Review

Towards LLM Enhanced Decision: A Survey on Reinforcement Learning based Ship Collision Avoidance

Yizhou Wu , Jin Liu * , Xingye Li , Junsheng Xiao , Tao Zhang , Haitong Xu , Lei Zhang *

Posted Date: 5 November 2025

doi: 10.20944/preprints202511.0267.v1

Keywords: reinforcement Learning; ship collision avoidance; multi-agent reinforcement learning; large language model; LLM agent; COLREGs compliance

Review

Towards LLM Enhanced Decision: A Survey on Reinforcement Learning based Ship Collision Avoidance

Yizhou Wu [1], Jin Liu [2,*], Xingye Li [2], Junsheng Xiao [2], Tao Zhang [1], Haitong Xu [3] and Lei Zhang [4,*]

[1] Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China
[2] College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China
[3] Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[4] Shanghai Science Center for Autonomous Intelligent Unmanned Systems, Tongji University, Shanghai 200092, China
* Correspondence: jinliu@shmtu.edu.cn_(J.L.); reizhg@tongji.edu.cn (L.Z.)

Abstract

This comprehensive review examines the works of reinforcement learning (RL) in ship collision avoidance (SCA) from 2014 to the present, analyzing the methods designed for both single-agent and multi-agent collaborative paradigms. While prior research has demonstrated RL's advantages in environmental adaptability, autonomous decision-making, and online optimization over traditional control methods, this study systematically addresses the algorithmic improvements, implementation challenges, and functional roles of RL techniques in SCA, such as Deep Q-Network (DQN), Proximal Policy Optimization (PPO), and Multi-Agent Reinforcement Learning (MARL). It also highlights how these technologies address critical challenges in SCA, including dynamic obstacle avoidance, compliance with Convention on the International Regulations for Preventing Collisions at Sea (COLREGs), and coordination in dense traffic scenarios, while underscoring persistent limitations such as idealized assumptions, scalability issues, and robustness in uncertain environments. Contributions include a structured analysis of recent technological evolution, and a Large Language Model (LLM) based hierarchical architecture integrating perception, communication, decision-making, and execution layers for future SCA systems, which prioritizes the development of scalable, adaptive frameworks that ensure robust and compliant autonomous navigation in complex, real-world maritime environments.

Keywords: reinforcement Learning; ship collision avoidance; multi-agent reinforcement learning; large language model; LLM agent; COLREGs compliance

1. Introduction

Vessel collision incidents have long been a major challenge in maritime safety, with catastrophic consequences that extend far beyond direct economic losses (as reported by the International Maritime Organization, a single major collision can result in average losses exceeding USD 30 million). More critically, these incidents pose severe threats to human life (with approximately 2,000 fatalities annually worldwide due to maritime accidents) and to the marine environment (collisions account for about 12% of marine oil pollution). Researches[1–3]have revealed that approximately 75% to 85% of maritime accidents can be attributed to human error, including misinterpretation of the COLREGs, delayed situational awareness under complex conditions, and inappropriate emergency maneuvers. These realities have spurred an urgent demand for intelligent ship collision avoidance (SCA) systems capable of autonomously generating optimal, regulation-compliant decisions in dynamic and uncertain environments.

Traditional SCA approaches primarily follow three technical paradigms:(1)Geometric algorithms. Methods such as the Velocity Obstacle (VO) approach[4] and its variants (e.g. Reciprocal VO (RVO)[5] and Hybrid RVO (HRVO)[6]), forecast collision risks by constructing velocity feasible regions. These methods are computationally inefficient, and incapable of modeling nonlinear vessel dynamics.(2)Rule-based expert systems. These systems, exemplified by fuzzy logic[7], operate based on predefined rule bases for decision-making. However, their generalizability is constrained by the completeness of the rule set, limiting their applicability in diverse scenarios.(3)Optimal control theory, notably Model Predictive Control (MPC)[8], generates collision-free trajectories by solving constrained optimal control problems online. Yet, the performance of MPC heavily depends on precise hydrodynamic modeling (such as added mass and damping coefficients) and lacks robustness to abrupt disturbances (e.g., sudden gusts or current shifts).

In recent years, reinforcement learning (RL)[9], as a promising model-free sequential decision-making framework, has been extensively studied to overcome the limitations of traditional SCA methods. Rooted in Markov Decision Process (MDP)[10] theory, RL enables an agent to interact with the environment, observe its state (e.g., own-vessel position, velocity vector, relative bearing of obstacles), take actions (e.g., rudder angle, engine throttle commands), and receive rewards (e.g., collision penalties, safety distance rewards). The agent ultimately learns a policy that maximizes the expected cumulative reward , which is a discount factor. Compared to traditional methods, RL offers three key advantages:(1)Environmental adaptability. End-to-end learning enables direct feature extraction from raw sensor data (AIS, radar point clouds, visual images) without explicit environment modeling.(2)Autonomous decision-making. Policy networks can learn complex, high-dimensional mappings, allowing for sophisticated maneuvers beyond human-defined rules (such as emergent evasive strategies in emergencies).(3)Online optimization. Techniques such as experience replay enable continuous policy improvement, enhancing robustness in uncertain scenarios.

As illustrated in Figure 1, the evolution of RL applications in SCA can be delineated into three distinct stages. (1)The initial stage (beginning around 2014) was characterized by the adoption of classical RL algorithms, such as Q-learning[11], for relatively simple scenarios involving single-vessel navigation and static obstacle avoidance, establishing a foundational methodological framework for subsequent advancements. (2)The second stage (from 2018 onward) marked a transition to more complex and realistic settings, with research efforts increasingly focusing on multi-vessel dynamic interaction scenarios. During this period, Deep Reinforcement Learning (DRL) methods, including Deep Q-Network (DQN)[12], Deep Deterministic Policy Gradient (DDPG)[13], and Proximal Policy Optimization (PPO)[14], were extensively explored to address the higher dimensionality and uncertainty inherent in dynamic maritime environments. Researches in this stage [15–17] also witnessed a growing emphasis on modeling vessel-to-vessel interactions and ensuring compliance with Convention on the International Regulations for Preventing Collisions at Sea (COLREGs) to enhance the generalizability and robustness of RL algorithms. (3)The latest stage (emerging since 2021) has been defined by a paradigm shift toward large-scale, multi-agent, and fully autonomous maritime scenarios. In this stage, driven by significant advances in computing power and the Multi-Agent Reinforcement Learning (MARL) framework[18], many studies[19–22] have begun to focus on achieving multi-vessel collaborative SCA, developing integrated perception and decision-making architectures, and achieving end-to-end autonomous navigation, which opening up new avenues for improving decision-making capabilities and safety performance in highly complex and uncertain marine environments.

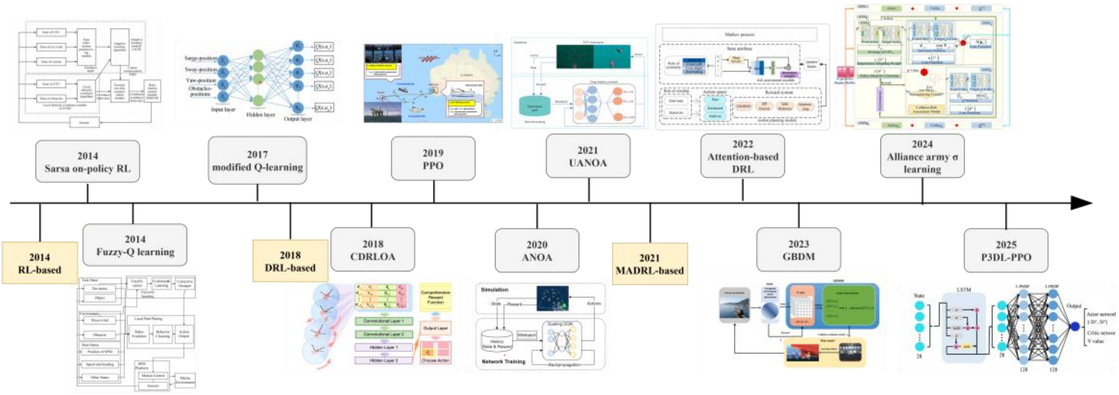


Figure 1. Evolution of Reinforcement Learning-based SCA Approach.

Despite the growing body of research in this field, a systematic literature review and bibliometric analysis specifically focused on RL-based SCA remains conspicuously lacking, as evidenced by comprehensive searches in the Web of Science database. This notable gap highlights the pressing need for a rigorous and structured synthesis of the existing literature to elucidate current trends, critically assess recent advancements, and inform future research directions. In response, this review aims to provide a comprehensive and systematic overview of the state-of-the-art in RL-driven SCA, examining prevailing trends, unresolved challenges, and emerging opportunities, and ultimately offering valuable insights to guide future innovation in this rapidly evolving domain.

2. Methods of the Current Study

While foundational reviews on SCA exist, a significant gap remains for a systematic analysis encompassing the rapid advancements in RL since 2018, particularly the pivotal shift toward MARL paradigms post-2021. Furthermore, previous syntheses have not explicitly deconstructed the problem according to the fundamental dichotomy of SCA scenarios: single-agent versus multi-agent, and collaborative versus non-collaborative environments. Therefore, the purpose of this study is twofold: first, to provide an up-to-date review of RL-based SCA from 2018 to the present, capturing the latest developments in deep RL and MARL; and second, to introduce a novel analytical framework that systematically categorizes and evaluates methods according to this scenario splitting, thereby offering a more structured and granular understanding of the field's evolution and current state of the art.

2.1. Search Strategy

To ensure comprehensive coverage, a search was performed across three major interdisciplinary databases: Web of Science (WOS), IEEE Xplore, and Scopus. The search query combined terms related to the technology ("reinforcement learning," "deep reinforcement learning," "multi-agent reinforcement learning") with terms related to the application ("SCA," "maritime autonomous navigation," "unmanned surface vehicle," "COLREGs")

2.2. Inclusion and Exclusion Criteria

This study established specific criteria for article selection to maintain a focus on recent and technically substantive contributions:

Inclusion Criteria:

- Publication date between 31 October, 2014, and March 17, 2025.
- Primary focus on RL or Deep RL algorithms for SCA.
- Empirical validation through simulation or real-world experiments.
- Peer-reviewed journal articles or conference proceedings published in English.

Exclusion Criteria:

- Studies published prior to 2018 or focusing solely on traditional non-RL methods.
- Applications in non-maritime domains (e.g., aerial or ground robots).
- Fully theoretical papers without algorithmic implementation or validation.

2.3. Selection Process

The literature selection followed a structured multi-stage process. Initial database searches yielded 346 records. After removing 189 articles that did not focus on RL for SCA, 157 articles were screened for relevance. The remaining full-text articles were assessed for eligibility based on the predefined criteria. This rigorous evaluation led to the exclusion of 132 articles, with common reasons being non-maritime application scope or lack of empirical RL focus. Ultimately, 25 studies met all inclusion criteria and formed the core corpus for this systematic review. The screening was conducted independently by two reviewers, with disagreements resolved through consensus or consultation with a third reviewer. Following this, Figure 2 visually present the overall structure and key points of the review.

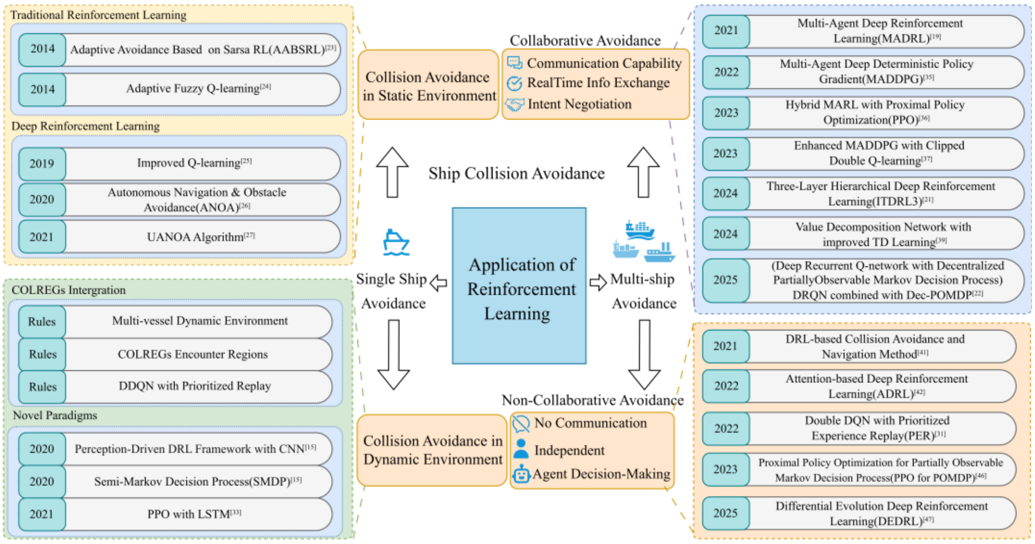

Traditional Reinforcement Learning
2014 Adaptive Avoidance Based on Sarsa RL(AABSRL)[23]
2014 Adaptive Fuzzy Q-learning[24]
Deep Reinforcement Learning
2019 Improved Q-learning[25]
2020 Autonomous Navigation & Obstacle Avoidance(ANOA)[26]
2021 UANOA Algorithm[27]
COLREGs Intergration
Rules Multi-vessel Dynamic Environment
Rules COLREGs Encounter Regions
Rules DDQN with Prioritized Replay
Novel Paradigms
2020 Perception-Driven DRL Framework with CNN[15]
2020 Semi-Markov Decision Process(SMDP)[15]
2021 PPO with LSTM[33]
Collision Avoidance in Static Environment
Collaborative Avoidance
Communication Capability
RealTime Info Exchange
Intent Negotiation
Ship Collision Avoidance
Single Ship Avoidance
Application of Reinforcement Learning
Multi-ship Avoidance
Non-Collaborative Avoidance
Collision Avoidance in Dynamic Environment
No Communication
Independent
Agent Decision-Making
2021 Multi-Agent Deep Reinforcement Learning(MADRL)[19]
2022 Multi-Agent Deep Deterministic Policy Gradient(MADDPG)[35]
2023 Hybrid MARL with Proximal Policy Optimization(PPO)[36]
2023 Enhanced MADDPG with Clipped Double Q-learning[37]
2024 Three-Layer Hierarchical Deep Reinforcement Learning(ITDRL3)[21]
2024 Value Decomposition Network with improved TD Learning[39]
2025 (Deep Recurrent Q-network with Decentralized PartiallyObservable Markov Decision Process) DRQN combined with Dec-POMDP[22]
2021 DRL-based Collision Avoidance and Navigation Method[41]
2022 Attention-based Deep Reinforcement Learning(ADRL)[42]
2022 Double DQN with Prioritized Experience Replay(PER)[31]
2023 Proximal Policy Optimization for Partially Observable Markov Decision Process(PPO for POMDP)[46]
2025 Differential Evolution Deep Reinforcement Learning(DEDRL)[47]

Figure 2. Evolution of Reinforcement Learning-based SCA Approach Taxonomy of Intelligent Collision Avoidance Approaches for Maritime Systems.

3. Single-Agent Collision Avoidance

As the foundation of autonomous maritime navigation, single-agent SCA primarily focuses on enabling an individual vessel to independently perceive its surroundings and safely maneuver through complex environments. The main objective is to dynamically adjust the heading and speed of the ego vessel to minimize the risk of collisions with both static and dynamic obstacles, including other vessels. Early research in this area leveraged traditional reinforcement learning algorithms to address the limitations of rule-based and optimization-based methods, laying an essential technological groundwork for subsequent developments. With the advent of deep reinforcement learning, the field has seen significant progress in handling more realistic scenarios involving moving obstacles and regulatory compliance. This section provides a comprehensive review of single-agent SCA methods, covering both static and dynamic environments, and analyzes the strengths and limitations of representative approaches.

3.1. Collision Avoidance in Static Environment

Early research on single-agent SCA centered on applying traditional RL algorithms to vessel navigation in environments with only static obstacles, aiming to overcome the limitations of conventional methods in addressing complex maritime environments with multiple influencing factors. These foundational studies established an important technological basis for subsequent developments in the field.

For example, Zhang et al.[23] addressed the challenge of static SCA under environmental disturbances, such as wind and currents, by proposing an Adaptive obstacle Avoidance algorithm Based on Sarsa on-policy RL (AABSRL) for Unmanned Surface Vehicles (USVs) in complex static environments. Specifically, in AABSRL, a Local Obstacle Avoidance Module is proposed, which utilizes an improved heading window algorithm to generate feasible headings and speeds while excluding directions that intersect with static obstacles. Concurrently, an adaptive learning module is introduced to leverage Sarsa RL algorithm[48] to adjust for heading deviations caused by environmental disturbances. Field trials demonstrated that the algorithm achieved a 42.3% reduction in average course deviation compared to traditional methods, facilitating safe navigation at speeds of up to 20 knots in clustered static obstacle fields. However, its dependence on complete prior knowledge of obstacles and the necessity for pre-training in disturbance compensation limit its adaptability, particularly in partially unknown or highly cluttered environments. To overcome the necessity for pre-training of disturbance compensation, Yang et al.[24] proposed a single-agent collision avoidance method based on fuzzy Q-learning[38]. This method simplifies environmental modeling by fuzzifying wind states, systematically navigates discrete compensation angles using chaotic logic mapping during the exploration phase, and introduces "similar states" to accelerate Q-table updates, thereby improving learning efficiency and solving the problem of wind-induced deviation in static obstacle avoidance. Simulations and field tests revealed a 37.5% reduction in heading error compared to vanilla methods. Nonetheless, this approach is constrained by the discrete nature of compensation actions, assumptions regarding wind continuity, and the exclusion of wave-induced disturbances, which can hinder performance in environments characterized by abrupt changes or high-accuracy requirements.

More recently, to enhance the applicability of SCA models in partially unknown or highly chaotic environments, Bhopale et al.[25] proposed an improved Q-learning algorithm for SCA in unknown environments with static obstacles. Their method introduces a "danger zone" around detected obstacles, enforcing a pure exploitation strategy within these zones and applying Q-value penalties to discourage unsafe actions. Additionally, a neural network function approximator is used to manage continuous state spaces. However, this method requires a large number of iterations for the algorithm and neural network to converge, and retraining is required when the obstacle layout or environmental conditions change, resulting in high deployment costs and inefficiency. Wu et al.[26] addressed real-time static obstacle avoidance for USVs using the Autonomous Navigation and Obstacle Avoidance (ANOA) method based on a Dueling DQN. The ANOA defines a custom state space using environmental perception images and a discrete action space, while a multi-objective reward function drives target-reaching and penalizes collisions and boundary violations. Experiments reveal improvements in exploration efficiency and convergence over conventional DQN, but the method's applicability is limited to grid-based simple obstacles, and its performance degrades under partial observability or real-time requirements. To further address the challenge of partial observability, Yan et al.[27] proposed the Autonomous Navigation and Obstacle Avoidance in USVs (UANOA) algorithm, which employs a double Q-network for end-to-end static obstacle avoidance in USVs. Raw sensor data (position and LiDAR) are processed to output discrete rudder commands. The reward function integrates components for SCA, target proximity, and penalizing ineffective navigation. While effective in simulation, the method's reliance on fixed throttle settings constrains maneuverability and adaptability in complex environments.

Although the early attempts of RL methods for SCA in static environments have brought significant performance improvements, these approaches remain limited by their dependence on complete or predefined knowledge of obstacle distributions, reliance on discrete compensation strategies, and simplified environmental assumptions. These limitations notably restrict their adaptability to complex, partially known, and rapidly changing real-world ocean scenarios.

3.2. Collision Avoidance in Dynamic Environment

With the advent of DRL architecture, research on SCA has progressively focused on addressing the demands of real-world navigation, particularly dynamic SCA in complex maritime environments with moving obstacles, other vessels, and regulatory requirements.

For instance, to overcome the limitations of traditional static methods, Cheng and Zhang[28] proposed a Concise DRL Obstacle Avoidance (CDRLOA) algorithm for underactuated unmanned vessels (UMVs) operating under unknown disturbances. In CDRLOA, obstacle avoidance, target approach, speed adjustment, and attitude correction are integrated into a unified reward function, with control strategies learned directly from environmental interactions using a DQN. While effective in simulation, the CDRLOA is limited to single-agent scenarios, does not account for COLREGs, and employs a fixed detection radius, which restricts its applicability in real-world maritime traffic. Shen et al.[29], to better adapt their model to COLREGs and SCA strategies in multi-vessel encounter scenarios, innovatively integrated COLREGs into the reward function of their proposed DQN-based multi-vessel dynamic SCA framework. Meanwhile, by introducing multi-vessel dynamic interaction modeling, such as vessel bumper, vessel domain, and predicted area of danger, their framework demonstrated significant performance improvements in complex maritime environments with a variable number of dynamic targets. However, the need for manual parameter adjustment and discrete action space limits its adaptability in highly congested waters such as inland waterways. Focusing on the dynamic SCA in high-density multi-vessel environments, Zhao and Roh [30] proposed a policy-gradient RL approach that classifies dynamic targets into encounter regions as defined by COLREGs. By considering only the nearest vessel in each region, the input space remains consistent regardless of the total number of obstacles, which enables simultaneous path following and SCA in multi-vessel scenarios. However, it assumes homogeneous vessel characteristics and does not account for communication delays. Zhai et al. [31] introduced an intelligent SCA algorithm based on Double DQN with prioritized experience replay, explicitly incorporating COLREGs and expert knowledge into the reward function. As summarized in Table 1, their composite reward employs a condition-triggered, multi-dimensional evaluation scheme that closely links the legitimacy and effectiveness of actions to regulatory compliance, which mainly consists of basic rewards (e.g., +10 for successful avoidance, -10 for collisions, -1 for potential danger, and 0 for neutral states) and additional weighted scores across five assessment dimensions. This finely tuned reward mechanism enables the agent to make decisions more closely aligned with human-like, rule-compliant SCA. Nevertheless, the discrete nature of the action space and the assumption of constant speed continue to constrain the algorithm's adaptability in complex scenarios involving variable speeds

Table 1. Overall Structure of the Composite reward in Intelligent SCA Algorithm Based on DDQN with Prioritized Experience Replay under COLREGs.

Scene type	Triggering condition	Reward value	Key parameter
Successful SCA	no collision risk	+10,$\sum(Wi\times Ri)$	Sub-reward weights：w_1=1,w_2=1, w_3=0.5,w_4=1 ,w_5=1.5
Collision penalty	distance to the target vessel d_i<0.3nm	-10	Critical distance threshold: 0.3 NM
Potential risk penalty	Predicting danger	-1	Observation vector dimension: 35 danger sectors
Neutral state	No significant events	0	N/A

More recently, to overcome the limitations of previous methods that relied heavily on idealized assumptions and thus lacking applicability in real-world scenarios, several studies have begun exploring a novel paradigms that combine deep learning with DRL framework.Woo and Kim [15] proposed a method based on visual state representation using Convolutional Neural Networks (CNN) and a semi-Markov decision process (SMDP). Unlike earlier approaches that relied on kinematic parameters as input, they designed a three-layer grid map to encode the target route, dynamic obstacles, and static obstacles, leveraging the visual feature extraction capabilities of CNN to better analyze complex and ambiguous situations commonly encountered at sea. This approach addresses the shortcomings of conventional DRL frameworks under non-ideal conditions to a certain extent. However, its performance is constrained by the fixed grid resolution and relies on the accuracy of visual detection, which may limit its effectiveness in low-visibility conditions such as fog or heavy rain. Additionally, to address the limitations of existing DRL methods that rely on discrete action spaces, Sawada et al. [32] proposed a continuous-action DRL framework based on PPO and Long Short-Term Memory (LSTM) networks, referred to as Inside OZT. By expanding the obstacle risk zones and employing LSTM units to process sequential data, Inside OZT can effectively capture near-field vessels information and enhance the SCA safety margins. However, Inside OZT incurs high computational complexity and depends on fixed risk zone parameters. In addition, the introduction of a continuous action space results in notable instability in heading control, which may lead to unpredictable risks.

Despite the considerable progress brought by the introduction of DRL framework to dynamic SCA methods in handling moving obstacles and regulatory constraints, they remain constrained by several fundamental limitations. Most existing methods still rely on idealized assumptions such as homogeneous vessel characteristics, fixed action or observation spaces, and perfect perception, which reduce their robustness and generalizability in highly dynamic and uncertain maritime environments. Moreover, the inherent complexity of multi-vessel interactions and the unpredictability of non-ego vessel behaviors in high-traffic scenarios often make DRL choose solution of suboptimal decision-making or reduced safety margins. These challenges underscore the urgent need for more scalable and adaptive solutions, thereby motivating the exploration of collaborative SCA methods based on multi-agent reinforcement learning paradigms.

4. Multi-Agent Collision Avoidance

With the increasing complexity of maritime traffic and the growing presence of autonomous vessels, single-agent collision avoidance approaches are often insufficient to ensure safety and efficiency in dense or highly interactive scenarios. Consequently, in recent year, research attention has been slowly shifting towards multi-agent SCA paradigms, which explicitly consider the interactions, cooperation, or competition among multiple autonomous vessels. The core objective is to develop scalable and adaptive strategies that account for the collective dynamics and decentralized decision-making inherent in real-world maritime environments. This section aims to review the recent advancements in multi-agent collision avoidance, focusing on both collaborative and non-collaborative settings and highlighting the unique challenges and opportunities introduced by multi-agent interactions.

4.1. Collision Avoidance in Collaborative Environment

Research on multi-agent collision avoidance in collaborative environments emphasizes the development of strategies that enable multiple vessels to share information, coordinate actions, and jointly optimize navigation decisions. The primary objective is to enhance overall safety and efficiency by leveraging inter-vessel communication and cooperation, thereby addressing the challenges posed by dense traffic, complex interactions, and dynamic maritime conditions. These approaches lay the groundwork for intelligent and scalable maritime navigation systems capable of adaptive and collective decision-making.

As one of the early explorations of MARL for SCA, Chen et al.[34] pioneered a collaborative approach where each vessel was modeled as an independent agent controlled by the DQN algorithm. To enhance practicability, they integrated the Mathematical Model Group vessel motion model and introduced cooperation coefficients to distinguish between three types of collaborative relationships, demonstrating effectiveness in COLREGs-compliant scenarios. However, their work was constrained to simple two-vessel encounters, did not consider static obstacles, wind, or current disturbances, and utilized a limited action space without speed adjustments. Furthermore, the training complexity increased sharply when scaling to more than two vessels. To overcome some of these limitations, Wen et al.[35] extended multi-agent cooperation to the joint optimization of dynamic navigation and area assignment for multiple USVs, leveraging a Multi-Agent Deep Deterministic Policy Gradient algorithm. This framework broke the conventional separation of trajectory optimization, obstacle avoidance, and task coordination, enabling integrated multi-objective optimization. However, their study did not deeply embed COLREGs rules, relied solely on the Gym simulation platform[40] without real-world validation, and exhibited insufficient robustness when facing dynamic obstacles.

Recently, based on previous work, some researchers have begun to explore hybrid architectures that combine traditional methods with MARL framework. Nantogma et al.[36] addressed multi-USV dynamic navigation and target capture in complex environments by proposing a hybrid MARL framework with heuristic guidance. They generated navigation subgoals using expert knowledge and utilized an immune network-based model for subgoal selection, followed by actor-critic PPO [14]for policy learning. While this approach improved the interpretability of group coordination, it assumed independent subgoals and only handled single intruders with regularized escape strategies, limiting its adaptability to more complex interactions. Concurrently, Gu et al.[37] introduced a virtual leader to calibrate USV formation, enhanced Multi-agent DDPG[13] with clipped double Q-learning and target policy smoothing to mitigate overestimation, and incorporated an artificial potential field into the reward function for better SCA awareness. Nevertheless, their work involved only limited vessels'formation switching scenarios, lacked analysis of communication delays, and required faster SCA responses in dense obstacle environments.

More recently, with the continuous progress of MARL algorithms, many studies have begun to focus on specific pain points such as local minima and partial observability in complex obstacle environments. Zheng et al.[21] presented a three-layer hierarchical deep reinforcement learning method (ITDRL3) to tackle multi-agent collaborative navigation in U-shaped obstacle scenarios. By decoupling target selection, right-turn strategies, and close-range avoidance, ITDRL3 reduced both the action space and training time. However, it was optimized primarily for U-shaped obstacles, exhibited poor adaptability to other complex scenarios, and saw diminished training efficiency with more than five agents. To address the challenge of task complexity and learning efficiency, Zhang et al.[39] decomposed multi-USV planning into task allocation and autonomous SCA using a value decomposition network and improved temporal difference learning. They also introduced hierarchical and regional division mechanisms to limit the exploration space. Yet, this method presupposed complete information sharing, did not account for incomplete information scenarios, thus insufficiently modeled dynamic environmental disturbances. To address the inherent partial observability challenges of dynamic environments, Wang et al.[22] combined the Deep Recurrent Q-network (DRQN) with the Decentralized Partially Observable Markov Decision Process (Dec-POMDP). They improved the DRQN update mechanism and adopted a centralized training and decentralized execution architecture, validating their approach in real-world waters and demonstrating superiority over DQN, DDQN, and APF baselines. Nonetheless, computational complexity increased substantially with more than four agents, the maneuverability differences between USVs of various tonnages were neglected, and further refinement of collaborative rules was needed for complex scenarios.

Despite significant progress, current collaborative multi-agent collision avoidance methods remain constrained by limited scalability, insufficient robustness to environmental uncertainties, and inadequate handling of partial observability and heterogeneous vessel dynamics. Future research should focus on developing more generalizable and adaptive frameworks that can effectively address these challenges, thereby enabling safe and efficient deployment of autonomous vessels in complex, real-world maritime environments.

doi:10.20944/preprints202511.0267.v1

4.2. Collision Avoidance in Non-Collaborative Environment

Unlike collaborative SCA, non-collaborative multi-agent collision avoidance focuses on scenarios where vessels navigate independently, often without explicit communication or coordination. The core challenge lies in designing robust strategies to cope with the unpredictable behavior of other agents and the inherent uncertainty in decentralized decision-making. Research in this area aims to improve the safety and adaptability of autonomous vessels operating in competitive, adversarial, or information-constrained maritime environments.

As one of the early foundational work, Liu et al.[41] pioneered the application of DRL to multi-vessel non-collaborative SCA, constructing simulated environments for both open waters and narrow waterways and integrating vessel motion characteristics to design a hierarchical state space, discrete action space, and scenario-dependent reward function. While their approach demonstrated effectiveness for dynamic SCA in two- and three-vessel scenarios, several limitations were evident: decision confusion arose in multi-vessel encounters due to the lack of prioritized avoidance responsibility, managing curved narrow waterways proved difficult owing to the challenge of balancing boundary and dynamic vessel constraints, and the absence of COLREGs integration led to non-compliant maneuvers. With these insights, Jiang et al.[42] addressed the insufficient risk assessment accuracy in multi-vessel environments by introducing an attention-based deep reinforcement learning (ADRL) algorithm. This method decomposed SCA into two modules: a risk assessment module, which simulated the attention distribution of human officers via an 8×8 local map to generate real-time attention weights, and a motion planning module, which combined supervised and reinforcement learning to accelerate exploration using historical avoidance data. This approach improved risk differentiation and learning efficiency, but the action space remained limited to heading adjustments without speed control, and the unpredictable behaviors of target vessels were not modeled, thus reducing robustness in dynamic coastal waters. Recognizing the need for enhanced COLREGs integration and scenario specificity, Guan et al.[43] proposed a Generalized Behavior Decision-Making (GBDM) model based on Q-learning. They introduced the Obstacle Zone by Target (OZT) to quantify collision risk, employed a 20×8 grid sensor to vectorize environmental data, and developed a navigation situation judgment mechanism to distinguish stand-on from give-way vessels—ensuring only give-way vessels performed avoidance and rewarding starboard turns to comply with COLREGs. While trained across a wide range of multi-vessel scenarios, this model did not account for vessel maneuverability differences and relied on fixed grid parameters, limiting its applicability in long-range open-water situations.

Recently, focusing on specialized COLREGs scenarios, Li et al.[44] targeted the risk of reduced closest point of approach in starboard-to-starboard head-on encounters. Leveraging DQN, the study defined state spaces by inter-vessel distance, speeds, and heading, and designed a discrete action space of turning angles, integrating closest point of approach (CPA) metrics into the reward function. Although this improved timing and angle of avoidance in specific scenarios, the solution was restricted to two-vessel encounters and failed to consider environmental disturbances, leading to idealized assumptions and limited generalizability. To address algorithmic shortcomings such as Q-value overestimation, Niu et al.[45] proposed an improved approach combining Double DQN and Prioritized Experience Replay (PER). By constructing a multi-agent system with up to 12 vessels and simplifying the state space via a grid-based quantification of danger areas, the model achieved more stable learning and balanced safety and economy through a multi-metric reward function. Nevertheless, the approach still relied on simplified vessel models and PER prioritization based solely on temporal difference error. Addressing the challenge of incomplete sensor information in real maritime environments, Zheng et al.[46] introduced the PPO for Partially Observable Markov Decision Process (POMDP) with guidelines under dense reward. Using high-resolution local images as state input and a dense reward function to accelerate training, combined with a route guidance mechanism to preserve original courses, this method enhanced adaptability in mixed-obstacle, multi-vessel scenarios. However, dependence on visual sensors and risks of local optimality persisted.

Most recently, in response to the path redundancy and local planning inefficiencies of traditional algorithms, Shen et al.[47] proposed the Differential Evolution Deep Reinforcement Learning (DEDRL) algorithm, integrating Differential Evolution (DE) with DQN within a two-layer optimization framework. Global path planning utilized DQN for collision-free path search and DE for node optimization, while local planning applied a course-punishing reward mechanism with a quaternion vessel domain and COLREGs compliance. Although this approach improved global-local coordination and regulatory adherence, DE-induced computational latency and reliance on global path direction for local turning angles reduced real-time responsiveness in rapidly changing multi-vessel encounters.

Despite notable progress, non-collaborative multi-agent collision avoidance methods remain fundamentally constrained by the lack of explicit communication and coordination, which often leads to conservative or conflicting maneuvers, increased path redundancy, and suboptimal navigation efficiency in dense traffic. Additionally, these methods struggle to accurately predict and respond to the highly dynamic and sometimes adversarial behaviors of other vessels, especially under partial observability and heterogeneous decision-making strategies. Future research should prioritize the development of adaptive algorithms that can better infer intent, dynamically adjust to unpredictable interactions, and robustly balance safety with efficiency in complex, large-scale maritime environments.

5. Challenges and Future Directions

5.1. Limitations and Open Challenges of Present RL Solutions

To date, reinforcement learning–based SCA has achieved encouraging progress in both single-agent and multi-agent frameworks. Nevertheless, despite these advancements, existing approaches remain fundamentally limited in scalability, adaptability, and robustness when facing the complexities of real-world maritime environments. Their performance still depends heavily on simplified assumptions, idealized sensing conditions, and static interaction models, restricting their practical deployment in dynamic, uncertain, and heterogeneous multi-vessel scenarios. The main limitations and open challenges can be summarized as follows:

▪ **Scalability and Generalization Gap.** Single-agent RL methods achieve good performance in simplified settings but fail to generalize to dynamic, large-scale maritime environments. Their reliance on ideal assumptions—such as perfect perception or homogeneous vessel models—limits scalability and adaptability across diverse operational scenarios.

▪ **Real-Time Decision Constraints.** Due to noisy and heterogeneous sensing data, computational complexity of DRL and MARL, real-time safety-critical decision-making at sea is difficult to achieve.

▪ **Insufficient Environmental Adaptation.** Most RL approaches overlook real-world disturbances like wind, current, and sensor noise, while regulatory RL rules are often simplified into static rewards. Such abstraction reduces robustness and interpretability in complex, uncertain maritime conditions.

▪ **Deficient Collaboration in MARL.** MARL enables distributed cooperations, but they usually were designed upon unrealistic communication conditions. Without the reasoning ability, vessel agents cannot infer non-collaborative ones' intentions. These weaknesses frequently result in conservative or conflicting maneuvers in real world dense traffic scenarios.

Considering the above limitations, these challenges suggest that existing RL-based approaches are reaching their capacity limits in handling the cognitive complexity of real-world maritime decision-making. Overcoming these issues requires a new paradigm that can integrate perception, reasoning, and regulation in a unified, interpretable, and adaptive manner.

5.2. LLM Enhanced Decision for SCA

doi:10.20944/preprints202511.0267.v1

Recent advances in RL have driven increasingly sophisticated applications of autonomous agents in the maritime domain, characterized by enhanced collaborative perception and distributed decision-making capabilities. However, despite these notable achievements, a fundamental limitation persists: the absence of a unified, robust, and scalable architecture that capable of integrating perception, adaptive interaction modeling, and regulatory compliance in highly dynamic and uncertain maritime environments.

To address these challenges, a promising direction for future research is the design of deeply decoupled, hierarchical decision-making architectures for multi-agent cooperative SCA.

As illustrated in Figure. 3, a typical multi-agent collision avoidance process begins when vessel agent A1 detects an approaching vessel B1 through perception. Agent A1 then shares this risk information with agent A2 via vessel-to-vessel (V2V) communication. Upon receiving the information, A2 integrates it with its own trajectory planning and the COLREGs rules to assess the potential collision risk and, if necessary, executes a give-way maneuver, while A1 and B1 maintain stand-on status. This procedure highlights the potential of multi-agent collaboration for achieving joint situational awareness and coordinated decision-making.

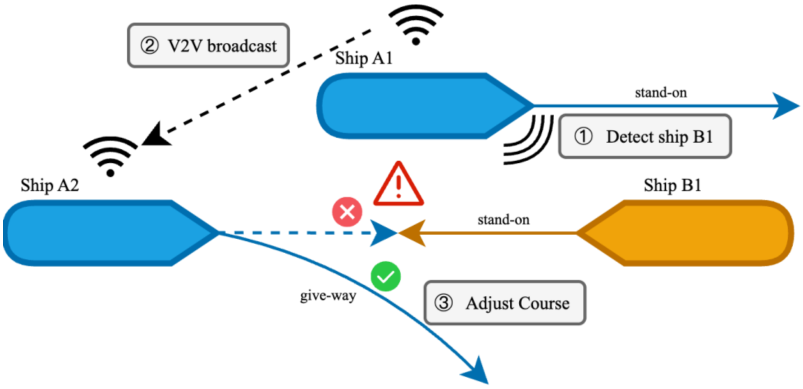


Figure 3. A schematic diagram of multi-agent SCA.

This general process can be further abstracted into hierarchical and modular architecture, as depicted in Figure. 4. The proposed architecture consists of four principal layers.

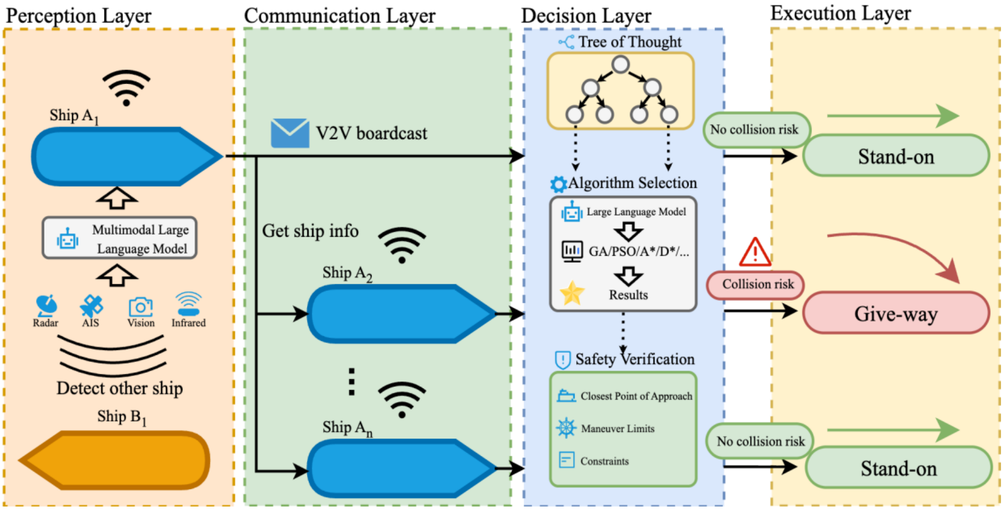


Figure 4. Highly decoupled hierarchical multi-agent collaborative architecture for multi-agent SCA.

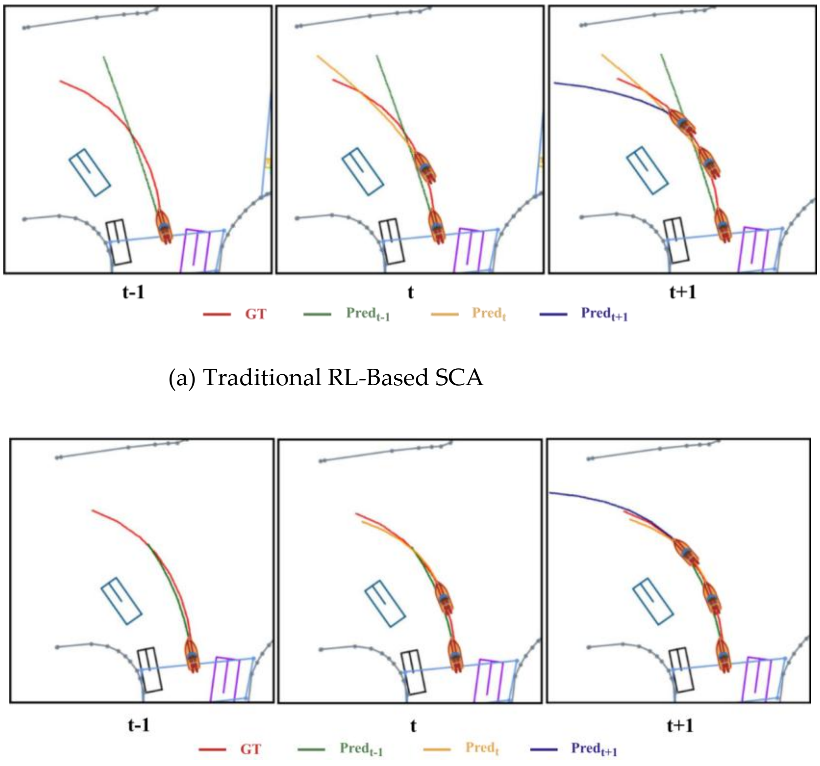
Perception Layer. The primary objective of the perception layer is to transform raw observational data into actionable features. Within our proposed architecture, information from radar, AIS, optical, and infrared sensors is no longer merely listed as discrete coordinates and velocities. Instead, we integrate a multi-modal Large Language Model (MLLM) to convert these heterogeneous data into industry-standard, natural language summaries of the maritime scene. Specifically, these summaries are enriched with structured fields, such as bearing, range, relative course, confidence intervals, and potential conflict points, and explicitly indicate which vessels occupy particular regions, their headings, whether their separation distances are decreasing, and if any vessels are approaching restricted areas. Notably, in situations where data ambiguity arises due to factors such as sea fog, solar glare, or wave clutter, this approach enables explicit delineation of uncertain areas, overcoming the limitations of traditional frameworks that often provide misleadingly precise numerical estimates.

Communication Layer. The communication layer is responsible for sharing the spatial distribution and behavioral intent of maritime agents across the multi-vessel environment, with the critical objective of achieving maximal information clarity and utility through minimal data transmission. To this end, a dual strategy of structured templates and semantic compression is adopted: every vessel broadcasts information using a standardized template comprising conclusions, evidence, key parameters, intent, and operational constraints, ensuring semantic alignment across various agents. Additionally, a lightweight semantic aggregation mechanism is introduced to synthesize consensus descriptions from repeated inter-vessel exchanges, minimizing redundant broadcasts and enhancing information consistency. When conflicting information arises, the communication layer triggers a minimal confirmation protocol, which leverages rule-based and temporal reasoning to rapidly achieve agreement, ensuring global coordination, determinacy, and predictability.

Decision Layer. Building on the structured information provided by the perception and communication layers, the decision layer is tasked with evaluating collision risk and generating avoidance strategies. Specifically, to enhance reasoning consistency and interpretability, the decision layer integrates a Tree-of-Thought (ToT) reasoning mechanism[49], using the LLM as the logical control core. In ToT reasoning, potential maneuver paths are organized as branching structures, with each node mapping actions, rationales, and supporting evidence in the context of triggers, applicable rules, and operational constraints. Here, the LLM is not used for natural language generation, but as an engine for logical planning and regulatory integration, establishing consistent reasoning chains among rules, constraints, and situational data, and yielding a structured, traceable set of candidate solutions. Multi-criteria assessment is then performed, and hierarchical logic is applied to filter and validate feasible plans.

Execution Layer. The execution layer translates decision results into low-level control commands and continuously monitors and adjusts execution in real time. Based on vessel dynamics, the control module converts heading and speed directives into rudder and propulsion commands, and closes the loop for deviation correction. To accommodate dynamic environmental changes, the execution layer integrates rolling optimization and feedforward compensation; if deviations exceed thresholds due to external disturbances, the system automatically initiates local refinements or replans maneuvers to maintain continuity and safety margins.

To verify the feasibility of the aforementioned framework, we conducted preliminary experiments to verify its feasibility and necessity in ship collision-avoidance tasks. The results are shown in Figure 5. Traditional RL methods rely on limited state features for policy updates and often fail to capture inter-ship interactions and rule constraints, resulting in unstable and biased trajectory predictions over time. With the introduction of an LLM-based Agent, the system performs semantic-level reasoning on the navigational context before decision-making, extracting latent intentions and risk cues to provide high-level semantic priors for RL. The experimental results indicate that this framework effectively enhances the stability of decision-making and the accuracy of trajectory prediction, demonstrating the practical necessity of integrating LLM-based Agents into intelligent maritime collision-avoidance decision systems.



(a) Traditional RL-Based SCA

(b) LLM-Agent Guided SCA

Figure 5. Comparison between the traditional RL-based collision avoidance method and the proposed LLM-Agent guided decision framework.

In conclusion, through the integration of multi-modal perception, semantic communication, hierarchical decision-making, and closed-loop execution, the LLM enhanced architecture endows vessel agents with self-perception, self-understanding, self-decision, and self-regulation capabilities. The multi-modal LLM will serve as the core cognitive unit, delivering semantic-level maritime scene comprehension in the perception layer, structured intent sharing in the communication layer, unified reasoning and optimization in the decision layer via ToT and computational reasoning, and real-time adaptive control in the execution layer. Inter-layer information flow and constraint feedback are tightly coupled through semantic alignment and data communication, ensuring consistent situational understanding and stable decision-making even in highly dynamic maritime environments. Thus, LLM-based Agents seem to be very promising in future ship collision avoidance research and practical autonomous systems.

Abbreviations

The following abbreviations are used in this manuscript:

AABSRL	Avoidance Algorithm Based on Sarsa on-policy Reinforcement Learning
ADRL	Attention-Based Deep Reinforcement Learning
ANOA	Autonomous Navigation and Obstacle Avoidance
CDRLOA	Concise DRL Obstacle Avoidance
CNN	Convolutional Neural Networks
COLREGs	Convention on The International Regulations for Preventing Collisions At Sea
CPA	Closest Point of Approach
DDPG	Deep Deterministic Policy Gradient
DE	Differential Evolution
DEDRL	Differential Evolution Deep Reinforcement Learning
Dec-POMDP	Decentralized Partially Observable Markov Decision Process

DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
DRQN	Deep Recurrent Q-network
GBDM	Generalized Behavior Decision-Making
HRVO	Hybrid Reciprocal Velocity Obstacle
ITDRL3	Three-layer Hierarchical Deep Reinforcement Learning
LLM	Large Language Model
LSTM	Long Short-Term Memory
MARL	Multi-Agent Reinforcement Learning
MDP	Markov Decision Process
MLLM	Multi-Modal LLM
MPC	Model Predictive Control
OZT	Obstacle Zone by Target
PER	Prioritized Experience Replay
POMDP	Partially Observable Markov Decision Process
PPO	Proximal Policy Optimization
RL	Reinforcement Learning
RVO	Reciprocal Velocity Obstacle
SCA	Ship Collision Avoidance
SMDP	Semi-Markov Decision Process
ToT	Tree-of-Thought
UANOA	Autonomous Navigation and Obstacle Avoidance in USVs
UMVs	Underactuated Unmanned Vessels
USVs	Unmanned Surface Vehicles
V2V	Vessel-to-Vessel
VO	Velocity Obstacle
WOS	Web of Science

References

1. Uflaz, E.; Akyuz, E.; Arslan, O. et al. Analysing human error contribution to ship collision risk in congested waters under the evidential reasoning SPAR-H extended fault tree analysis. Ocean Engineering, 2023, 287, 115758.
2. Ma, L.; Ma, X.; Wang, T. et al. A data-driven approach to determine the distinct contribution of human factors to different types of maritime accidents. Ocean Engineering, 2024, 295,116874.
3. Guo, X.; qian, Zheng, Q.; Guo, Y. Maritime accident causation: A spatiotemporal and HFACS-Based approach. Ocean Engineering, 2025, 340, 122329.
4. Fiorini, P.; Shiller, Z. Motion planning in dynamic environments using velocity obstacles. The international journal of robotics research, 1998, 17(7), 760-772.
5. Van, den, Berg, J.; Lin, M.; Manocha, D. Reciprocal velocity obstacles for real-time multi-agent navigation. In Proceedings of 2008 IEEE international conference on robotics and automation, Pasadena, CA, USA, 19-23 May 2008; pp. 1928-1935.
6. Snape, J.; Van, Den, Berg, J.; Guy, S. J. et al. The hybrid reciprocal velocity obstacle. IEEE Transactions on Robotics, 2011, 27(4), 696-706.
7. Kao, S. L.; Lee, K. T.; Chang, K. Y. et al. A fuzzy logic method for collision avoidance in vessel traffic service. The journal of navigation, 2007, 60(1), 17-31.
8. Caldwell, C. V.; Dunlap, D. D.; Collins, E. G. Motion planning for an autonomous underwater vehicle via sampling based model predictive control. In Proceedings of OCEANS 2010 MTS/IEEE SEATTLE, Seattle, WA, USA, 20-23 September 2010; pp.1-6.
9. Kaelbling, L. P.; Littman, M. L.; Moore, A. W. Reinforcement learning: A survey. Journal of artificial intelligence research, 1996, 4, 237-285.
10. Puterman, M. L. Markov decision processes. Handbooks in operations research and management science, 1990, 2, 331-434.
11. Watkins C J C H, Dayan P. Q-learning. Machine learning, 1992, 8(3), 279-292.

12. Mnih, V.; Kavukcuoglu, K.; Silver, D. et al. Human-level control through deep reinforcement learning. nature, 2015, 518(7540), 529-533.
13. Lillicrap, T. P.; Hunt, J. J.; Pritzel, A. et al. Continuous control with deep reinforcement learning. arXiv 2015, arXiv:1509.02971.
14. Schulman, J.; Wolski, F.; Dhariwal, P.; et al. Proximal policy optimization algorithms. arXiv 2017, arXiv:1707.06347.
15. Woo, J.; Kim, N. Collision avoidance for an unmanned surface vehicle using deep reinforcement learning. Ocean Engineering, 2020, 199, 107001.
16. Meyer, E.; Heiberg, A.; Rasheed, A. et al. COLREG-compliant collision avoidance for unmanned surface vehicle using deep reinforcement learning. Ieee Access, 2020, 8, 165344-165364.
17. Shen, H.; Hashimoto, H.; Matsuda, A. et al. Automatic collision avoidance of multiple ships based on deep Q-learning. Applied Ocean Research, 2019, 86, 268-288.
18. Busoniu, L.; Babuska, R.; De, Schutter. B. Multi-agent reinforcement learning: A survey. In Proceedings of 2006 9th international conference on control, automation, robotics and vision, Singapore, 05-08, December 2006; pp.1-6.
19. Chen, C.; Ma, F.; Xu, X. et al. A novel ship collision avoidance awareness approach for cooperating ships using multi-agent deep reinforcement learning. Journal of Marine Science and Engineering, 2021, 9(10), 1056.
20. Yoshioka, H.; Hashimoto, H.; Matsuda, A. Artificial Intelligence for Cooperative Collision Avoidance of Ships Developed by Multi-Agent Deep Reinforcement Learning. In Proceedings of International Conference on Offshore Mechanics and Arctic Engineering, Singapore, 09-14, June, 2024;pp. 125392-125400.
21. Zheng, S.; Luo, L. Multi-Agent Cooperative Navigation with Interlaced Deep Reinforcement Learning. In Proceedings of 2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Hangzhou, China, 23-25, October, 2024; pp. 409-414.
22. Wang, Z.; Chen, P.; Chen, L. et al. Collaborative Collision Avoidance Approach for USVs Based on Multi-Agent Deep Reinforcement Learning. IEEE Transactions on Intelligent Transportation Systems, 2025, 26(4), 4780-4794.
23. Zhang, R.; Tang, P.; Su, Y. et al. An adaptive obstacle avoidance algorithm for unmanned surface vehicle in complicated marine environments. IEEE/CAA Journal of Automatica Sinica, 2014, 1(4), 385-396.
24. Yang, Y.; Pang, Y.; Li, H. et al. Local path planning method of the self-propelled model based on reinforcement learning in complex conditions. Journal of Marine Science and Application, 2014, 13(3), 333-339.
25. Bhopale, P.; Kazi, F.; Singh, N. Reinforcement learning based obstacle avoidance for autonomous underwater vehicle. Journal of Marine Science and Application, 2019, 18(2), 228-238.
26. Wu, X.; Chen, H.; Chen, C. et al. The autonomous navigation and obstacle avoidance for USVs with ANOA deep reinforcement learning method. Knowledge-Based Systems, 2020, 196, 105201.
27. Yan, N.; Huang, S.; Kong, C. Reinforcement Learning-Based Autonomous Navigation and Obstacle Avoidance for USVs under Partially Observable Conditions. Mathematical Problems in Engineering, 2021, 2021(1), 5519033.
28. Cheng, Y.; Zhang, W. Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels. Neurocomputing, 2018, 272, 63-73.
29. Shen, H.; Hashimoto, H.; Matsuda, A. et al. Automatic collision avoidance of multiple ships based on deep Q-learning. Applied Ocean Research, 2019, 86, 268-288.
30. Zhao, L.; Roh, M. I. COLREGs-compliant multiship collision avoidance based on deep reinforcement learning. Ocean Engineering, 2019, 191, 106436.
31. Zhai, P.; Zhang, Y.; Shaobo, W. Intelligent ship collision avoidance algorithm based on DDQN with prioritized experience replay under COLREGs. Journal of Marine Science and Engineering, 2022, 10(5), 585.
32. Alayrac, J. B.; Donahue, J.; Luc, P. et al. Flamingo: a visual language model for few-shot learning. Advances in neural information processing systems, 2022, 35, 23716-23736.
33. Sawada, R.; Sato, K.; Majima, T. Automatic ship collision avoidance using deep reinforcement learning with LSTM in continuous action spaces. Journal of Marine Science and Technology, 2021, 26(2), 509-524.

34. Chen, C.; Ma, F.; Xu, X. et al. A novel ship collision avoidance awareness approach for cooperating ships using multi-agent deep reinforcement learning. Journal of Marine Science and Engineering, 2021, 9(10), 1056.
35. Wen, J.; Liu, S.; Lin, Y. Dynamic navigation and area assignment of multiple USVs based on multi-agent deep reinforcement learning. Sensors, 2022, 22(18), 6942.
36. Nantogma, S.; Zhang, S.; Yu, X. et al. Multi-USV dynamic navigation and target capture: A guided multi-agent reinforcement learning approach. Electronics, 2023, 12(7), 1523.
37. Gu, Y.; Wang, X.; Cao, X. et al. Multi-USV Formation Control and Obstacle Avoidance Under Virtual Leader.In Proceedings of 2023 China Automation Congress (CAC), Chongqing, China, 17-19, November 2023; pp. 3411-3416.
38. Glorennec, P. Y.; Jouffe, L.; Fuzzy Q-learning. In Proceedings of 6th international fuzzy systems conference, Barcelona, Spain, 05-05, July, 1997;pp. 659-662.
39. Zhang, J.; Ren, J.; Cui, Y. et al. Multi-USV task planning method based on improved deep reinforcement learning. IEEE Internet of Things Journal, 2024, 11(10), 18549-18567.
40. Brockman, G.; Cheung, V.; Pettersson, L. et al. Openai gym. arXiv 2016, arXiv:1606.01540.
41. Liu, J.; Xiao, Y. Intelligent ships collision avoidance and navigation method based on deep reinforcement learning.In Proceedings of 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI), Kunming, China, 17-19, September, 2021;pp. 573-578.
42. Jiang, L.; An, L.; Zhang, X.; et al. A human-like collision avoidance method for autonomous ship with attention-based deep reinforcement learning. Ocean Engineering, 2022, 264, 112378.
43. Guan, W.; Zhao, M.; Zhang, C. et al. Generalized behavior decision-making model for ship collision avoidance via reinforcement learning method. Journal of Marine Science and Engineering, 2023, 11(2), 273.
44. Li, H.; Weng, J.; Zhou, Y. Ship collision avoidance method in starboard-to-starboard head-on situations.In Proceedings of 2023 7th International Conference on Transportation Information and Safety (ICTIS), Xi'an, China, 04-06, August, 2023;pp. 609-614.
45. Niu, Y.; Zhu, F.; Zhai, P. An autonomous decision-making algorithm for ship collision avoidance based on DDQN with prioritized experience replay. In Proceedings of 2023 7th International Conference on Transportation Information and Safety (ICTIS), Xi'an, China, 04-06, August, 2023;pp.1174-1180.
46. Zheng, K.; Zhang, X.; Wang, C. et al. A partially observable multi-ship collision avoidance decision-making model based on deep reinforcement learning. Ocean & Coastal Management, 2023, 242, 106689.
47. Shen, Y.; Liao, Z.; Chen, D. Differential Evolution Deep Reinforcement Learning Algorithm for Dynamic Multiship Collision Avoidance with COLREGs Compliance. Journal of Marine Science and Engineering, 2025, 13(3), 596.
48. Singh, S. P.; Sutton, R, S. Reinforcement learning with replacing eligibility traces. Machine learning, 1996, 22(1), 123-158.
49. Yao, S.; Yu, D.; Zhao, J. et al. Tree of thoughts: Deliberate problem solving with large language models. Advances in neural information processing systems, 2023, 36, 11809-11822.

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.