Expert Systems With Applications 254 (2024) 124388

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A human-like collision avoidance method for USVs based on deep reinforcement learning and velocity obstacle

Xiaofei Yang [a,*], Mengmeng Lou [a], Jiabao Hu [a], Hui Ye [a], Zhiyu Zhu [a], Hao Shen [b], Zhengrong Xiang [c], Bin Zhang [d]

[a] School of Automation, Jiangsu University of Science and Technology, Zhenjiang 212100, China
[b] School of Electrical and Information Engineering, Anhui University of Technology, Maanshan 242032, China
[c] School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China
[d] Department of Electrical Engineering, University of South Carolina, Columbia 29224, USA

ARTICLE INFO

Keywords:
Deep reinforcement learning
Unmanned surface vehicles
Obstacle avoidance
Velocity obstacle
Safety path
COLREGs

ABSTRACT

Collision avoidance is critical in unmanned surface vehicles (USVs) and is especially challenging in scenarios involving ships of different intelligence levels. To address this problem, inspired by risk analysis and collision avoidance habits of seafarers, this paper proposes a human-like collision avoidance method based on deep reinforcement learning (DRL) and velocity obstacle (VO). It first introduces a navigation impact factor (NIF) calculation module based on fuzzy theory to simulate human beings' attention mechanisms when facing multiple ships. To be compatible with human beings' regulations, the convention on the International Regulations for Preventing Collisions at Sea (COLREGs) is first to be integrated into the calculation of the NIF calculation module, and the VO algorithm is incorporated into the reward function, which can reduce collision risks in paths and improve safety requirements. In addition, a series of reward functions are carefully designed to balance safety, smoothness, and driving operation. To validate the performance, experiments are conducted on our virtual simulation platform. The results show that our algorithm can accurately assess the impact of target ships (TS) on own ship (OS) in complex environments and can obey the COLREGs. Collision avoidance can be achieved effectively and the path is smoother.

1. Introduction

In recent years, there has been a huge development in unmanned surface vehicles (USVs) (Wang & Xu, 2020; Yan, Wang, Ma, Liu, &Wang, 2020; Zhang, Liu, Huang, & Zhang, 2022; Yang et al., 2024). The International Maritime Organization (IMO), as an official agency responsible for formulating maritime rules, believes that full autonomy USVs will not come true soon. This indicates that ships with various levels of autonomy, including semi-autonomous, remote control, and manned ships, will coexist within the maritime traffic system. Therefore, the safety issue should focus more on obstacle avoidance from the view of the USVs (Wang, Zhang, Zhang, & Gao, 2023). For example, when encountering other ships at sea, it is critical to make USVs understand human beings' driving patterns and intentions and achieve safe collision avoidance.

Collision avoidance for ships is generally categorized as local path planning (Wang, Zhang, Ahn, & Xu, 2022), which is addressed in many literatures. The artificial potential field (APF) algorithm was initially introduced by Khatib (Khatib, 1986), and was adopted to avoid obstacles in real-time (Liu et al., 2023). The environment is also considered and combined with dynamic virtual ships to avoid collision and track the desired path, which enhances the performance of USVs (Zhang, Han, Li, & Zhang, 2022). However, the issues of local minima and goal unreachability should be considered, especially in multiple obstacles environments. Dynamic window approach (DWA) is a common method for local path planning (Fox, Burgard, & Thrun, 1997). In Guan and Wang (2023), the weight coefficient of the objective function is trained by the deep Q network, which improves the success rate of collision avoidance. The rapidly exploring random tree (RRT) algorithm is effective in path planning in high-dimensional spaces (Chiang & Tapia, 2018). Random sampling is used to explore the search space to eliminate the need for explicitly constructing high-precision maps. It is not

* Corresponding author.
E-mail addresses: yxfei_0809@just.edu.cn (X. Yang), 211110303115@stu.just.edu.cn (M. Lou), 221110303111@stu.just.edu.cn (J. Hu), yehuicc@just.edu.cn (H. Ye), zhuzy@just.edu.cn (Z. Zhu), haoshen@ahut.edu.cn (H. Shen), xiangzr@njust.edu.cn (Z. Xiang), zhangbin@cec.sc.edu (B. Zhang).

https://doi.org/10.1016/j.eswa.2024.124388
Received 13 November 2023; Received in revised form 27 March 2024; Accepted 1 June 2024
Available online 6 June 2024

sensitive to obstacles in the environment and when dynamic obstacles exist, the path generated is longer and non-smooth. Moreover, the velocity obstacle (VO) method can instantly predict potential collision scenarios. It is also a real-time and efficient local path planning method (Fiorini & Shillert, 1998), especially in uncertain environments. However, high computational complexity in the case of multiple obstacles has limited its application (Xie & Dames, 2023).

With the development of intelligent technology, they have been adopted to cope with the increasingly complex environment. Such as the ant colony optimization (ACO) and DWA algorithm are combined to solve the path planning and collision avoidance of unmanned sailboats (Shen, Ding, Liu, & Yu, 2022). ACO is a typical swarm intelligence algorithm. A path-searching-based method is proposed, which generates trajectories that comply with International Regulations for Preventing Collisions at Sea (COLREGs) based on the bell shape of normal distribution (Wang, Yu, Liang, & Li, 2018). However, its effectiveness in complex environments has not been verified. A collision avoidance method based on fuzzy logic and DWA is proposed, fuzzy controllers are designed to adjust the weight parameters in real-time, which can enhance the robustness of the algorithm (Yao et al., 2024). Ahn et al. (2012) have designed a collision avoidance system based on a fuzzy inference system and an expert system, and the neural networks are used to calculate the risk of collisions. Xie et al. (2019) have proposed a model predictive method that uses neural networks to achieve collision avoidance in complex situations. However, its performance depends on the established model and environmental characteristics, and generality is its weakness. As such, the deep reinforcement learning (DRL) method draws more and more attention (Zhang, Cheng, Lin, Zhang, & Xie, 2023). It has been applied for cargo ships, and relevant reward functions are designed based on obstacles, non-navigation zones, and distance to train the navigation strategies (Chen, Chen, Ma, Zeng, & Wang, 2019). However, the rationality of the planned path has not been considered, such as the smoothness and length of the path. In addition, the study overlooked the moving ships within the navigation area. For dynamic and static obstacles, an improved deep deterministic policy gradient (DDPG) algorithm is proposed to form obstacle avoidance strategies (Zhou, Wang, Wang, & He,2022). The neural network and experience pool are optimized to accelerate the learning process. Unfortunately, the problem of collision avoidance in complex dynamic scenarios is not considered, and the COLREGs are not considered a constraint on decision-making, leading to some risks when interacting with other ships (Cho, Han, & Kim, 2022). In Woo and Kim (2020), the visual recognition capabilities of deep neural networks are utilized to perceive the environment, and a DRL based on a semi-Markov decision process model is designed to avoid collisions. Its performance is only verified in a simple environment. It is worth noting that the combination of DRL with traditional algorithms is a new trend. For instance, DRL and the reciprocal velocity obstacle (RVO) algorithm were combined to address collision avoidance (Xue, Wu, Yamashita, & Li, 2023). In complex environments, it's beneficial to assess the impact of incoming ships for safe navigation. Unfortunately, it's not discussed in the aforementioned research. The closest point of approach (CPA) is the common method used to calculate the impact of incoming vessels (Sun, Wang, Fan, & Mu, 2023). Although simple, it is not sufficient for comprehensive evaluation. Chun et al. (2021) have incorporated the ship's domain into the CPA to enhance the accuracy of the evaluation. To enhance the comprehensiveness, fuzzy logic has been further considered in the risk assessment. Human experience is translated into mathematical models by fuzzy logic. However, they tend to evaluate the impact of incoming ships based on objective factors such as speed and distance, and the impact of seafarers' attention and maritime rules are ignored (Bukhari, Tusseyeva, Lee, & Kim, 2013; Zheng, Xie, & Yuan, 2023).

Although intelligent algorithms are quite extensive, there are still some shortcomings. Firstly, some models are constrained by predefined rules and are overly dependent on specific situations, leading to machine decision-making that resorts to simplistic "brute force" methods (French & Douglas, 1991). Secondly, traditional artificial intelligence methods often lack a deeper comprehension of encounter situations, which may result in incorrect decisions such as in the risk assessment and interaction with other vessels. Before the emergence of intelligent ships, humans, as the primary decision-makers, had amassed extensive practical experience in navigation. Therefore, integrating the experience of seafarers with intelligent algorithms can facilitate decision-making that is more grounded in reality (Li, Cao, Shi, Bai, & Chen, 2022). An intelligent collision avoidance model inspired by human thinking patterns of fast and slow is proposed (Li, Peng, & Zheng, 2023). Nevertheless, this approach conducts risk assessments without clarifying collision avoidance responsibilities, thus yielding inaccurate risk assessments. Similarly, the human attention mechanism has been incorporated into DRL to tackle the challenge of multi-ship collision avoidance (Jiang, An, Zhang, Wang, & Wang, 2022). However, this method fails to account for the impact of COLREGs. Additionally, it disregards the potential risks associated with collision avoidance actions, which is inconsistent with typical collision avoidance practices. Planning local safety paths to avoid obstacles for USVs, especially in the hybrid environment with various autonomy levels of ships, remains an open and challenging problem. Motivated by the aforementioned literature, this paper proposes a safe human-like collision avoidance method based on deep reinforcement learning and the velocity obstacle method. The main contributions of the paper are as follows,

1) To adapt to the hybrid environment with various autonomy levels of ships, especially manned ships, a human-like collision avoidance method is proposed by simulating the analysis and avoidance habits of seafarers. To analyze the impact of TSs on OS comprehensively, in addition to objective factors of navigation, COLREGs that are regarded as a consensus among seafarers are innovatively taken into consideration. To estimate and avoid potential risks like human beings, the DRL is the mainframe and VO is introduced into the reward function design to facilitate the USV collision avoidance with the manned ships.

2) To balance safety and rationality, a more practical reward function combination is designed by dynamic obstacle avoidance reward based on VO and path optimization reward. The VO-based reward encourages safer rather than risky operations. The path optimization reward aims to optimize the rationality of the path, including smoothness and compliance with COLREGs. The practical application value of the decision-making model is enhanced by addressing the sparse reward and considering the practical navigation needs.

3) To assess the threat of multiple ships and sequentially avoid collision at sea, the navigation impact factor (NIF) is introduced to construct the state space. It is calculated based on fuzzy theory. The COLREGs are first integrated into NIF to assess the TS' threat from the view of seafarers, and incorporated into our training model to enhance the reliability of navigation decisions. In addition, a series of experiments are conducted to verify the performance and advantage of our algorithm.

The rest of the study is organized as follows. Section 2 presents the problem statement. Section 3 fully develops the proposed method based on DRL and VO. Section 4 provides experimental verification and comparison studies. Section 5 gives the concluding remarks and future works.

2. Problem statement

2.1. Description of the problem

Environmental perception and collision avoidance are two major aspects affecting navigation safety. In the scene of multiple ships, the primary focus should be on assessing the impact of TS on the OS. However, it is often overlooked in existing works, which can lead to inappropriate collision avoidance actions. Rules 13–16 of COLREGs respectively stipulate the encounter scenarios of ships and the division of avoidance responsibilities. Seafarers usually conduct environmental assessments and collision avoidance based on their responsibilities. (He et al., 2017; Zhao & Roh, 2019). However, navigation risk assessment often pays more attention to objective factors such as speed and distance, while neglecting the influence of human factors represented by COLREGs. This may result in difficulty in distinguishing the most threatening obstacles under the constraints of maritime rules (Namgung & Kim, 2021). Moreover, a higher success rate in collision avoidance does not necessarily equate to enhanced safety. Some studies with successful, collision avoidance show that the actions taken are sometimes risky, which is not in line with the decision-making tendencies of human beings (Xu, Cai, Cao, Chu, Zhu, & Zhang, 2023; Kuwata, Wolf, Zarzhitsky, & Huntsberger, 2014).

In today's hybrid traffic systems with ships of various autonomy levels, it is a trend for USVs to make human-like decisions. Generally, seafarers excel in environmental awareness and collision avoidance decision-making based on practical experience. The adoption of human-like operations in USVs offers the potential to make more friendly interactions with manned ships and lead to improved navigation safety. To this end, this paper studies a human-like collision avoidance method, which has a primary emphasis on safety and adherence to maritime regulations, rather than simply achieving a high success rate in collision avoidance. The problem is converted into a multi-constraint optimization problem, including safety constraints and rationality constraints. Safety constraints are established to avert the execution of risky actions. Rationality constraints prioritize the smoothness of the path and the promotion of driving compliance, aiming to prevent unreasonably large turns and unlawful interactions.

For navigation in complex dynamic environments, USVs need to make decisions to reach destinations without collision based on the navigation state and COLREGs. The navigation task is translated into a Markov decision process defined by tuples $< S, A, T, R, \gamma >$. S is the state space, which includes the state of the USV, the state of obstacles, and the navigation impact factors. Especially, the calculation of navigation impact factors focuses on COLREGs. A represents the action space composed of the thrust and rudder torque of the unmanned boat. T represents the transition of the state, that is, the next state is determined by the current state s and action a. R represents a reward function used to evaluate the immediate reward obtained by executing a in s, which is related to the safety constraints and rationality constraints. γ is a discount factor. The goal of this task is to obtain a strategy π^* that maximizes the cumulative reward $V_\pi(s)$ obtained in long-term operation, π^* and $V_\pi(s)$ are defined as follows:

$$\begin{aligned} \pi^* &= \underset{\pi}{\operatorname{argmax}} V_\pi(s), \\ V_\pi(s) &= E_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | S_t = s\right] \\ &= \sum_{a \in A} \pi(a|s) Q_\pi(s,a) \end{aligned} \tag{1}$$

where $Q_\pi(s,a)$ is the action-value function.

2.2. COLREGs

COLREGs are mandatory regulations ruled by IMO to reduce ship collisions. Rules 13–15 of COLREGs specify encounter situations and assign collision avoidance responsibilities, as displayed in Fig. 1(a). The vessel which has the responsibility to give way must maintain a proper lookout at all times and take early action to keep clear of the other vessels stipulated by Rule 16 of COLREGs. COLREGs have been extensively explained by Li et al. (2021), and their definitions are adopted in the paper. For convenience, we will consider Rules 13–15, as follows:

(a) Overtaking

When OS is positioned at the port or starboard of TS within $[112.5^\circ, 180^\circ]$ and there is a tendency to overtake TS, the TS is a stand-on vessel and the OS should give way to TS, as displayed in Fig. 1(b).

(b) Head-on

When TS is positioned at the port or starboard of OS within $[0^\circ, 5^\circ]$ and there is a risk of collision, both vessels should turn to starboard and pass each other port-to-port, as displayed in Fig. 1(c).

(c) Crossing give-way

When TS is positioned on the starboard side of OS within $[5^\circ, 112.5^\circ]$, and there is a risk of collision, the OS should give way to TS and turn to the starboard to pass on the port side of TS, as displayed in Fig. 1(d). The TS is a stand-on vessel.

(d) Crossing stand-on

When TS is positioned on the port side of OS within $[5^\circ, 112.5^\circ]$, and there is a risk of collision, the TS should give way to OS and turn to starboard to pass on the port side of OS, as displayed in Fig. 1(e). The OS is a stand-on vessel.

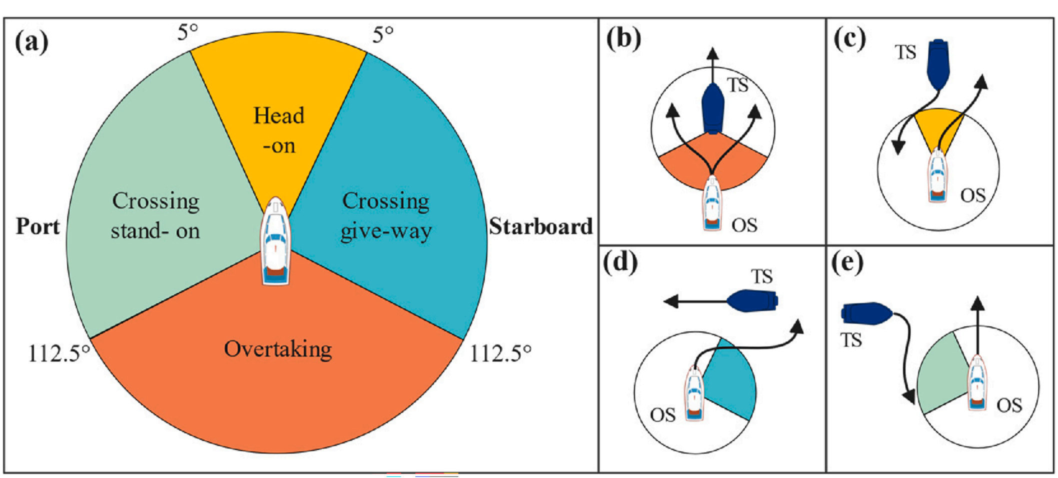


Fig. 1. (a) Collision avoidance action zones; (b) Overtaking; (c) Head-on; (d)Crossing give-way; (e) Crossing stand-on.

3. A human-like collision avoidance algorithm based on DRL and VO

As mentioned in Section 2, the safety collision avoidance of USVs can be converted into a multi-constraint optimization problem. Such a complex problem, although challenging for classical optimization methods, can be structured as a Markov decision process (MDP) that is suitable for the DRL method (Yang, Shi, Liu, Hui, Zhong, & Xiang, 2022). In this research, the DRL is adopted as the central mainframe, which incorporates elements, including the impact assessment of the TS on OS and the COLREGs, to form our human-like method. The paper uses Dueling DQN (Gao, Kang, Zhang, Liu, & Zhao, 2022), and the overall framework of the method is shown in Fig. 2.

3.1. The overall framework of the method

As shown in Fig. 2, the method starts by extracting to construct a state space. Since the TS and OS are ruled by COLREGs, it is difficult to directly use objective factors such as speed and heading to train the effective collision avoidance strategy, without consideration of the constraints of COLREGs. For instance, the seafarers will allocate avoidance responsibilities under the COLREGs and evaluate the risk of the incoming vessels based on their responsibilities. They further determine the priority for evasion based on the risk of the incoming vessels. However, traditional methods lack such navigational expertise, potentially leading to confusion in the avoidance relationship. Therefore, such a human-like decision-making mechanism is introduced, and an NIF calculation module is designed to assess the impact of TS on OS. The NIFs are integrated into the state space of the DRL with the navigation information. Once the state information is obtained, the USV takes appropriate collision avoidance actions. Due to the limit of the Dueling DQN and the requirement for flexibility, the action space must be discretized for feasible implementation. In the phase of reward function design, the VO algorithm is integrated into the dynamic obstacle avoidance reward function. We aim to predict risks like humans and avoid potential risks. The combination of DRL and the VO algorithm enables the design of a human-like collision avoidance strategy with lower risk. To make the decisions more reasonable, the smoothness and the legality of collision avoidance are also taken into consideration.

3.2. State space with NIF

3.2.1. Design of the State Space

In the proposed algorithm, the actions taken by the USV directly depend on the state space. To ensure safety, it is necessary to obtain obstacle information. Therefore, the state space consists of the state of OS and TS, respectively represented as S_O and S_T. The 2D spatial coordinates and heading angles are the basic information in S_O and S_T. To enhance the accuracy of the assessment of the incoming ship and the reliability of the model, the practical experience of the seafarer has been incorporated, and the human-like components are incorporated into the state space. So, the NIF calculation module is adopted to simulate the cognitive assessment of human beings. The scores are used to indicate the impact of the TS on OS. It will be described in detail in the next section.

The coordinates of OS and TS are presented as (x_O,y_O) and (x_T,y_T), and the heading angles are φ_O and φ_T, respectively. NIF is the navigation impact factor. To improve the stability, the linear normalization method is used to normalize the state of OS and TS, as follows:

$$\begin{cases} o_1 = \frac{x_O}{h}, x_O \in [0, h] \\ o_2 = \frac{y_O}{w}, y_O \in [0, w] \\ o_3 = \frac{\varphi_O}{\pi}, \varphi_O \in [0, \pi] \\ o_4 = \frac{x_T}{h}, x_T \in [0, h] \\ o_5 = \frac{y_T}{w}, y_T \in [0, w] \\ o_6 = \frac{\varphi_T}{\pi}, \varphi_T \in [0, \pi] \end{cases} \tag{2}$$

where h and w are the length and width of the navigation environment. The state of OS S_O is (o_1, o_2, o_3). The normalized state of TSs S_T is (o_4^1, o_5^1, o_6^1, NIF^1, …, o_4^n, o_5^n, o_6^n, NIF^n), where n is the total number of TSs. Hence, the state space can be represented as $S = (S_O, S_T)$.

3.2.2. NIF calculation module

When analyzing the impact of the incoming ship, the seafarers often incorporate their understanding of maritime regulations, rather than just focusing on objective factors such as ship speed. To simulate the impact assessment of human beings and obtain more accurate features, the factors considered include distance, collision risk, and the situation compliant with COLREGs. Since the reasoning process of humans is relatively vague, the fuzzy theory is adopted to design the module. The details of the factors are discussed as follows:

Distance Dv is the most intuitive factor for human beings to assess the possibility of collision. Generally, the smaller the distance, the greater the impact of TS on the OS, and the maneuverability of the OS should be restricted further. Five linguistic values for fuzzy are given. *DVN* is very near, *DN* is quite near, *DM* is near, *DF* is far, and *DVF* is very far. The membership functions for Dv are shown in Fig. 3(a).

Collision risk Cr is the possibility of a collision between two ships. This paper uses the method proposed in Huang and Van Gelder (2020). It

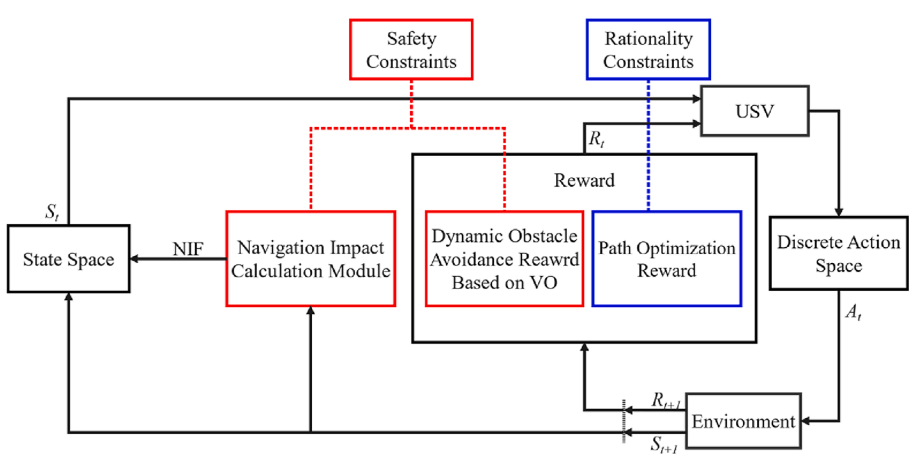


Fig. 2. Mainframe of the human-like method based on DRL and VO.

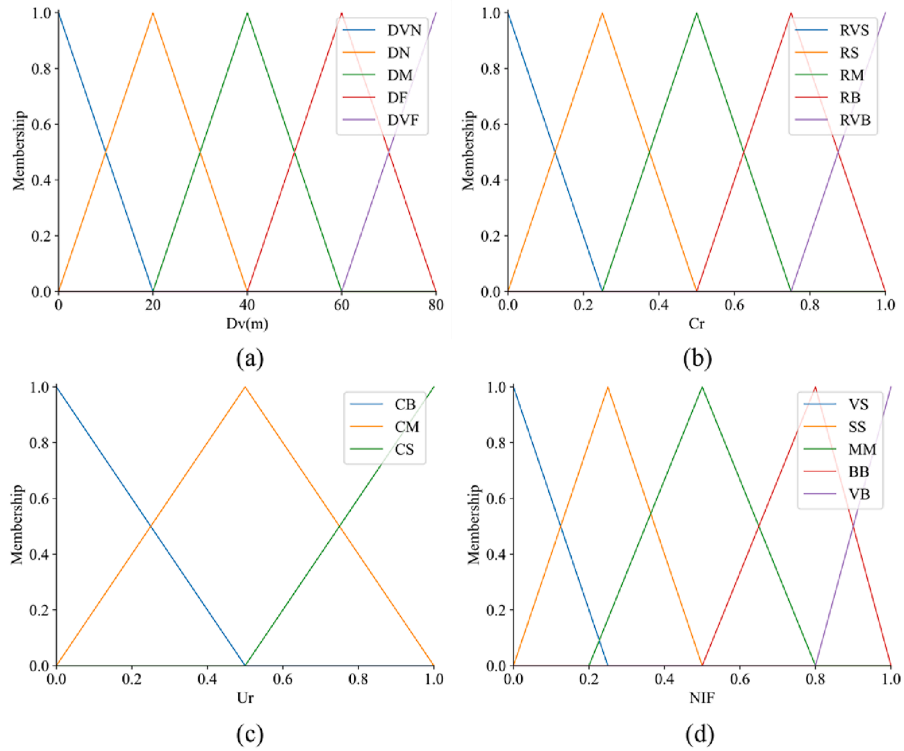


Fig. 3. (a) Membership of distance; (b) Membership of collision risk; (c) Membership of urgency in collision; (d) Membership of NIF.

considers the risk of the TSs and the capability of the OS to prevent collisions, and can effectively calculate *Cr*. To prevent redundancy, a safe distance metric δ is set. *Cr* is calculated only when the distance is less than the δ. To quantify the collision risk, a set of fuzzy linguistic values is defined. *RVS* is very low risk, *RS* is low risk, *RM* is moderate risk, *RB* is high risk, and *RVB* is very high risk. The membership functions for *Cr* are illustrated in Fig. 3(b).

COLREGs are the regulations of the collision avoidance of maritime vessels and have a significant impact on the actions of seafarers. According to it, when TSs and OS form situations of overtaking, head-on, or crossing giving-way, OS is obligated to take actions to avoid collision. Intuitively, the more urgent the situation, the greater the impact on the OS. Bearing angle ψ of OS relative to TS can facilitate the human beings to assess the urgency, and the value of ψ is $[0, \pi]$. With this, the urgency *Ur* can be defined as:

$$Ur = \begin{cases} 1 - \dfrac{\psi}{\pi}, if \text{ According to COLREGs, OS gives way} \\ 0, else \end{cases} \tag{3}$$

In general, a smaller bearing angle means the TS and the OS are closer to each other's course, making them more prone to collisions. The influence of *Ur* is disregarded when the encounter does not conform to the avoidance situation specified in COLREGs. A set of fuzzy linguistic values is defined. *CS* is a low urgency, *CM* is a moderate urgency, and *CB* is a high urgency. The membership functions are illustrated in Fig. 3(c).

For the fuzzy inference to calculate NIF, the inputs are distance, collision risk, and urgency. The higher the NIF, the greater the impact on the navigation of the OS. The fuzzy linguistic values of NIF are defined as shown in Fig. 3(d). VS is the very small impact, *SS* is the small impact, *MM* is the moderate impact, *BB* is large, and *VB* is the very large impact.

In this work, a multi-input single-output fuzzy system is defined. In addition, the fuzzy rules are designed, as shown in Table 1. The Mamdani fuzzy inference method is adopted to obtain the fuzzy set of outputs based on the fuzzy rules (Mamdani & Assilian, 1999).

The output of fuzzy inference is NIF, which is a specific value. The centroid method is employed for defuzzification (Wu, Yip, Yan, & Soares, 2019), as follows:

$$\text{NIF} = \frac{\int I_{NIF}\mu_{I_{NIF}}(I_{NIF})d(I_{NIF})}{\int \mu_{I_{NIF}}(I_{NIF})d(I_{NIF})} \tag{4}$$

where $\mu_{I_{NIF}}(I_{NIF})$ is the membership function, I_{NIF} is input. The centre of the area surrounded by $\mu_{I_{NIF}}(I_{NIF})$ and the abscissa is calculated, which is the NIF.

Table 1
Fuzzy inference rule base.

Ur	*Dv*	*Cr*				
		RVS	RS	RM	RB	RVB
CS	**DVN**	VS	VS	MM	BB	BB
	DN	VS	VS	SS	MM	BB
	DM	VS	VS	VS	MM	MM
	DF	VS	VS	VS	SS	MM
	DVF	VS	VS	VS	SS	MM
CM	**DVN**	SS	SS	MM	BB	VB
	DN	VS	SS	MM	BB	BB
	DM	VS	VS	SS	MM	MM
	DF	VS	VS	SS	SS	MM
	DVF	VS	VS	VS	VS	SS
CB	**DVN**	SS	MM	BB	VB	VB
	DN	SS	MM	MM	BB	VB
	DM	VS	SS	MM	MM	BB
	DF	VS	VS	SS	MM	MM
	DVF	VS	VS	VS	SS	SS

3.3. Design of the action space

It is hard to cope with the continuous action space by Dueling DQN algorithm (Zhan & Zeng, 2022). To expedite training speed, the action space of the USV is discretized. The dynamic models of USVs with the three Degrees of Freedom (DOFs) is commonly used (Yan et al., 2024). In this paper, the models of the USV in Yang et al. (2022) are completely adopted without modification. For underactuated USVs, to reduce the complexity, only thrust τ_u and rudder torque τ_r are considered. For simplicity, τ_u is set as a fixed value, while τ_r can change the yaw of the ship, thus directly affecting the change of the heading angle. Since collision avoidance is our primary focus, τ_r is mainly manipulated to change the heading angle. The size of the action space has been limited to 11, and actions vary uniformly in both magnitude and direction. After discretization, the action space can meet the requirement for flexibility of USV and offers a higher level of versatility. The discretized action space is designed as follows:

$$A = \{-5\Delta\tau_r, -4\Delta\tau_r, ..., 5\Delta\tau_r\} \tag{5}$$

where $\Delta\tau_r > 0$ and is a fixed value.

3.4. Design of reward functions

In the field of local path planning, the objects include not only reaching the destination, but also ensuring obstacle avoidance, optimizing the smoothness of the navigation path, and adhering to the regulations. To this end, dynamic avoidance rewards and path optimization rewards are designed, corresponding to the safety and rationality of the collision avoidance path, respectively. The dynamic obstacle avoidance reward is based on the VO, which is defined as r_{VO}. Path optimization rewards consist of target reward r_g, boundary reward r_c, dynamic approach reward r_{da}, static approach reward r_{sa}, turning reward r_t, and COLREGs reward r_{CLGs}.

Therefore, the total reward function R is divided into dynamic obstacle avoidance rewards based on the VO and path optimization rewards, which can be represented as follows:

$$R = r_{VO} + r_g + r_c + r_{da} + r_{sa} + r_t + r_{CLGs} \tag{6}$$

3.4.1. Dynamic obstacle avoidance reward based on VO

Although some specific actions can achieve collision avoidance, they may be at risk. Such actions should be avoided in practice. Therefore, it is a crucial task to eliminate risky actions to ensure safety. The classic VO algorithm can determine whether the velocity selected poses a collision risk and is very flexible (Xu et al., 2021). It is suitable for identifying actions in the action space that may pose risks. To reduce the collision risk, it is introduced into the obstacle avoidance reward function without modification. As shown in Fig. 4, $VO_{O|T}$ is represented as velocity obstacle space and v_o^i are the velocities induced by the discrete action space of the OS. Among them, those velocities that intersect with $VO_{O|T}$ are marked in red. They pose a collision risk and should be avoided. Conversely, those marked in black indicate no collision risk. In the case of multiple-ship encounters, various velocity obstacles should be constructed with different TSs separately, denoted as $VO_{O|T} = VO_{O|T1} \cup VO_{O|T2}$. Similarly, actions corresponding to velocities marked in black are considered optional and safe. The design of the r_{VO} is given as follows:

$$r_{VO} = \begin{cases} 0, if\, v_O outsides VO_{O|T} \text{ and } d_{OT} > \vartheta \\ \lambda_{VO1}, if\, v_O within VO_{O|T} \text{ and } d_{OT} > \vartheta \\ \lambda_{VO2}, if\, d_{OT} \leq \vartheta \end{cases} \tag{7}$$

where $d_{OT} = \sqrt{(x_O - x_T)^2 + (y_O - y_T)^2}$ is the distance between OS and TS, ϑ is the minimum distance at which a collision occurs, $\lambda_{VO1} < 0$, and $\lambda_{VO2} < 0$. v_o is the velocity of the OS. If it is not within $VO_{O|T}$ and the distance is greater than ϑ, there is no collision risk and no penalty is applied to the action. If the action causes v_o to fall into $VO_{O|T}$, and the distance is greater than ϑ, it indicates that the action may lead to a collision, but have not yet. In this case, the action has a potential risk, and a penalty λ_{VO1} is assigned. When the distance is less than ϑ, it means that they have collided, a penalty λ_{VO2} is assigned. Thus, actions involving collision risk are subject to penalties, which enhance the overall safety of navigation.

3.4.2. Path optimization reward

1. Target reward

Reaching the target is important for USV path planning. To train the USV to target, the target reward r_g is designed. Usually, when the distance is within a specific range, the task is considered to be completed. Therefore, this reward is related to distance, as follows:

$$r_g \stackrel{def}{=} \begin{cases} \lambda_g, if\, d_{OG} < d_{og} \\ 0, else \end{cases} \tag{8}$$

where $d_{OG} = \sqrt{(x_O - x_G)^2 + (y_O - y_G)^2}$ is the distance between the USV and the target, (x_G, y_G) is the target position, and $\lambda_g > 0$. d_{og} is a threshold, and if d_{OG} is less than this threshold, it is considered that the target has been reached. Upon successfully reaching the target, a reward λ_g is returned. Otherwise, a reward of 0 is returned.

2. Boundary reward

To prevent unrestricted movement of USVs, it is necessary to set boundaries in the training environment. To prevent collisions with the environmental boundaries, the boundary reward r_c is designed. Once

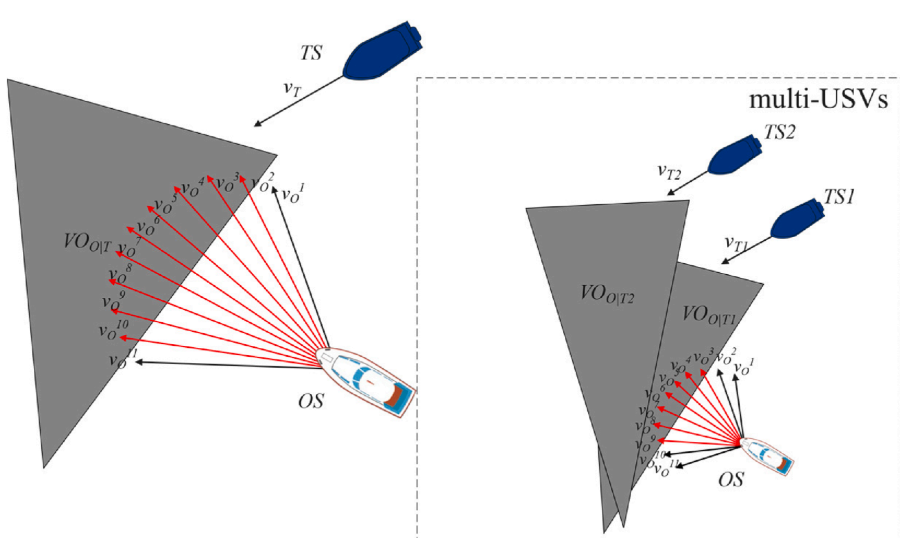


Fig. 4. Dynamic obstacle avoidance reward based on VO.

USV exceeds the range of the environment, a penalty is given. r_c is shown as (9).

$$r_c \stackrel{def}{=} \begin{cases} 0, & if(x_{\min} < x_O < x_{\max}) and (y_{\min} < y_O < y_{\max}) \\ \lambda_c, & else \end{cases} \tag{9}$$

where λ_c is a penalty, x_{min} and x_{max} are the minimum and maximum values of the horizontal boundary, y_{min} and y_{max} are the minimum and maximum values of the vertical boundary.

3 Dynamic approach reward

To quickly approach the target point with the correct posture, a dynamic approach reward r_{da} is designed. The USV should not only approach the target point but also maintain consistency with the expected heading, otherwise, it will cause an imbalance in decision-making. Therefore, r_{da} is designed based on the angle and distance, as follows:

$$r_{da} \stackrel{def}{=} k_1 (d_{pre} - d_{OG}) - \lambda_1 \varphi_d \tag{10}$$

(x_O^{bf}, y_O^{bf}) is coordinate of the USV before executing an action, $d_{pre} = \sqrt{(x_O^{bf} - x_G)^2 + (y_O^{bf} - y_G)^2}$ is its distance to the target, $\varphi_d = |\varphi_O - \varphi_G|$ is the absolute angle error between the actual heading angle φ_O and the ideal heading angle φ_G. k_1 and λ_1 are positive coefficients. After executing the action, if the distance between the USV and the target reduces, and the absolute angle error becomes smaller, it returns a greater reward. Since the reward is closely related to the state changes of the USV, it is considered a dynamic reward.

4 Static approach reward

To encourage the USV to approach the target, a static approach reward r_{sa} is designed. The reward is structured such that it decreases as the distance between the USV and the target grows larger, and it increases as the distance gets smaller. The reward is static as it is solely determined by the distance, independent of the state changes of the USVs. r_{sa} is designed as follows:

$$r_{sa} \stackrel{def}{=} \lambda_{sa} d_{OG} \tag{11}$$

where $\lambda_{sa} < 0$.

5 Turning reward

For USVs, excessive or large turning angles can result in unnecessary energy consumption or even capsizing. To mitigate this issue and get a more efficient turning range, a turning reward is designed as follows:

$$r_t \stackrel{def}{=} -k_t \left| \frac{\tau_r}{\Delta \tau_r} \right| \tag{12}$$

where τ_r is the rudder moment, and k_t is a positive. When $\tau_r = 0$, it means that the USV maintains the original heading without any turning, in such case, the reward is 0. When the τ_r is non-zero, it indicates a turn is executed. A larger value of τ_r corresponds to a wider turn, which results in a large punishment.

6 COLREGs reward

Rules 13–16 of COLREGs are mandatory rules. To ensure better interaction between USVs and manned ships and enhance driving compliance, a COLREGs reward is designed as follows:

$$r_{CLGs} \stackrel{def}{=} \begin{cases} 0, & ifobeyCOLREGs \\ \lambda_{CLGs}, & ifdisobeyCOLREGs \end{cases} \tag{13}$$

where $\lambda_{CLGs} < 0$. According to the COLREGs, when the OS is a giving-way ship, it is required to steer to the starboard to avoid TS. When OS complies with COLREGs, no penalty is imposed on OS. Any other actions are considered as a violation and will result in a negative reward.

3.5. The proposed human-like collision avoidance algorithm

The proposed human-like safe collision avoidance algorithm in a hybrid environment is illustrated in Fig. 5, which mainly consists of the dual network, the experience replay pool, the environmental platform, and the navigation impact factor calculation module. The dual networks receive state s and output corresponding action a or the strategy π. The experience replay pool is used to store the experience data generated by the interaction between USV and the environment. The environmental platform is used to provide states, receive action a, and return reward r. The navigation impact factor calculation module is used to calculate the navigation impact of surrounding obstacles on the USV.

The process is described as follows. The dual networks are given and they have the same structure. The training parameters are given, including the number of training sets, learning rate, and attenuation coefficient. The neural networks are initialized (line 3).

The second phase is environmental interaction (lines 5–9). At the beginning of each episode, the environment is initialized, and the NIF is obtained by fuzzy inference. The navigation state along with NIF forms the current state (line 5). The current state is input into the current value network, which calculates the Q-value corresponding to the action (line 7). Following the ε-greedy strategy, an action, denoted as a, is selected (line 8). The USV agent executes the action, interacts with the environment, and scores the dynamic obstacle avoidance process based on the velocity obstacle reward function. Turning reward and COLREGs reward are used as path optimization reward functions. Finally, the total reward r is determined, and the next state is obtained (line 9).

The third phase involves the neural networks update (lines 10–15). This process operates as follows: experiences gathered during the interaction are stored in the experience replay buffer (line 10). Each time, a random subset of data is selected from the buffer for training (line 12). Backpropagation is used to minimize the loss function, which updates the parameter θ of the current value network (line 14). Finally, the parameter θ' of the target network is updated periodically (line 15).

Algorithm 1 shows the pseudo-code of the proposed algorithm.

Algorithm 1 Human-like safe local path planning algorithm

1. **Input:** Number of episodes *M*; Maximum number of steps P; State Characteristics Dimension *n*; Action set *A*; Learning Rate α; Attenuation factor γ; Experience Replay buffer maximum size *D*.
2. **Output:** Target Value Network *Q'*.
3. Initialize the parameter θ of the current value network *Q*; Initialize the parameters $\theta' = \theta$ of the target value network *Q'*; Initialize replay buffer *R*.
4. **for** i = 1, 2... *M* **do**
5. Initialize the environment and obtain state *S*, which is composed of navigation impact factors and environmental data.
6. **for** t = 1, 2...*P* **do**
7. Use *s* as input in the current value network *Q*, the *Q*-value output corresponding to all actions of action set *A* is obtained.
8. Select action *a* from the *Q*-value output with the ε-greedy strategy.
9. The USV performs action *a* and interacts with the environment to get reward ***r*** and new state *s'*.
10. Store (*s*, *a*, *r*, *s'*) in the experience replay buffer.
11. $s = s'$.
12. Sample mini-batch of *m* transition (s_j, a_j, r_j, s_j') from *R*, calculating the target *Q* value y_j.
13. $y_j = \begin{cases} r_j & if\, s' \,is \text{ terminal} \\ r_j + \gamma Q'(s', \underset{a'}{\operatorname{argmax}} Q(s', a; \theta); \theta'), & otherwise \end{cases}$
14. Using the loss function $\frac{1}{m}\sum_{j=1}^{m}(y_j - Q(s_j, a_j, \theta))^2$, update the parameter θ of the current value network *Q* by the gradient direction reverse propagation of the neural network.
15. Replace target parameters every *C* steps: $\theta' \leftarrow \theta$
16. **if** *s'* is the terminal state, **do**
17. **break**
18. **end if**
19. **end for**
20. **end for**

4. Simulation and comparison

To validate the performance of our algorithm, a series of experiments are conducted. First, the simulation environment and training parameters are defined. Next, the proposed navigation impact assessment

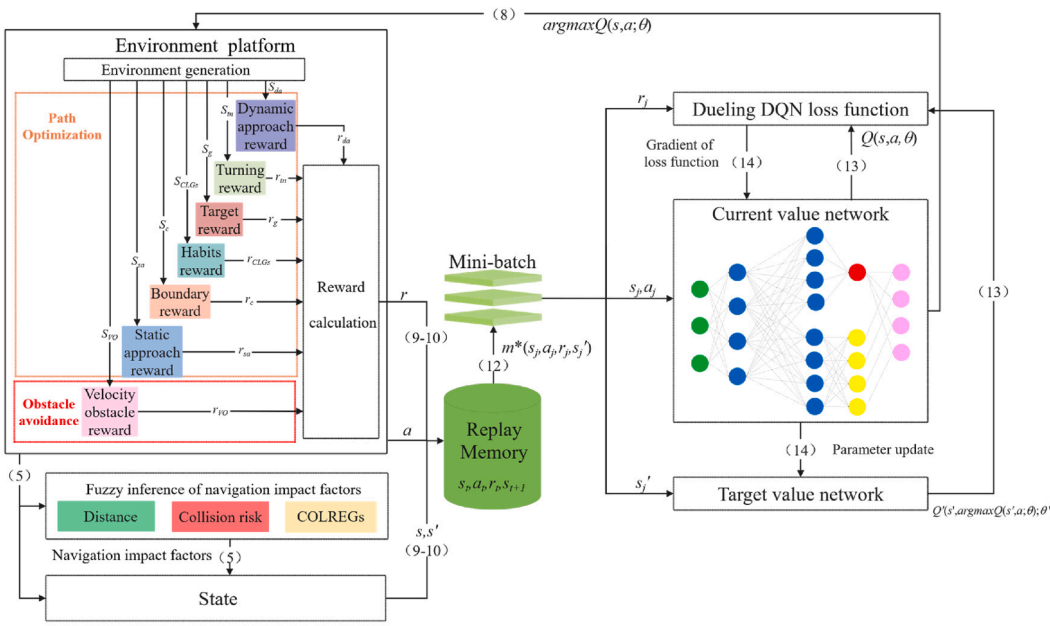


Fig. 5. The framework of the proposed algorithm.

method is compared with others to demonstrate its advantages. Typical scenes involving two ships' encounters are designed to verify compliance with the COLREGs. Subsequently, more experiments are constructed to verify its feasibility in a complex maritime environment.

4.1. Simulation environment and parameters

The computer is an Intel Core i5-9400 CPU @ 2.90 GHz 2.90 GHz, NVIDIA GeForce GTX 1660 Ti 6 GB, and 24 GB DDR4 memory. The simulation is hosted on our virtual simulation platform designed by Unity3D. Our algorithm is implemented in Python and TensorFlow. The collision avoidance range of vessels is usually restricted to local waters, and a rectangular environment of 40 m by 90 m is established. The navigation range of the USV and other ships is limited in the area, and the boundaries are considered static obstacles. The starting and target points are set on opposite sides of the area, and dynamic ships are located between them. Table 2 shows the key relevant parameters of this setup. Without loss of generality, the radius of ships is all set as 1 m. Collision avoidance is performed when the distance is less than the safety distance, which is set as 7 m. The collision distance is set as 2 m. Whenever a collision occurs or a ship goes beyond the boundaries, the current round ends, the environment is reset, and training proceeds to the next round.

4.2. Performance of navigation impact calculation methods

To verify the performance and advantage of the proposed human-like NIF calculation method, a comparative test is conducted with the methods in Jiang et al. (2022) and Shi, Zhen, and Liu (2022), which are named Method 1 and Method 2, respectively. It is crucial to use the NIF calculation method to judge the impact of incoming ships based on the COLREGs, just like seafarers do. To verify it, the complexity of ship encounters in real ports or waterways is set and simulated. In this experiment, the OS is restricted to sail in a straight line, while TS1 and TS2 met the OS from both sides, and different encounter scenarios are created, including Crossing stand-on and Crossing give-way.

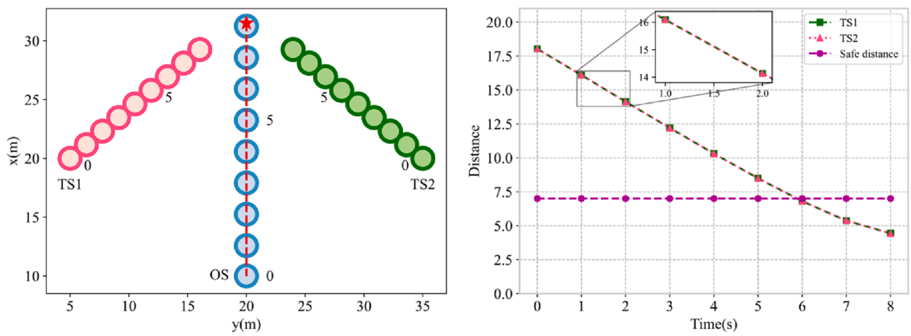
In the test, the OS has a starting point of (10, 20) and an initial heading angle of 0°, and it navigates toward the destination at a speed of 2.6 m/s. TS1 starts from (20, 5), and TS2 starts from (20, 35). They approach the OS at the same speed of 1.8 m/s with initial heading angles 45° and −45°, respectively. Fig. 6(a) shows their trajectories, and Fig. 6(b) shows the changes in distances between TS and OS. It can be seen that one line is for metric of safe distance and the two lines for TS1 and TS2 almost overlap. From 0 s to 5 s, the OS is not a giving-way ship for TS1, and the impact of TS1 on it is only related to the distance. As the distance decreases, the NIF of TS1 gradually increases. However, it is a giving-way relationship with TS2, the bearing angle of the OS relative to TS2 gradually decreases, and the NIF of TS2 gradually increases. At 6 s, the distances between OS and TS1, as well as between TS2 and OS are both less than the safety distance, which means that the collision risk is increasing. Consequently, the NIF is increased. As the two ships are gradually passed by the OS at 7 s, the NIFs are decreased. At 8 s, TS1 and TS2 are overtaken by the OS, the NIF decreases further.

The NIF calculated by our algorithm is shown in Fig. 7(a). From 0 s to 7 s, although the distance between TS1 and OS, as well as between TS2 and OS, is equal, TS2 and OS are consistently forming an encounter situation of crossing giving-way as COLREGs. And the OS is the giving-way ship. According to regulations, it should constantly observe TS2's behavior changes and take evasive action when necessary. Therefore, the collision risk of TS2 is greater than that of TS1.

The performance of Method 1 and Method 2 are illustrated in Fig. 7(b) and Fig. 7(c), respectively. It can be seen that from Fig. 7(b), the NIF of TS1 and TS2 fluctuate slightly around 0.5 during the 8 s, which means that the influences of TS1 and TS2 on the OS are almost identical. So, this method cannot reflect the collision risk defined by COLREGs. From Fig. 7(c), it can be also seen that the collision risks of both TS1 and TS2 are around 0.3, and there is no variation due to changes. The NIF

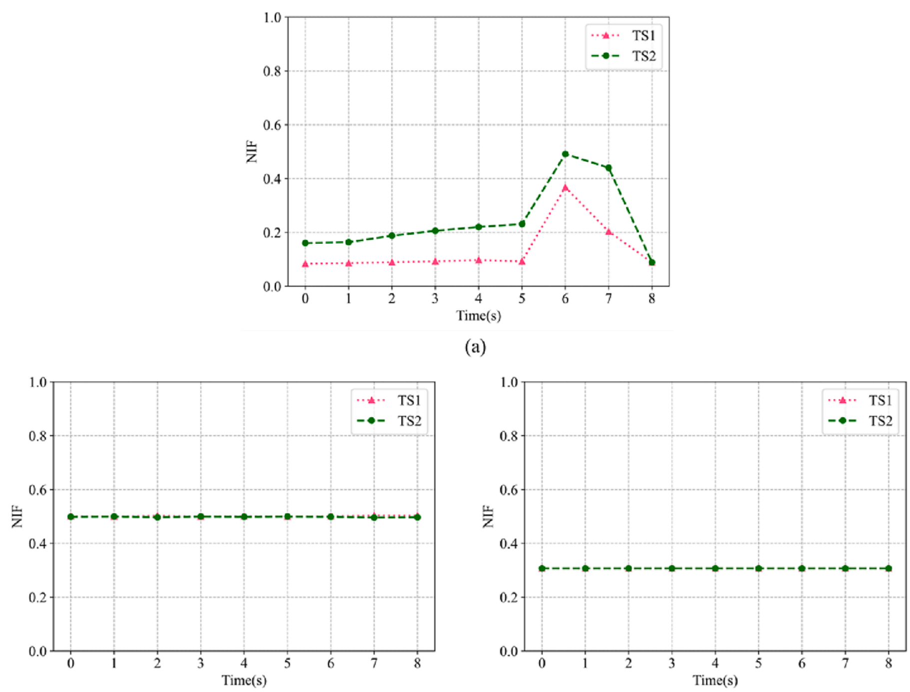
Table 2
Relevant parameters used in training.

Parameters Name	Value	Description
Action space size	11	Number of actions
Learning rate α	0.0003	Learning rate of NN
Attenuation factor γ	0.95	Discount factor
Replay buffer size D	1,000,000	Historical experiences data
Exploration rate ε	0.995	Exploration rate of action
Exploration decay rate	0.00005	Exploration decay rate of action
Batch size	1024	Size of extracted empirical data
Hidden layers size	4	Size of hidden layers
Number of neurons	128	Number in hidden layer
Activation function	Relu	Neuron activation function



(a) (b)

Fig. 6. (a) Trajectory of TSs and OS; (b) Distance between TSs and OS.



(a)

(b) (c)

Fig. 7. (a) NIF of proposed method; (b) NIF of method 1; (c) NIF of method 2.

calculated by method 2 in this case are not reasonable values.

Therefore, when situations defined by COLREGs are taken into consideration, our method can also accurately assess the potential impact of incoming ships. It is human-like and is more in line with the judgment of human beings.

4.3. Compliance test with the COLREGs

The COLREGs stipulate three typical encounter situations: Overtaking, Head-on, and Crossing give-way, and assign collision avoidance responsibilities. To verify the compliance of the algorithm, these scenarios are constructed. According to COLREGs, OS should give way to TS proactively. The experiment aims to verify if OS can conduct compliant collision avoidance according to COLREGs. Once the distance is less than the safety distance, collision avoidance is performed. The trajectories are illustrated in Fig. 8(a)-(c), with numerical labels indicating positions at specific time points. The changes in the distance are illustrated in Fig. 8(d).

The trajectory of the overtaking is shown in Fig. 8(a). The initial heading angle of OS is 0°, the TS is 0°. The OS is in the rear of TS and the field of view is between $-112.5^\circ$and 112.5°. It is compliant with the definition of overtaking of the COLREGs. As shown in Fig. 8(d), the distance between them is about to reach the safety distance. Once navigation continues, the collision will occur. The OS should change its heading angle to 30° and pass the TS from its port side. At 6 s, it successfully overtakes the TS and can turn leftwards to the target.

The trajectory of the head-on is shown in Fig. 8(b). The initial heading angle of OS is 0°, and TS is 180°. As shown in Fig. 8(d), the distance between them is greater than the safety distance over the first 3 s. The OS can navigate towards the target following its initial heading. At 3 s, the distance is close to the safety distance, and a collision may be occurred. In this situation, the OS is the giving-way ship and should turn to starboard in advance. At 4 s, the heading angle is turned to 45° and passes the TS on its port side. After 6 s, The OS steers towards the target leftward and reaches the target at 13 s. It adheres to the COLREGs.

The trajectory of the crossing giving-way is shown in Fig. 8(c). The initial heading angle of OS is 0°, and the TS is -90°. As shown in Fig. 8 (d), at 1 s, the distance is greater than the safety distance, and the OS continues its course. At 2 s, the distance is less than the safe distance, and TS is in the field of view between 5° and 112.5° on the starboard side of

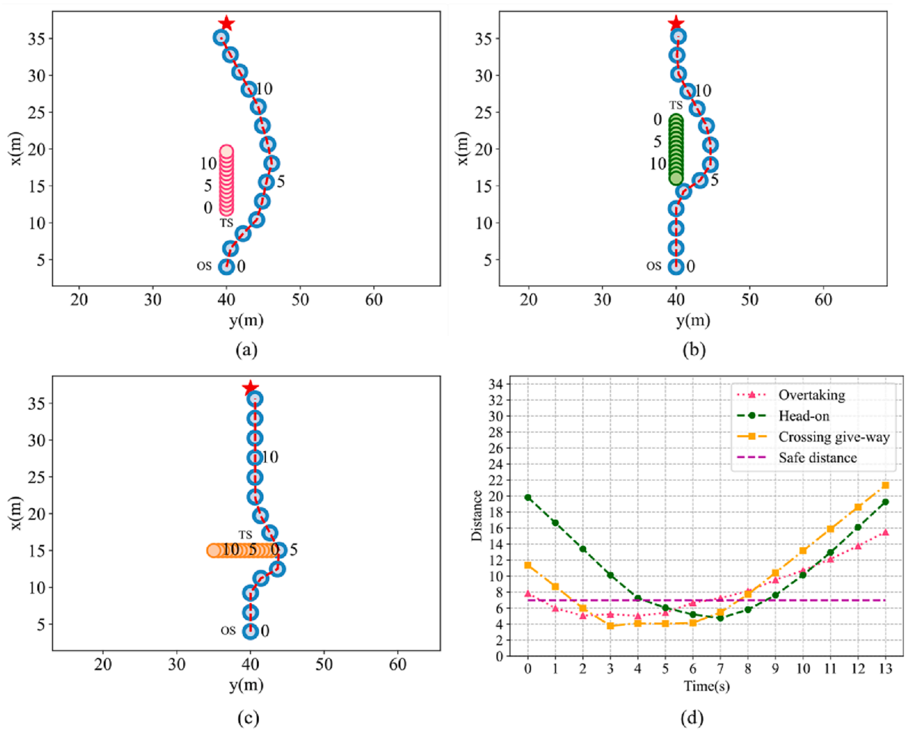


Fig. 8. (a) Overtaking; (b) Head-on; (c) Crossing give-way;(d) Distance between TSs and OS.

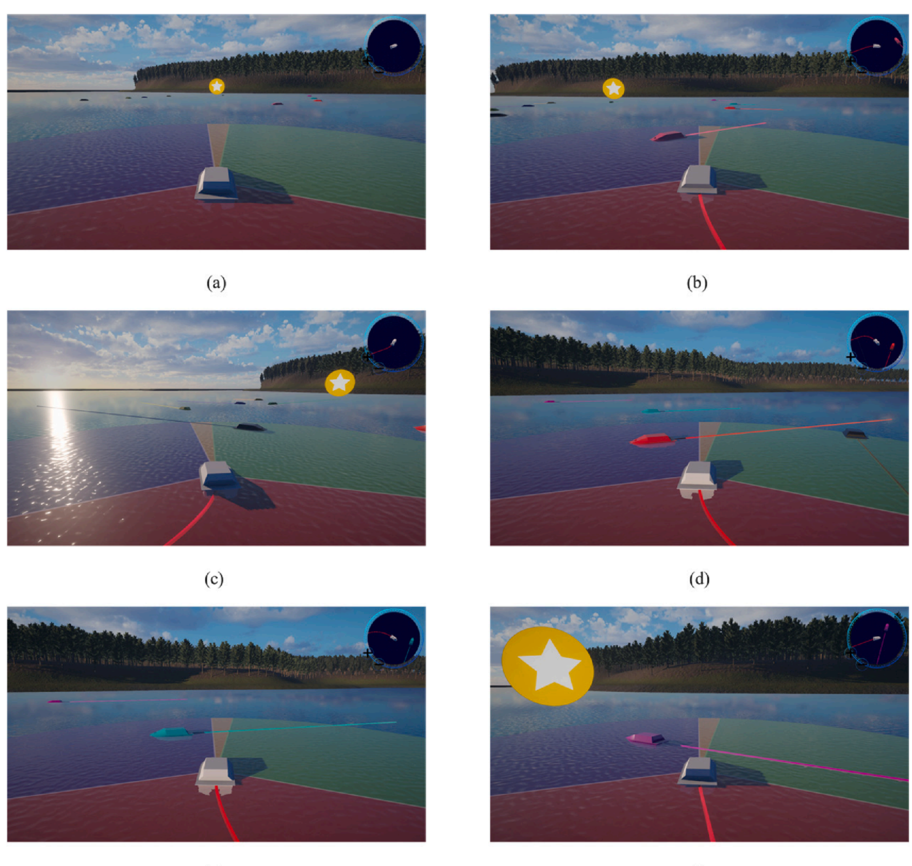


Fig. 9. (a) Overview of 3D Environment; (b) Collision avoidance process between OS and TS1; (c) Collision avoidance process between OS and TS6;(d) Collision avoidance process between OS and TS2; (e) Collision avoidance process between OS and TS3; (f) Collision avoidance process between OS and TS4.

the OS. It adheres to the definition of crossing giving-way, and the OS is the giving-way ship. The OS changes its heading angle to 45° and passes the TS on its port side. At 5 s, collision avoidance is completed, and the OS turns leftward to the target. It is compliant with the COLREGs.

4.4. Performance in the multi-ship encounter situation

Another advantage of our method is that it can make local path planning in complex maritime environments. Since the navigation environment is complex and full of dynamic obstacles, a simulation environment with multiple dynamic obstacles is built to verify the collision avoidance effect of the algorithm and the accuracy of the NIF. A multi-ship experiment is conducted, as shown in Fig. 9. Nine TSs are set in the platform and parameters are given in Table 3.

DRL is the main framework of our human-like method. In the test, 40,000 rounds of training are implemented, and the total rewards are recorded as in Fig. 10. The horizontal axis represents the number of rounds, and the vertical axis represents the total rewards. The red line is the moving average of rewards (every 200 rounds), and the shaded green area is the actual reward. In the early training stages, the USV primarily focuses on exploration, resulting in a higher frequency of erroneous attempts. As it continues to interact with the environment, it progressively transitions towards exploiting the learned experiences. The neural network has been further trained. After 20,000 rounds, our algorithm converges rapidly and the USV selects the optimal actions based on the environment.

The snapshot of the trajectories of our experiment is given in Fig. 11. As shown in Fig. 11(a), a crossing giving-way situation is formed by OS and TS7 at the first 5 s, and the OS is the giving-way ship. As their distance is gradually decreasing, TS7 has a higher influence on OS than others, as depicted in Fig. 12. However, it is not significant, and the OS continues to navigate towards the target. As the distance further decreases, the collision risk with OS starts to increase. At 6 s, OS starts to turn starboard to avoid collision. According to the COLREGs, it passes TS7 from its portside, as shown in Fig. 9(b). At 7 s, a giving-way relationship is established between the OS and TS2, and the influence of TS2 on OS gradually rises.

As shown in Fig. 11(b), TS6 is close to OS at 10 s. According to the COLREGs, TS6 is the giving-way ship, the OS can either hold its course or give way. For simplicity, giving-way action is made by OS, and a turn is initiated at 11 s. It changes its course to 45° to the starboard of the TS6 to avoid collision, as shown in Fig. 9(c).

After that, a crossing giving-way situation is formed between OS and TS2 over 12–13 s. The distance between them is close and the influence of TS2 on OS increases. OS is the giving-way ship and turns its heading angle to 160° to pass the portside of TS2 over 14–15 s, as shown in Fig. 9 (d). As shown in Fig. 11 (c), OS begins to turn left towards the target at 16 s. However, at this time, certain risks are introduced into the action space of OS, which is caused by TS2. So, the influence of TS2 on OS slightly increases. After collision avoidance with the TS2, the OS encounters the TS3 at the same time. As the COLREGs, a crossing giving-way situation is formed by them. At present, OS turns to starboard

Table 3
Initial information about vessels.

Ship number	Initial posit.	Initial orient.(deg)
OS	(4,7)	70
TS1	(25,70)	−95
TS2	(20,40)	20
TS3	(0,50)	20
TS4	(0,70)	15
TS5	(35,40)	180
TS6	(35,30)	160
TS7	(5,30)	−40
TS8	(35,45)	160
TS9	(35,55)	160

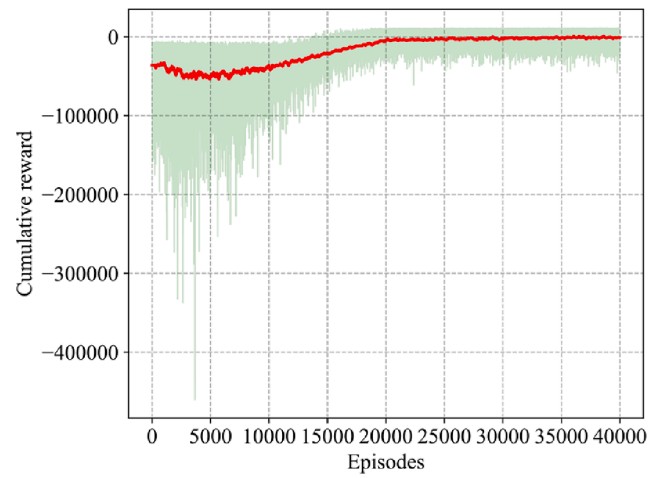


Fig. 10. Reward during the training of DRL.

and passes the portside of the TS3 over 18–20 s, as shown in Fig. 9(e). Subsequently, OS turns left to 45° at 21 s. At this time, TS3 is close and may collide with OS. The influence of TS3 on OS slightly increases. Over the 22–24 s, a crossing giving-way situation is gradually formed by TS4 and OS. Due to the distance being relatively long between them, its influence on the OS slightly increases over the 24–25 s. As shown in Fig. 11 (d), during the 27–28 s, TS4 is directly in front of OS, the collision risk between them sharply increases. At 28 s, as illustrated in Fig. 9(f), OS slightly turns to starboard to avoid collision with TS4. After 29 s, collision avoidance ends and the collision risk starts to decrease. At 30 s, OS turns left to the target. After that, OS reaches the target at the 36 s.

Through the above experiments and analysis, the following conclusions can be drawn. In a complex environment, our algorithm is convergent and can obtain a collision avoidance strategy through interaction with the environment. It is feasible and compliant with the COLREGs. It can accurately assess the influence of TSs on OS and take good collision avoidance actions.

4.5. Performance comparative experiments

To verify the advantages of our algorithm in terms of safety and smoothness, comparative experiments are conducted with classical Dueling DQN, DQN, DWA, VO, and DWA-VO methods. In this experiment, the settings of the velocity and heading of the TSs are consistent with the experiment in section 4.4, with only the starting points of each ship being changed. 100 rounds are conducted by the aforementioned methods. Collision avoidance success rate, average path risk, and average large steering count are the metrics, and the comparison results are shown in Table 4. In terms of collision avoidance success rate, our algorithm is with the highest success rate. Compared to the classical Dueling DQN and the DQN, the integration of a NIF calculation module and obstacle avoidance rewards based on VO increase the success rate by at least 20 %. The path risk p_r is designed based on Huang et al. (2020). to comprehensively evaluate the potential risks, and formulated as follows:

$$p_r = \begin{cases} 1, ifcollisionhappens \\ \frac{n_e}{n_{vo}}, elseifVOisconstructed \end{cases} \tag{14}$$

where n_{vo} is the number of obstacle areas constructed in a path, and n_e denotes the entry number of OS into the obstacle area. According to the aforementioned experiments, once collision avoidance action leads OS entry into the obstacle area, then it has a certain risk. Therefore, the risk of the path can be determined by the entry number of OS to the obstacle area. When a collision occurs, the path risk is assigned as 1. By checking

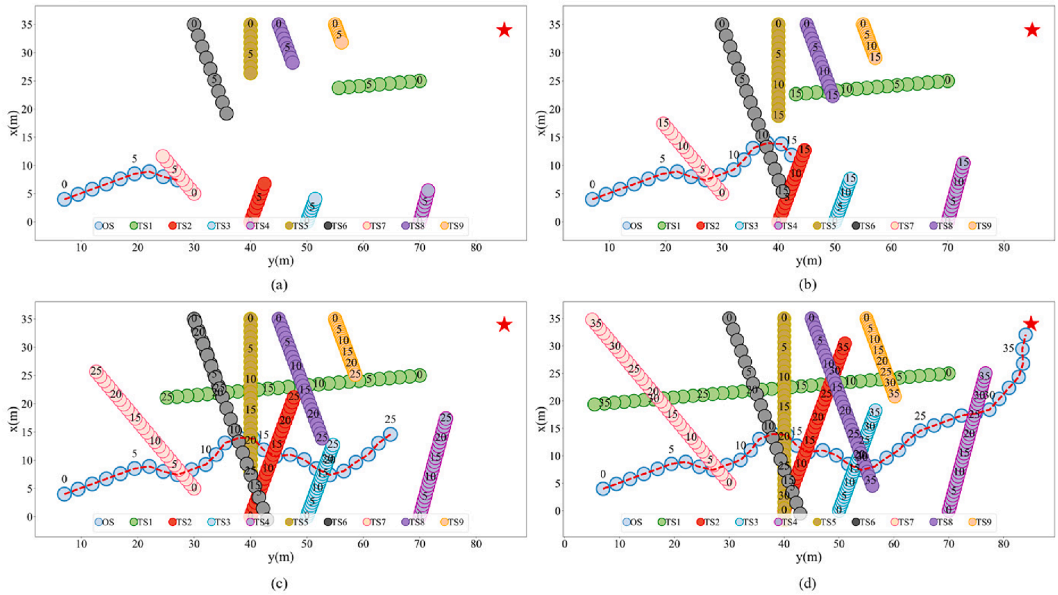


Fig. 11. (a) Snapshot of trajectories (t = 8 s); (b) Snapshot of trajectories (t = 15 s); (c) Snapshot of trajectories (t = 25 s);(d) Snapshot of trajectories (t = 36 s).

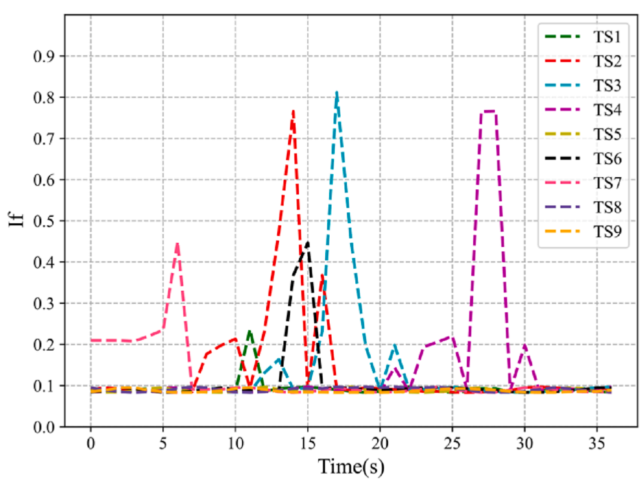


Fig. 12. Navigational impact factors of TSs.

Table 4
Analysis of collision avoidance results.

Method	Success rate (%)	Avera. path risk (%)	Avera. large steering count (%)
Proposed algorithm	0.86	0.23	0.28
Dueling DQN	0.58	0.48	0.38
DQN algorithm	0.44	0.61	0.59
DWA algorithm	0.56	0.49	0.94
VO algorithm	0.50	\	0.61
DWA-VO algorithm	0.67	\	1.11

the values of the average risk of the paths generated by each algorithm, our method has a lower collision risk. It is more closely to the expectations of a seaman for a safe path.

In addition, more significant turns in a path may lead to capsizing. Although the turning range is not strictly restricted in COLREGs, in conjunction with the model of the USV, it can be considered a risky action when the heading angle changes by more than 80°. In view of the average number of significant turns generated by different algorithms, it can be seen that our algorithm has the lowest significant turns. Hence, it is the smoothest and safest path.

To compare the performance of collision avoidance in different environments, experiments are conducted to compare with the aforementioned methods. The success rates are illustrated in Fig. 13. The horizontal axis represents the number of ships, and the vertical axis represents the success rate of collision avoidance. As the ships increase, the success rates of the other algorithms significantly fluctuate. In contrast, our algorithm keeps the highest performance and stability.

5. Conclusion

The paper discusses the collision avoidance problem for USVs in hybrid maritime environments, in which ships with different autonomy levels coexist, and proposes a human-like collision avoidance method based on DRL and VO. The proposed method has the following features to achieve human-like collision avoidance: 1) The navigation impact factor calculation module based on fuzzy theory is introduced to simulate human beings' attention mechanism when facing the situations of multi-ships; 2) To obey the human beings' rules of COLREGs, collision avoidance situations based on COLREGs are for the first time integrated into the navigation impact factor calculation module; 3) The VO algorithm is also incorporated into the reward function of the DRL, which can reduce collision risks and improve safety; and 4) a series of reward functions are carefully designed to balance safety, smoothness, and good driving operation. The issue of sparse rewards is also addressed. With these features, the proposed method ensures that the planned paths are highly practical and suitable for real-world applications. The performance of the proposed method is validated through experiments and comparison studies. The results show that our method can make better collision avoidance and has advantages over other methods in success rate, compliance with COLREGs, and smoothness.

Our future works will focus on the following aspects. First, the action space of the USVs is discretized, which may compromise the maneuverability and flexibility of USVs. The collision avoidance problem in a continuous action space will be considered in the future. Second, the training speed of the DRL will be discussed further. Typically, DRL requires a significant amount of time for training, the efficiency of training is another focus in the future. Moreover, the impact of environmental uncertainty and inaccurate sensor data are not considered, they will be

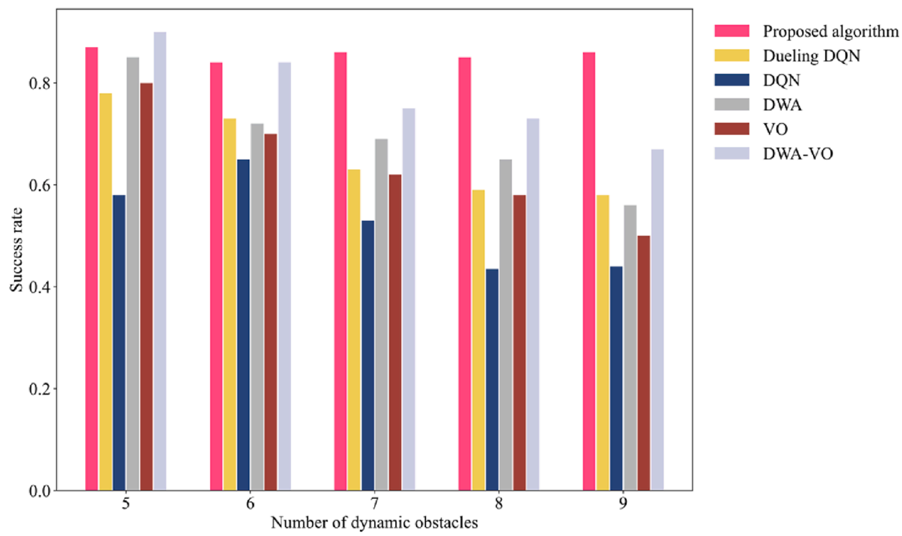


Fig. 13. Comparison of collision avoidance success rate.

incorporated into the model to enhance the performance in practical application.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 62373191 and the Jiangsu Province Research and Practice Innovation Plan Project under Grant SJCX23_2129. (Corresponding author: Xiaofei Yang).

References

Wang, N., & Xu, H. (2020). Dynamics-constrained global-local hybrid path planning of an autonomous surface vehicle. *IEEE Transactions on Vehicular Technology, 69*(7), 6928–6942.

Yan, X., Wang, S., Ma, F., Liu, Y., & Wang, J. (2020). A novel path planning approach for smart cargo ships based on anisotropic fast marching. *Expert Systems with Applications, 159*, Article 113558.

Zhang, G., Liu, S., Huang, J., & Zhang, W. (2022). Dynamic event-triggered path-following control of underactuated surface vehicle with the experiment verification. *IEEE Transactions on Vehicular Technology, 71*(10), 10415–10425.

Yang, X., She, H., Lou, M., Ye, H., Guan, J., Li, J., et al. (2024). A Joint Ship Detection and Waterway Segmentation Method for Environment-Aware of USVs in Canal Waterways. *IEEE Transactions on Automation Science and Engineering*, 1558–3783.

Wang, S., Zhang, Y., Zhang, X., & Gao, Z. (2023). A novel maritime autonomous navigation decision-making system: Modeling, integration, and real ship trial. *Expert Systems with Applications, 222*, Article 119825.

Wang, N., Zhang, Y., Ahn, C. K., & Xu, Q. (2022). Autonomous pilot of unmanned surface vehicles: Bridging path planning and tracking. *IEEE Transactions on Vehicular Technology, 71*(3), 2358–2374.

Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research, 5*(1), 90–98.

Liu, W., Qiu, K., Yang, X., Wang, R., Xiang, Z., Wang, Y., et al. (2023). COLREGS-based collision avoidance algorithm for unmanned surface vehicles using modified artificial potential fields. *Physical Communication, 57*, Article 101980.

Zhang, G., Han, J., Li, J., & Zhang, X. (2022). APF-based intelligent navigation approach for USV in presence of mixed potential directions: Guidance and control design. *Ocean Engineering, 260*, Article 111972.

Fox, D., Burgard, W., & Thrun, S. (1997). The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine, 4*(1), 23–33.

Guan, W., & Wang, K. (2023). Autonomous collision avoidance of unmanned surface vehicles based on improved A-Star and dynamic window approach algorithms. *IEEE Intelligent Transportation Systems Magazine, 15*(3), 36–50.

Chiang, H.-T.-L., & Tapia, L. (2018). COLREG-RRT: An RRT-based COLREGS-compliant motion planner for surface vehicle navigation. *IEEE Robotics and Automation Letters, 3*(3), 2024–2031.

Fiorini, P., & Shillert, Z. (1998). Motion planning in dynamic environments using velocity obstacles. *International Journal of Robotics Research, 17*(7), 760–772.

Xie, Z., & Dames, P. (2023). DRL-VO: Learning to navigate through crowded dynamic scenes using velocity obstacles. *IEEE Transactions on Robotics, 39*(4), 2700–2719.

Shen, Z., Ding, W., Liu, Y., & Yu, H. (2022). Path planning optimization for unmanned sailboat in complex marine environment. *Ocean Engineering, 269*, Article 113475.

Wang, Y., Yu, X., Liang, X., & Li, B. (2018). A COLREGs-based obstacle avoidance approach for unmanned surface vehicles. *Ocean Engineering, 169*, 110–124.

Yao, M., Deng, H., Feng, X., Li, P., Li, Y., & Liu, H. (2024). Improved dynamic windows approach based on energy consumption management and fuzzy logic control for local path planning of mobile robots. *Computers & Industrial Engineering, 187*, Article 109767.

Ahn, J. H., Rhee, K. P., & You, Y. H. (2012). A study on the collision avoidance of a ship using neural networks and fuzzy logic. *Applied Ocean Research., 37*, 162–173.

Xie, S., Garofano, V., Chu, X., & Negenborn, R. R. (2019). Model predictive ship collision avoidance based on Q-learning beetle swarm antenna search and neural networks. *Ocean Engineering, 193*, Article 106609.

Zhang, C., Cheng, P., Lin, B., Zhang, W., & Xie, W. (2023). DRL-based target interception strategy design for an underactuated USV without obstacle collision. *Ocean Engineering, 280*, Article 114443.

Chen, C., Chen, X., Ma, F., Zeng, X., & Wang, J. (2019). A knowledge-free path planning approach for smart ships based on reinforcement learning. *Ocean Engineering, 189*, Article 106299.

Zhou, C., Wang, Y., Wang, L., & He, H. (2022). Obstacle avoidance strategy for an autonomous surface vessel based on modified deep deterministic policy gradient. *Ocean Engineering, 243*, Article 110166.

Cho, Y., Han, J., & Kim, J. (2022). Efficient COLREG-compliant collision avoidance in multi-ship encounter situations. *IEEE Transactions on Intelligent Transportation Systems, 23*(3), 1899–1911.

Woo, J. H., & Kim, N. (2020). Collision avoidance for an unmanned surface vehicle using deep reinforcement learning. *Ocean Engineering, 199*, Article 107001.

Xue, D., Wu, D., Yamashita, A. S., & Li, Z. (2023). Proximal policy optimization with reciprocal velocity obstacle based collision avoidance path planning for multi-unmanned surface vehicles. *Ocean Engineering, 273*, Article 114005.

Sun, X., Wang, G., Fan, Y., & Mu, D. (2023). Collision avoidance control for unmanned surface vehicle with COLREGs compliance. *Ocean Engineering, 263*, Article 113263.

Chun, D. H., Roh, M. I., Lee, H. W., Ha, J., & Yu, D. (2021). Deep reinforcement learning-based collision avoidance for an autonomous ship. *Ocean Engineering, 234*, Article 109216.

Bukhari, A. C., Tusseyeva, I., Iee, B. G., & Kim, Y. G. (2013). An intelligent real-time multi-vessel collision risk assessment system from VTS view point based on fuzzy inference system. *Expert Systems with Applications, 40*(4), 1220–1230.

Zheng, R., Xie, H., & Yuan, K. (2023). Autonomous collision avoidance system in a multi-ship environment based on proximal policy optimization method. *Ocean Engineering, 272*, Article 113779.

French, R. M., & Douglas, R. H. (1991). Tabletop: A stochastic, emergent model of analogy-making. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 708–713). Hillsdale, NJ: Lawrence Erlbaum.

Li, X. H., et al. (2022). A survey of data-driven and knowledge-aware eXplainable AI. *IEEE Transactions on Knowledge and Data Engineering, 34*(1), 29–49.

Li, Y., Peng, Y., & Zheng, J. (2023). Intelligent ship collision avoidance model integrating human thinking experience. *Ocean Engineering, 286*(1), Article 115510.

Jiang, L., An, L., Zhang, X., Wang, C., & Wang, X. (2022). A human-like collision avoidance method for autonomous ship with attention-based deep reinforcement learning. *Ocean Engineering, 264*, Article 112378.

He, Y., Jin, Y., Huang, L., Xiong, Y., Chen, P., & Mou, J. (2017). Quantitative analysis of COLREG rules and seamanship for autonomous collision avoidance at open sea. *Ocean Engineering, 140*, 281–291.

Zhao, L., & Roh, M. I. (2019). COLREGs-compliant multiship collision avoidance based on deep reinforcement learning. *Ocean Engineering, 191*, Article 106436.

Namgung, H., & Kim, J.-S. (2021). Collision risk inference system for maritime autonomous surface ships using COLREGs rules compliant collision avoidance. *IEEE Access, 9*, 7823–7835.

Xu, X., Cai, P., Cao, Y., Chu, Z., Zhu, W., & Zhang, W. (2023). Real-time planning and collision avoidance control method based on deep reinforcement learning. *Ocean Engineering, 281*, Article 115018.

Kuwata, Y., Wolf, M. T., Zarzhitsky, D., & Huntsberger, T. L. (2014). Safe maritime autonomous navigation with COLREGS, using velocity obstacles. *IEEE Journal of Oceanic Engineering, 39*(1), 110–119.

Li, L., Wu, D., Huang, Y., & Yuan, Z. (2021). A path planning strategy unified with a COLREGS collision avoidance function based on deep reinforcement learning and artificial potential field. *Applied Ocean Research, 113*, Article 102759.

Yang, X., Shi, Y., Liu, W., Hui, Y., Zhong, W., & Xiang, Z. (2022). Global path planning algorithm based on double DQN for multi-tasks amphibious unmanned surface vehicle. *Ocean Engineering, 266*, Article 112809.

Gao, M., Kang, Z., Zhang, M., Liu, J., & Zhao, F. (2022). MASS autonomous navigation system based on AIS big data with dueling deep Q networks prioritized replay reinforcement learning. *Ocean Engineering, 249*, Article 110834.

Huang, Y., & Van Gelder, P. (2020). Time-varying risk measurement for ship collision prevention. *Risk Analysis, 40*(1), 24–42.

Mamdani, E. H., & Assilian, S. (1999). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Human-Computer Studies, 51*(2), 135–147.

Wu, B., Yip, T. L., Yan, X., & Soares, C. G. (2019). Fuzzy logic based approach for ship-bridge collision alert system. *Ocean Engineering, 187*, Article 106152.

Zhan, C., & Zeng, Y. (2022). Energy minimization for cellular-connected UAV: From optimization to deep reinforcement learning. *IEEE Transactions on Wireless Communications, 21*(7), 5541–5555.

Yan, X., Yang, X., Feng, B., Liu, W., Ye, H., Zhu, Z., et al. (2024). A navigation accuracy compensation algorithm for low-cost unmanned surface vehicles based on models and event triggers. *Control Engineering Practice, 146*, Article 105896.

Yang, X., Yan, X., Liu, W., Ye, H., Du, Z., & Zhong, W. (2022). An improved stanley guidance law for large curvature path following of unmanned surface vehicle. *Ocean Engineering, 266*, Article 112797.

Xu, M., Xie, X., Lv, P., Niu, J., Wang, H., Li, C., et al. (2021). Crowd behavior simulation with emotional contagion in unexpected multihazard situations. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51*(3), 1567–1581.

Shi, Z., Zhen, R., & Liu, J. (2022). Fuzzy logic-based modeling method for regional multi-ship collision risk assessment considering impacts of ship crossing angle and navigational environment. *Ocean Engineering, 259*, Article 111847.