. A single USV or multiple USVs can achieve complete coverage according to the divided task search region, without repeated coverage, avoiding obstacles and covering high-probability regions first. However, the paths generated by the other algorithms cannot achieve complete coverage and cannot prioritize the high-probability regions.



(**a**) (**b**)

(**c**) (**d**)

Figure 13. The cumulative POS for each step. (**a**) Single USV; (**b**) USV 1; (**c**) USV 2; (**d**) USV 3.

Table 6. Experimental results.

Algorithm	USV	Coverage (%)	Repeated Coverage (%)	Step	Priority Search for High-Probability Regions
PA	Single USV	0.96	0	46	×
	Multiple USVs	0.64	0	64	×
CA	Single USV	0.98	0	46	×
	Multiple USVs	1	0	97	×
A*	Single USV	1	0	47	×
	Multiple USVs	1	0	97	×
DQN	Single USV	1	0	47	×
	Multiple USVs	1	0	97	×
POS-DQN	Single USV	1	0	47	✓
	multiple USVs	1	0	97	✓

5. Conclusions

This paper has established a maritime SAR path planning model. It has utilized real-time data from the ocean region and inferred drift trajectories to calculate the POC based on the particle distribution at that time. Grid unit colors have been assigned to generate a POC distribution map, and this map has been combined with the grid map of the ENC to build a maritime SAR environment model. Using an improved task allocation algorithm, task regions have been divided among USVs. The USV agent has been trained to generate the optimal path by setting the parameters of the POS-DQN algorithm. Comparative experimental results have shown that the path generated by the algorithm training in this paper can avoid obstacles, prioritize high-probability regions, and achieve complete coverage without a repetition rate.

However, this paper only considers the static environment at the current moment, which has significant limitations. During the SAR process, the SAR objectives will be affected by wind, current, and other weather conditions and dynamic drift. The water temperature in different regions will affect the total time that SAR objectives can wait for rescue. In future work, the model in this paper will adjust the search region according to changes in time and consider the impact of water temperature on SAR time to improve SAR efficiency. Ultimately, the optimal path planning for maritime SAR can be achieved to minimize casualties, environmental pollution, and property losses in the ocean.

Author Contributions: Conceptualization, L.L. and Q.S.; methodology, L.L. and Q.S.; software L.L. and Q.S.; validation, L.L. and Q.S., formal analysis, L.L.; investigation, L.L.; resources, Q.X. and Q.S.; data curation, L.L.; writing—original draft preparation, L.L.; writing—review and editing, L.L. and Q.S.; visualization, L.L. and Q.S.; supervision, Q.X. and Q.S.; project administration, Q.X. and Q.S.; funding acquisition, Q.S. and Q.X.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 52371360), the Fundamental Research Funds for Central Universities (Grant No. 3132024118), and the Zhejiang Lab Open Research Project (Grant No. K2022QA0AB03).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We thank the hard-working editors and reviewers for their valuable comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Teng, F.; Ban, Z.X.; Li, T.S.; Sun, Q.Y.; Li, Y.S. A privacy-preserving distributed economic dispatch method for integrated port microgrid and computing power network. *IEEE Trans. Ind. Inform.* **2024**, *in press*. [CrossRef]
2. Sun, Y.; Ling, J.; Chen, X.Q.; Kong, F.C.; Hu, Q.Y.; Biancardo, S. Exploring maritime search and rescue resource allocation via an enhanced particle swarm optimization method. *J. Mar. Sci. Eng.* **2022**, *10*, 906. [CrossRef]
3. Li, J.Q.; Zhang, G.Q.; Jiang, C.Y.; Zhang, W.D. A survey of maritime unmanned search system: Theory, applications and future directions. *Ocean. Eng.* **2023**, *285*, 115359. [CrossRef]
4. Wang, H.L.;Yin, C.Y.; Lu, L.Y.; Wang, D.; Peng, Z.H. Cooperative path following control of UAV and USV cluster for maritime search and rescue. *Chin. J. Ship Res.* **2022**, *17*, 157–165.
5. Gao, S.N; Peng, Z.H.; Liu, L.; Wang, H.L.; Wang, D. Coordinated target tracking by multiple unmanned surface vehicles with communication delays based on a distributed event-triggered extended state observer. *Ocean. Eng.* **2021**, *227*, 108283. [CrossRef]
6. Zhang, H.; Huang, Y.Y; Qin, H.C; Geng, Z. USV search mission planning methodology for lost target rescue on sea. *Electronics* **2023**, *12*, 4584. [CrossRef]
7. Mariyasagayam, M.N.; Menouar, H.; Lenardi, M. GeoNet: A project enabling active safety and IPv6 vehicular applications. In Proceedings of the IEEE International Conference on Vehicular Electronics and Safety, Columbus, OH, USA, 22–24 September 2008; pp. 312–316.
8. Noguchi, S.; Tsukada, M.; Ernst, T.; Inomata, A.; Fujikawa, K. Location-aware service discovery on IPv6 GeoNetworking for VANET. In Proceedings of the 11th IEEE International Conference on ITS Telecommunications, St. Petersburg, Russia, 23–25 August 2011; pp. 224–229.
9. Cai, C.; Chen, J.F.; Saad, A.M.; Liu, F. A task allocation method for multi-AUV search and rescue with possible target area. *J. Mar. Sci. Eng.* **2023**, *11*, 804. [CrossRef]
10. Koopman, B.O. The theory of search: III. The optimum distribution of searching effort. *Oper. Res.* **1957**, *5*, 613–626. [CrossRef]
11. Kratzke, T.M.; Stone, L.D.; Frost, J.R. Search and rescue optimal planning system. In Proceedings of the Information Fusion (FUSION), 2010 13th Conference on IEEE, Edinburgh, UK, 26–29 July 2010.
12. Chen, Z.K,; Liu, H.; Tian, Y.L,; Wang, R.; Xiong, P.S.; Wu, G.H. A particle swarm optimization algorithm based on time-space weight for helicopter maritime search and rescue decision-making. *IEEE Access* **2020**, *8*, 81526–81541. [CrossRef]
13. Ma, M.; Gu, N.; Dong, J.W.; Yin, Y.; Han, B.; Peng, Z.H. Area coverage path planning of multiple ASVs based on ECDIS. *Chin. J. Ship Res.* **2024**, *19*, 211–219.

14. Ma, Y.; Li, B.; Huang, W.T.; Fan, Q.Q. An improved NSGA-II based on multi-task optimization for multi-UAV maritime search and rescue under severe weather. *J. Mar. Sci. Eng.* **2023**, *11*, 781. [CrossRef]
15. Xing, S.W. Research on Global Optimization Model and Simulation of Joint Aeronautical and Maritime Search. Doctoral Dissertation, Dalian Maritime University, Dalian, China, 2012.
16. Zhou, X.; Cheng, L.; Li, W.D.; Zhang, C.R.; Zhang, F.L.; Zeng, F.X.; Yan, Z.J.; Ruan, X.G.; Li, M.C. A comprehensive path planning framework for patrolling marine environment. *Appl. Ocean. Res.* **2020**, *100*, 102155. [CrossRef]
17. Shu, Y.Q.; Zhu, Y.J.; Xu, F.; Gan, L.X.; Li, P.T.W.; Yin, J.C.; Chen, J.H. Path planning for ships assisted by the icebreaker in ice-covered waters in the Northern Sea Route based on optimal control. *Ocean. Eng.* **2023**, *267*, 113182. [CrossRef]
18. Cai, C.; Chen, J.F.; Yan, Q.L.; Liu, F. A multi-robot coverage path planning method for maritime search and rescue using multiple AUVs. *Remote Sens.* **2023**, *15*, 93. [CrossRef]
19. Hayat, S.; Yanmaz, E.; Bettstetter, C.; Brown, T.X. Multi-objective drone path planning for search and rescue with quality-of-service requirements. *Auton. Robots* **2020**, *44*, 1183–1198. [CrossRef]
20. Lv, M.M.; Zhang, Q. Automatic search mode of ship's dynamic sector based on MMG model. *J. Shandong Jiaotong Univ.* **2018**, *26*, 83–88.
21. Hu, X.Y.; Li, X.K.; Zhang, Y.J. Fast filtering of LiDAR point cloud in urban areas based on scan line segmentation and GPU acceleration. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 308–312.
22. Guo, W.L.; Liu, C.; Sun, T.; Cococcioni, M. Cooperative maritime search of multi-ship based on improved robust Line-of-Sight guidance. *J. Mar. Sci. Eng.* **2024**, *12*, 105. [CrossRef]
23. Tan, X.Q.; Han, L.H.; Gong, H.; Wu, Q.W. Biologically inspired complete coverage path planning algorithm based on Q-Learning. *Sensors* **2023**, *23*, 4647. [CrossRef] [PubMed]
24. Tan, C.S.; Mohd-Mokhtar, R.; Arshad, M.R. A comprehensive review of coverage path planning in robotics using classical and heuristic algorithms. *IEEE Access* **2021**, *9*, 119310–119342. [CrossRef]
25. Bartumeus, F.; da Luz, M.G.E.; Viswanathan, G.M.; Catalan, J. Animal search strategies: A quantitative random-walk. *Ecology* **2005**, *86*, 3078–3087. [CrossRef]
26. Wen, J.; Yang, J.C.; Wei, W.; Lv, Z.H. Intelligent multi-AUG ocean data collection scheme in maritime wireless communication network. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 3067–3079. [CrossRef]
27. Wang, X.L.; Yin, Y.; Jing, Q.F. Maritime search path planning method of an unmanned surface vehicle based on an improved bug algorithm. *J. Mar. Sci. Eng.* **2024**, *11*, 2320. [CrossRef]
28. Ma, Y.; Zhao, Y.J.; Li, Z.X.; Yan, X.P.; Bi, H.X.; Krolczyk, G. A new coverage path planning algorithm for unmanned surface mapping vehicle based on A-star based searching. *Appl. Ocean. Res.* **2022**, *123*, 103163. [CrossRef]
29. Chen, Y.L.; Bai, G.Q.; Zhan, Y.; Hu, X.Y.; Liu, J. Path planning and obstacle avoiding of the USV based on improved ACO-APF hybrid algorithm with adaptive early-warning. *IEEE Access* **2021**, *9*, 40728–40742. [CrossRef]
30. Yang, S.X; Luo, C. A neural network approach to complete coverage path planning. *IEEE Trans. Syst. Man, Cybern. Part B* **2004**, *34*, 718–724. [CrossRef]
31. Liu, X.; Zhong, W.Z.; Wang, X.; Duan, H.T.; Fan, Z.X.; Jin, H.W.; Huang, Y.; Lin, Z.P. Deep reinforcement learning-based 3D trajectory planning for cellular connected UAV. *Drones* **2024**, *8*, 199. [CrossRef]
32. Xing, B.W.; Wang, X.; Liu, Z.C. The wide-area coverage path planning strategy for deep-sea mining vehicle cluster based on deep reinforcement learning. *J. Mar. Sci. Eng.* **2024**, *12*, 316. [CrossRef]
33. Teng, F.; Zhang, Y.X.; Yang, T.K.; Li, T.S.; Xiao, Y.; Li, Y.S. Distributed optimal energy management for We-Energy considering operation security *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 225–235.
34. Anaya, J.J.; Talavera, E.; Jimenez, F.; Serradilla, F.; Naranjo, J.E. Vehicle to vehicle GeoNetworking using wireless sensor networks. *Ad Hoc Netw.* **2015**, *27*, 133–146. [CrossRef]
35. Chang, Y.; Han, G.H.; Yan, L.L. Trust evaluation model based on optimal search theory. In Proceedings of the 2010 6th International Conference on Wireless Communications Networking and Mobile Computing, Chengdu, China, 23–25 September 2010.
36. Xiong, P.S.; Liu, H.; Yang, H. Helicopter maritime search area planning based on a minimum bounding rectangle and K-means clustering. *Chin. J. Aeronaut.* **2021**, *34*, 554–562. [CrossRef]
37. IAMSAR. *International Aeronautical and Maritime Search and Rescue Manual II*; Mission Coordination; IMO/International Civil Aviation Organization Publications; ICAO: London, UK; Montreal, QC, Canada, 2022.
38. Zhang, J.J.; Liu, Y.L.; Zhou, W.D. Adaptive sampling path planning for a 3D marine observation platform based on evolutionary deep reinforcement learning. *J. Mar. Sci. Eng.* **2024**, *11*, 2313. [CrossRef]
39. Guo, S.Y.; Zhang, X.G.; Du, Y.Q.; Zheng, Y.S.; Cao, Z.Y. Path planning of coastal ships based on optimized DQN reward function. *J. Mar. Sci. Eng.* **2021**, *9*, 210. [CrossRef]
40. Szwaykowska, K.; Zhang, F. Controlled Lagrangian particle tracking: Error growth under feedback control. *IEEE Trans. Control. Syst. Technol.* **2018**, *26*, 874–889. [CrossRef]

41. Xing, B.W.; Wang, X.; Yang, L.; Liu, Z.C.; Wu, Q.Y. An algorithm of complete coverage path planning for unmanned surface vehicle based on reinforcement learning. *J. Mar. Sci. Eng.* **2023**, *11*, 645. [CrossRef]
42. Shu,Y.Q.; Xiong, C.H.; Zhu, Y.J.; Liu, K.; Liu, R.W.; Xu, F.; Gan, L.X.; Zhang, L. Reference path for ships in ports and waterways based on optimal control. *Ocean. Coast. Manag.* **2024**, *253*, 107168. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.