Journal of
Marine Science and Engineering

MDPI

Article

Ship Collision Avoidance Decision-Making Using Multi-Agent Deep Reinforcement Learning with MMG Manoeuvring Dynamics

Junheng Zhao [1,2], Jiongjiong Liu [1,2], Jinfen Zhang [1,2,3,*], Zhepeng Han [1,2] and Wuliu Tian [3,4]

[1] State Key Laboratory of Maritime Technology and Safety, Wuhan University of Technology, Wuhan 430062, China; 361612@whut.edu.cn (J.Z.); jiongjiongliu@whut.edu.cn (J.L.); hanzhepeng@whut.edu.cn (Z.H.)
[2] Intelligent Transportation System Research Center, Wuhan University of Technology, Wuhan 430062, China
[3] Guangxi Laboratory of Modern Canal, Nanning 530029, China; tianwuliu@bbgu.edu.cn
[4] Maritime College, Beibu Gulf University, Qinzhou 535000, China
* Correspondence: jinfen.zhang@whut.edu.cn; Tel.: +86-13476860906

Abstract

With the rapid advancement of Maritime Autonomous Surface Ships (MASSs), developing intelligent decision-making systems that ensure navigation safety in complex waters has become a core priority for the maritime industry. To address the limitations of oversimplified ship dynamics, the instability in game-based strategies, and the presence of non-compliant ships in multi-ship encounters, a novel decision-making framework is developed based on Multi-Agent Deep Reinforcement Learning (MADRL). A three-degree-of-freedom (3-DOF) manoeuvring modelling group (MMG) model is incorporated to replace conventional constant-speed assumptions. By explicitly modelling the hydrodynamic forces acting on the hull, propeller, and rudder, the proposed framework captures the intrinsic coupling between the speed and heading, thereby ensuring that the generated manoeuvres conform to the physical and operational constraints. To achieve stable decision-making in multi-ship encounters, an MATD3-based decision-making module is integrated within a Centralised Training and Distributed Execution (CTDE) paradigm. This architecture enables ships to derive robust and decentralised policies, while benefiting from global information during training. In addition, the proposed method demonstrates a promising adaptability in the investigated multi-ship encounter scenarios. Simulations are conducted across a set of encounter scenarios restricted to open waters. The results demonstrate that the proposed framework achieves safe collision avoidance performance while generating COLREGs-consistent behaviours in basic encounters involving standard power-driven vessels under open-water conditions.

Keywords: multi-ship encounter; collision avoidance; deep reinforcement learning; COLREGs; MMG dynamic model

Check for updates

Academic Editor: Chia-Hung Yeh

Received: 24 June 2026
Revised: 20 July 2026
Accepted: 23 July 2026
Published: 24 July 2026

1. Introduction

Maritime transport constitutes a fundamental pillar of global economic development, accounting for more than 80% of international trade by volume [1,2]. With the rapid development of the global shipping industry, increasingly, traffic density and the growing number of ships have rendered the navigation environment progressively more congested and complex [3]. Ship collisions remain one of the most critical safety concerns in maritime traffic, which sometimes result in catastrophic consequences, including severe loss of life, substantial economic damage, and irreversible marine environmental pollution [4–6].

Statistical analyses indicate that human factors, such as misjudgement situations, inadequate communication, and the limited decision-making capability of watchkeeping officers, contribute to approximately 60.6% of maritime accidents [7,8]. Advances in artificial intelligence and automation technologies have accelerated the development of unmanned surface vehicles (USVs). By leveraging high-precision perception, cognition, and autonomous decision-making, USVs offer the potential to significantly reduce human-induced errors and enhance navigational safety [9]. However, in complex multi-ship encounter scenarios, conventional collision avoidance methods often struggle to adapt to dynamically evolving conditions [10,11]. Consequently, the development of collision avoidance algorithms with a high autonomy and strong coordination capability is essential for next-generation intelligent waterborne transport systems [12].

Despite the substantial progress in ship collision avoidance, significant challenges remain in addressing complex multi-ship encounters. First, many conventional approaches rely on non-cooperative game assumptions, treating surrounding ships as dynamic obstacles moving at constant speeds with fixed trajectories. Such simplifications fail to capture the coupled and interactive nature of multi-ship decision-making [13]. Second, although deep-reinforcement-learning (DRL)-based methods have demonstrated good performance, many existing studies employ simplified kinematic models that neglect ship inertia and hydrodynamic effects, thereby limiting the practical applicability of the generated strategies to real ships [14]. Furthermore, a lot of research assumes ideal cooperative behaviour among ships and does not adequately consider high-risk scenarios involving non-compliant or rule-violating ships [15]. In real-world operations, the absence of appropriate avoidance actions may compromise navigational safety, highlighting the need for decision-making frameworks that can consider realistic ship dynamics and non-cooperative behaviours.

To address these limitations, this study develops a multi-ship collision avoidance decision-making framework for open-water navigation by integrating a high-fidelity ship dynamic model with an MATD3-based decision-making module. A three-degree-of-freedom (3-DOF) MMG model is incorporated to represent hull forces, propeller thrust, and rudder forces, enabling physically feasible manoeuvre generation. Meanwhile, the MATD3-based framework provides decentralised decision-making under a Centralised Training and Distributed Execution (CTDE) paradigm, while the reward design incorporates COLREGs-oriented guidance and considers non-cooperative target behaviours. The primary objective of this study is to assess the effectiveness and robustness of the proposed approach in the presence of non-cooperative disturbances under open-water navigation conditions.

The remainder of this paper is organised as follows. Section 2 reviews the relevant literature on ship collision avoidance and identifies key research gaps. Section 3 presents the proposed multi-ship collision avoidance framework, including MMG-based dynamic modelling, collision avoidance decision-making, and the implementation of the MATD3 algorithm. Section 4 validates the proposed method through a series of case studies. Finally, Section 5 concludes this study and outlines future research.

2. Literature Review

Autonomous ship collision avoidance decision-making has always been a hot topic in intelligent maritime navigation. Extensive efforts have been devoted to this field, leading to the development of a wide range of collision avoidance algorithms. Among the traditional approaches, Artificial Potential Field (APF), Velocity Obstacle (VO), and Model Predictive Control (MPC) are widely adopted. APF is commonly applied in unmanned ship path planning due to its clear mathematical formulation and low computational cost, which enables real-time implementation [16,17]. However, as a local planning method, APF is prone to becoming trapped in local minima, particularly in complex environments with multiple obstacles or dynamic encounter scenarios [18]. To overcome these limitations, VO and its variants have been introduced into ship collision avoidance. By constructing collision cones in the velocity space, VO can effectively address dynamic obstacles and has been successfully applied in USV navigation tasks [19,20]. Nevertheless, VO-based methods typically rely on linear motion assumptions, where obstacles are considered to move at constant speeds along straight trajectories [21]. This simplification neglects the substantial inertia and complex hydrodynamic characteristics of ships. In addition, VO methods often exhibit limited robustness under sensor noise and environmental disturbances, restricting their applicability in realistic maritime conditions [22]. Beyond geometric and potential-field-based methods, MPC has attracted considerable attention due to its capability to explicitly incorporate system constraints and nonlinear dynamics [23]. By solving online optimisation problems over a finite horizon, MPC can generate trajectories that comply with COLREGs while satisfying ship kinematic constraints [24]. However, its performance is highly dependent on the accuracy of the underlying model, and parameter uncertainties can significantly degrade control accuracy [25]. Moreover, solving optimisation problems for nonlinear ship dynamics imposes substantial computational demands [26]. In dense multi-ship encounter scenarios, the increasing dimensionality of the state space makes it challenging for MPC to guarantee millisecond-level real-time decision-making, thereby limiting its applicability in highly dynamic environments [27].

To alleviate the dependence of traditional methods on accurate mathematical models and to reduce computational burden, Deep Reinforcement Learning (DRL) has emerged as a promising data-driven paradigm in recent years [28]. Unlike optimisation-based approaches, DRL adopts a model-free framework and learns collision avoidance strategies through interaction with the environment [29]. Once trained, the resulting policy enables an end-to-end mapping from sensory inputs to control actions, thereby supporting real-time decision-making [30]. Early applications of DRL in ship collision avoidance primarily relied on value-based methods, such as Deep Q-Network (DQN) and its variants [31,32]. However, these approaches require the discretisation of the inherently continuous action space (e.g., heading and speed), which often leads to oscillatory control behaviour and unnatural zigzag trajectories, reducing the manoeuvring smoothness and potentially increasing the mechanical wear [33].

To address these issues, policy gradient methods have been introduced to enable the continuous control of ship motion. Among them, Proximal Policy Optimisation (PPO) and Deep Deterministic Policy Gradient (DDPG) have become the dominant approaches. PPO, as an on-policy algorithm, improves the training stability by constraining policy updates and has been widely applied in multi-ship interaction tasks [34–36]. In contrast, DDPG adopts an off-policy framework with experience replay, enabling the more efficient utilisation of historical data [37]. Given that ship simulation environments often involve computationally intensive hydrodynamic calculations, data sampling is costly. Consequently, PPO suffers from a relatively low sample efficiency, whereas off-policy methods such as DDPG and its variants are better suited to ship control applications due to their superior data efficiency [38–40].

Despite their advantages, DDPG-based approaches are inherently designed for single-agent settings. In multi-ship collision avoidance scenarios, all ships update their policies simultaneously, leading to a continuously changing environment from the perspective of each agent. This phenomenon, known as environmental non-stationarity, violates the Markov assumption and makes convergence difficult, often resulting in policy oscillation or even training failure [41,42]. To address this issue, value-based methods have been extended to multi-agent systems, leading to the development of Multi-Agent DQN

(MADQN) [43]. Based on this framework, Chen et al. [44] incorporated a profit-and-loss distribution mechanism into the reward function to encourage cooperative and rule-compliant behaviour among ships. Additionally, Niu et al. [30] proposed a PER-DDQN-based method that introduces a non-collaborative factor to model the behaviour of other ships, thereby improving robustness. For continuous action spaces, Lowe et al. [41] proposed the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm within an Actor–Critic framework, incorporating the Centralised Training and Distributed Execution (CTDE) paradigm. This approach leverages the global information during training to stabilise the value function estimation while enabling decentralised decision-making based on local observations during execution. Building upon this, Wang and Zhao [42] further introduced an inter-agent communication mechanism into the MADDPG framework, allowing ships to exchange information explicitly and thereby improving the decision-making performance in scenarios with limited visibility or incomplete information.

Although MADDPG and its extensions have significantly improved multi-agent coordination, they inherit the Q-value overestimation issue of DDPG [45]. This bias arises from the maximisation operation in value updates and may accumulate during training, leading to suboptimal or overly aggressive policies. To address this limitation, Ackermann et al. [46] extended the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm to multi-agent systems, resulting in the MATD3 algorithm. By employing twin critic networks and delayed policy updates, MATD3 effectively mitigates the overestimation bias and enhances the training stability in complex interaction scenarios.

Despite its advantages, the application of MATD3 to multi-ship collision avoidance remains constrained by several simplifying assumptions in existing studies. First, ship dynamic models are often oversimplified, neglecting inertia and physical constraints, which limits the feasibility of the generated strategies in real-world manoeuvring. Second, control inputs are frequently restricted to heading adjustments, ignoring speed variations and the strong coupling between these variables. Third, most studies assume fully cooperative environments and fail to consider non-compliant ships that do not adhere to COLREGs, which is a critical factor in practical operations.

To address these gaps, this study proposes a multi-ship collaborative collision avoidance framework integrating an MATD3-based decision-making module with high-fidelity ship dynamics modelling. The main contributions are summarised as follows:

- **High-fidelity dynamic modelling:** A 3-DOF MMG model is incorporated to explicitly represent ship forces and inertia, ensuring that collision avoidance actions remain consistent with realistic manoeuvring constraints.
- **Coupled speed–heading control:** Both speed and heading are jointly optimised, with their physical coupling explicitly modelled, thereby enabling physically feasible and flexible manoeuvring decisions during multi-ship encounter scenarios.
- **Robustness to non-cooperative behaviour:** The proposed framework demonstrates an effective collision avoidance capability in the investigated scenarios involving rule-violating ships, improving safety under non-cooperative encounter conditions.

3. Methodology and Modelling

3.1. Research Framework

This study develops a multi-ship collaborative collision avoidance framework based on a high-fidelity dynamic model and multi-agent deep reinforcement learning. The overall architecture is illustrated in Figure 1.

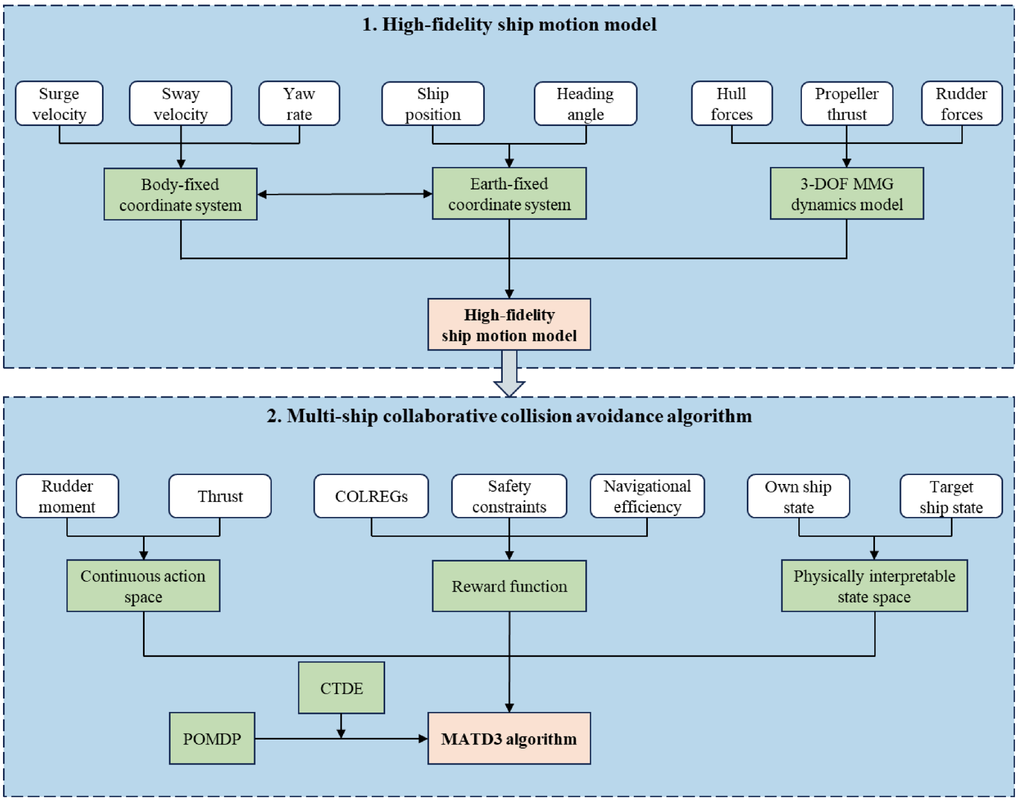

Figure 1. Flowchart of the proposed method.

To accurately characterise ship motion under complex sea conditions, Earth-fixed and body-fixed coordinate systems are first established to represent the kinematic and dynamic states of ships. A 3-DOF MMG model is then incorporated to construct high-fidelity dynamic equations, in which the nonlinear hydrodynamic coupling effects of the hull, propeller, and rudder are explicitly modelled (Section 3.2).

Based on the dynamic environment, multi-ship encounter problem is formulated as a Partially Observable Markov Decision Process (POMDP), with a continuous action space and a physically interpretable state space defined in terms of own-ship motion states and relative parameters of target ships (Section 3.3). To ensure rule-compliant and goal-oriented collision avoidance, a composite reward function is designed by integrating COLREGs requirements, safety constraints, and task-completion objectives (Section 3.4).

Finally, the Multi-Agent Twin Delayed Deep Deterministic Policy Gradient (MATD3) framework is employed to address the non-stationarity inherent in multi-agent learning. By incorporating twin critic networks and a delayed policy update mechanism, the proposed method achieves stable training and robust collaborative collision avoidance performance (Section 3.5).

3.2. Ship Motion Model

To realistically simulate the dynamic response of ships during navigation, a 3-DOF MMG approach is adopted. The simulated ship is based on the standard KVLCC2 tanker model, and the principal particulars and hydrodynamic coefficients follow the MMG standard method proposed by Yasukawa and Yoshimura [47]. Although a 6-DOF model offers a more comprehensive description of ship motion, the present study employs a 3-DOF formulation, as collision avoidance manoeuvres are predominantly confined to horizontal plane motions, namely, surge, sway, and yaw. In addition, the reduced model complexity significantly alleviates the computational burden associated with deep-reinforcement-learning training, which typically requires a large number of environment interactions to achieve convergence. Within the MMG framework, the hydrodynamic forces acting on the hull, propeller thrust, and rudder forces are modelled separately. This modular formulation enables an accurate representation of ship manoeuvring behaviour, particularly the nonlinear dynamic characteristics observed during large-angle steering.

To describe ship motion, two coordinate systems are defined: an Earth-fixed coordinate system (x_0, y_0), attached to the Earth, and a body-fixed coordinate system (x, y), located at the ship's centre of gravity. The Earth-fixed coordinate system is used to describe the ship position and heading angle ψ, whereas the body-fixed coordinate system is employed to represent the surge velocity u, sway velocity v, and yaw rate r. The transformation between these two coordinate systems is given as follows:

$$\begin{cases} \dot{x}_0 = u\cos\psi - v\sin\psi \\ \dot{y}_0 = u\sin\psi + v\cos\psi \\ \dot{\psi} = r \end{cases} \tag{1}$$

Based on rigid-body dynamics, the 3-DOF equations of motion, incorporating the effects of added mass, can be expressed as follows:

$$\begin{cases} (m+m_x)\dot{u} - (m+m_y)vr - (mx_G)r^2 = X \\ (m+m_y)\dot{v} + (m+m_x)ur + (mx_G + J_z)\dot{r} = Y \\ (I_{zz} + J_{zz})\dot{r} + (mx_G)(\dot{v} + ur) = N \end{cases} \tag{2}$$

where m represents the ship mass, m_x and m_y denote the added mass in the longitudinal and transverse directions, I_{zz} is the moment of inertia, J_{zz} denotes the added moment of inertia, and x_G represents the longitudinal coordinate of the centre of gravity. X, Y, and N represent the total forces and moment acting on the ship hull, which are modularly decomposed into three primary contributions: the ship hull (H), the propeller (P), and the rudder (R).

$$\begin{cases} X = X_H + X_R + X_P \\ Y = Y_H + Y_R \\ N = N_H + N_R \end{cases} \tag{3}$$

In the present study, the environmental disturbances caused by wind, waves, and ocean currents are not considered, and all simulations are conducted under calm-water conditions. Accordingly, the external environmental forces and moments are omitted from the current MMG formulation. The hull hydrodynamic forces are represented by nondimensional hydrodynamic derivatives as follows:

$$\begin{cases} X_H = \frac{1}{2}\rho LdU^2 X'_{\mathrm{H}}(u', v', r') \\ Y_H = \frac{1}{2}\rho LdU^2 Y'_{\mathrm{H}}(v', r') \\ N_H = \frac{1}{2}\rho LdU^2 N'_{\mathrm{H}}(v', r') \end{cases} \tag{4}$$

where ρ denotes the seawater density, L is the ship length, d represents the draft, and U denotes the resultant velocity. The variables u', v', and r' represent the non-dimensional surge velocity, sway velocity, and yaw rate, respectively. X'_{H}, Y'_{H}, and N'_{H} are non-linear functions composed of first-order and higher-order hydrodynamic derivatives.

The longitudinal thrust generated by the propeller, denoted as X_P, is treated as a control input to the system. The corresponding physical thrust T_{thrust} is obtained by mapping the normalised action output of the learning agent, and the relationship is defined as follows:

$$X_P = T_{thrust} \tag{5}$$

The forces and moments generated by the rudder depend on the effective inflow velocity and the effective angle of attack. The normal rudder force F_{N} is calculated as follows:

$$F_{\mathrm{N}} = \frac{1}{2}\rho A_R U_R^2 f_\alpha \sin(\alpha_R) \tag{6}$$

where A_R is the rudder area and α_R is the effective rudder angle.

The contributions of the rudder forces to ship motion (X_{R},Y_{R},N_{R}) are determined by the thrust deduction coefficient t_{R}, the hull transverse force coefficient a_H induced by the rudder force, and the longitudinal position of the rudder x_{R}, as follows:

$$\begin{cases} X_{\mathrm{R}} = -(1 - t_{\mathrm{R}})F_{\mathrm{N}} \sin\delta \\ Y_{\mathrm{R}} = -(1 + a_{\mathrm{H}})F_{\mathrm{N}} \cos\delta \\ N_{\mathrm{R}} = -(x_{\mathrm{R}} + a_{\mathrm{H}}x_{\mathrm{H}})F_{\mathrm{N}} \cos\delta \end{cases} \tag{7}$$

For clarity and reproducibility, the main physical parameters of the simulated ship are summarised in Table 1. The nondimensional hydrodynamic derivatives and MMG interaction coefficients are adopted from the standard KVLCC2 model reported in Ref. [47]. The actuator constraints, including the maximum propeller thrust and rudder moment, are determined according to the physical limitations implemented in the simulation environment.

Table 1. Main physical parameters of the KVLCC2 model used in simulations.

Parameter	Symbol	Value
Ship length	L	320 m
Draft	d	20.8 m
Ship mass	m	3.20×10^8 kg
Centre of gravity position	x_G	11.2 m
Water density	ρ	1025 kg/m^3
Rudder area	A_R	112.5 m^2
Maximum propeller thrust	$T_{\max}$	5.0 MN
Maximum rudder moment	$N_{\max}^{\mathrm{rud}}$	15 MN·m

3.3. Ship Decision, Action, and State Space

To enable ships to effectively perceive their surrounding environment and make accurate decisions, a high-dimensional state space is constructed, comprising both the own-ship state and the states of target ships. In addition, a continuous action space is defined to facilitate fine-grained control of ship manoeuvring behaviour.

3.3.1. Ship Decision Action Space

To achieve fine-grained control of ships, a continuous action space is adopted. The policy network outputs a normalised action vector constrained within the interval [−1,1], which is subsequently mapped to the rudder moment and propeller thrust, respectively, as follows:

$$A = (a_{\mathrm{rudder}}, a_{\mathrm{thrust}}) \tag{8}$$

To map the outputs of the policy network to the actual control inputs of the physical environment, two linear mapping functions are adopted. The rudder and thrust control are expressed in Equations (9) and (10):

$$N_{\mathrm{rudder}} = a_{\mathrm{rudder}} \cdot N_{\max}^{\mathrm{rud}} \tag{9}$$

$$T_{\mathrm{prop}} = a_{\mathrm{thrust}} \cdot T_{\max} \tag{10}$$

where N_{rudder} is the turning moment applied to the ship hull, and $N_{\max}$ is the maximum rudder moment. T_{prop} represents the longitudinal thrust generated by the propeller, and $T_{\max}$ is the maximum thrust. Furthermore, the actual inputs are subject to physical rate limits, specifically restricted to a maximum steering rate of 2.3°/s for the rudder angle and a maximum thrust variation rate of 83.33 kN/s for the propeller acceleration.

3.3.2. Ship State Space

The system state S_t consists of the own ship state vector S_{own} and the state matrix S_{target}. To enhance the convergence rate and training stability, all state variables are normalised prior to being input into the policy network.

The own ship state vector $S_{\text{own}} \in \mathbb{R}^8$ includes navigation information, kinematic parameters, and action feedback from the previous time step:

$$S_{\text{own}} = \left(x_{\text{g}}, y_{\text{g}}, \psi_{\text{o}}, u_{\text{o}}, v_{\text{o}}, r_{\text{o}}, a^{t-1}_{\text{rudder}}, a^{t-1}_{\text{thrust}}\right) \tag{11}$$

where x_{g} and y_{g} represent the longitudinal and transverse distances of the destination relative to the ship's centre of gravity, respectively. ψ_{o} denotes the current heading angle of the ship, while u_{o}, v_{o}, and r_{o} are the longitudinal velocity, transverse velocity, and yaw rate in the body-fixed coordinate system. a^{t-1}_{rudder} and a^{t-1}_{thrust} refer to the normalised rudder and thrust actions executed by the ship in the previous time step. For each target ship i in the environment, the state vector $S^i_{\text{target}} \in \mathbb{R}^7$ observed by the own ship includes relative position, relative motion, and collision risk indicator:

$$S^i_{\text{target}} = \left(x^i_{\text{rel}}, y^i_{\text{rel}}, C^i_{\text{rel}}, v^i_{x\_\text{rel}}, v^i_{y\_\text{rel}}, \text{DCPA}^i, \text{TCPA}^i\right) \tag{12}$$

where x^i_{rel} and y^i_{rel} represent the relative position coordinates of the target ship relative to the own ship. C^i_{rel} is the relative heading angle, $v^i_{x\_\text{rel}}$, $v^i_{y\_\text{rel}}$ are the relative velocity of the target ship in the Earth-fixed coordinate system. DCPA^i and TCPA^i represent the distance to the closest point of approach and the time to the closest point of approach. In this study, all ships involved in the simulation are assumed to be standard power-driven ships underway with equivalent navigational status. Therefore, the COLREGs-related decision-making mainly focuses on the encounter situations among power-driven ships, including head-on, crossing, and overtaking scenarios. The influence of special ship categories defined in COLREGs Rule 18, such as ships engaged in fishing, ships restricted in their ability to manoeuvre, and towing ships, is beyond the scope of this study.

3.4. Ship Decision Reward Function

The reward function is crucial for guiding deep-reinforcement-learning ships toward learning desired behaviours. In this study, a multi-objective composite reward function is formulated to balance goal-reaching performance, safety, and compliance with COLREGs.

The destination proximity reward R_{des} is introduced to encourage the ship to minimise its distance to the target position at each time step. Specifically, a higher reward is assigned as the ship moves closer to the destination over successive time steps. Let d_t denote the Euclidean distance between the own ship and the destination at time step t. R_{des} is defined as shown below:

$$R_{\text{des}} = k_{\text{p}} \cdot (d_t - d_{t-1}) \tag{13}$$

As shown in Equation (14), R_{head} represents the heading reward. ψ_{o} denotes the current heading angle of the own ship, while ψ_{g} is the heading angle required to point directly toward the destination. To prevent the ship from making unnecessary detours, a larger reward is granted as the difference between ψ_o and ψ_g decreases:

$$R_{\text{head}} = -k_h \cdot \frac{|\psi_o - \psi_g|}{180^\circ} \tag{14}$$

As shown in Equations (15)–(17), R_{cri} denotes the reward function for collision risk. The collision risk is characterised by DCPA and TCPA. A larger negative reward is imposed as the collision risk increases when the ship and the target ship approach the CPA:

$$\mu_d(\text{DCPA}) = \begin{cases} 1, & \text{DCPA} \le d_1 \\ \left(\frac{d_2 - \text{DCPA}}{d_2 - d_1}\right)^k, & d_1 < \text{DCPA} < d_2 \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

$$\mu_t(\text{TCPA}) = \begin{cases} 1, & \text{TCPA} \le t_1 \\ \left(\frac{t_2 - \text{TCPA}}{t_2 - t_1}\right)^m, & t_1 < \text{TCPA} < t_2 \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

$$R_{\text{cri}} = -k_{\text{cri}} \cdot [\alpha \cdot \mu_d + (1 - \alpha)\mu_t] \tag{17}$$

R_{rule} represents the reward associated with COLREGs-consistent behaviours among power-driven ships. In this study, COLREGs consistency among power-driven ships is treated as a soft constraint in the reward design. Therefore, in normal encounter situations, a small positive reward is assigned when the ship adopts a manoeuvre consistent with the corresponding COLREGs requirements. This encourages the learned policy to generate standard, predictable, and rule-compliant collision avoidance behaviours while maintaining reasonable progress toward the destination. Nevertheless, navigational safety is regarded as a hard constraint. Therefore, if the ship takes a dangerous action during navigation, such as manoeuvring toward the target ship, reducing the predicted passing distance, crossing ahead at insufficient clearance, or failing to take necessary avoidance action, a large negative reward is imposed. Similarly, behaviours that violate COLREGs and increase collision risk are strongly penalised. In this way, the reward design distinguishes between the soft guidance effect of COLREGs compliance and the hard safety requirement of collision avoidance. In the current framework, the basic responsibility assignment between give-way and stand-on vessels is considered according to the identified encounter situations. However, some detailed COLREGs requirements, such as the temporal requirement of early action under Rule 8 and different vessel-status hierarchies, are simplified into a general compliance guidance term. Accordingly, R_{rule} is defined as follows:

$$R_{\text{rule}} = \begin{cases} -1, & \text{if against rules} \\ 0.1, & \text{if complying with rules} \end{cases} \tag{18}$$

As shown in Equation (19), R_{speed} is the speed reward function, where V_{eco} denotes the economic speed and U represents the current resultant speed of the ship. To maintain the ship at the economic speed and avoid unnecessary acceleration or deceleration, a larger negative reward is imposed as the difference between V_{eco} and U increases:

$$R_{\text{speed}} = -k_s \cdot \frac{|V_{\text{eco}} - U|}{V_{\text{eco}}} \tag{19}$$

The goal reward R_{goal} is designed to encourage the ship to reach the destination, as shown in Equation (20). d_{og} represents the distance between the ship and the destination, while d_{goal} is the distance threshold used to determine whether the ship has arrived at the destination. If d_{og} is less than d_{goal}, the ship is considered to have reached the destination and is granted a positive reward r_{g}; otherwise, the reward is 0:

$$R_{\mathrm{goal}} = \begin{cases} r_{\mathrm{g}}, & d_{\mathrm{og}} \leq d_{\mathrm{goal}} \\ 0, & d_{\mathrm{og}} > d_{\mathrm{goal}} \end{cases} \tag{20}$$

The collision reward R_{coll} is designed to keep the ship away from target ships, as expressed in Equation (21). d_{ot} denotes the distance between the own ship and the target ship, while d_{c} is the distance threshold used to determine whether a collision has occurred. If d_{ot} is less than d_{c}, a negative reward k_{c} is imposed to penalise the collision event:

$$R_{\mathrm{coll}} = \begin{cases} k_{\mathrm{c}}, & d_{\mathrm{ot}} \leq d_{\mathrm{c}} \\ 0, & d_{\mathrm{ot}} > d_{\mathrm{c}} \end{cases} \tag{21}$$

Based on the above reward functions, the total reward R_i for each ship i at each time step can be expressed as follows:

$$R_i = R_{\mathrm{des}} + R_{\mathrm{head}} + R_{\mathrm{cri}} + R_{\mathrm{rule}} + R_{\mathrm{speed}} + R_{\mathrm{goal}} + R_{\mathrm{coll}} \tag{22}$$

The total system reward R^{tot} at each time step is obtained by summing the total rewards R_i for each ship i, as defined in Equation (23):

$$R^{\mathrm{tot}} = \sum_i R_i \tag{23}$$

3.5. Implementation of the Deep-Reinforcement-Learning Algorithm

To ensure the stability and convergence of the MATD3 algorithm in multi-ship encounter scenarios, the hyperparameters should be systematically tuned. The configuration is designed to balance the trade-off between exploration and exploitation, while effectively addressing the inherent non-stationarity of multi-ship environments. Key parameters, including the network architecture, learning rates of the Actor and Critic networks, batch size, and the capacity of the experience replay buffer, were carefully selected to enhance training efficiency and overall performance. The detailed hyperparameter settings adopted in the experiments are summarised in Table 2.

Table 2. Parameters in training process.

Parameter	Symbol	Value
Hidden layer units	H	128
Optimiser	-	Adam
Batch size	B	512
Replay buffer size	D	3×10^5
Discount factor	γ	0.96
Actor learning rate	lr_{a}	5×10^{-4}
Critic learning rate	lr_{c}	1×10^{-3}
Soft update rate	τ	0.001

The parameter values in the reward function are shown in Table 3.

The reward coefficients listed in Table 3 were determined through a manual tuning procedure based on the physical meaning of each reward component and preliminary training observations. The relative magnitudes of the coefficients were adjusted to establish an appropriate priority among competing objectives. Specifically, safety-related objectives, including collision avoidance and COLREGs compliance, were assigned higher priority, while speed maintenance and destination reaching were considered secondary objectives to ensure that the generated manoeuvres remain both safe and operationally reasonable. No automatic hyperparameter optimisation or grid search was applied in this study.

Table 3. Parameters in reward function.

Symbol	Value
k_p	8
k_h	6
k_{cri}	1.5
α	0.5215
d_1	0.5 n mile
d_2	1.5 n mile
t_1	360 s
t_2	1800 s
k	2
m	2
k_s	1
r_g	100
k_c	−150

The algorithmic process of MATD3 is presented in Algorithm 1, which comprises the following three stages.

Algorithm 1: MATD3 for Collaborative Collision Avoidance

Initialise actor network π_ϕ, twin critic networks Q_{θ_1}, Q_{θ_2} with random parameters ϕ, θ_1, θ_2.
Initialise target networks ϕ', θ'_1, θ'_2 with weights equal to main networks.
Initialise replay buffer D and exploration noise σ_{exp}.
for episode = 1 to M **do**
  Initialise environment and receive initial state S^1.
  for t = 1 to $T_{\max}$ **do**
    Select joint action with exploration noise: $A^t = \text{clip}\left(\pi_\phi(S^t) + \varepsilon_{\text{exp}}, [-1,1]\right)$, where $\varepsilon_{\text{exp}} \sim \mathcal{N}(0, \sigma_{\text{exp}})$.
    Execute A^t in simulation environment.
    Observe next state S^{t+1} and calculate total reward R^t.
    Store transition $\left(S^t, A^t, R^t, S^{t+1}, Done^t\right)$ in replay buffer D.
    if $\text{len}(D) > B$ **then**
      Sample a random minibatch of B transitions $\left(S^i, A^i, R^i, S^{i+1}, Done^i\right)$ from D.
      Sample noise: $\varepsilon \sim \text{clip}(\mathcal{N}(0, \widetilde{\sigma}), [-c, c])$.
      Compute smoothed target action: $\widetilde{A}^{i+1} \leftarrow \text{clip}\left(\pi_{\phi'}\left(S^{i+1}\right) + \varepsilon, [-1,1]\right)$.
      Compute target value: $y^i \leftarrow R^i + \gamma\left(1 - Done^i\right)\min_{j=1,2} Q_{\theta'_j}\left(S^{i+1}, \widetilde{A}^{i+1}\right)$
      Update critics θ_1, θ_2 by minimising loss: $L = \frac{1}{B}\sum_i\left(y^i - Q_{\theta_j}\left(S^i, A^i\right)\right)^2$
      if t mod $d_{\text{freq}} = 0$ **then**
        Update policy ϕ by maximising: $\nabla_\phi J \approx \frac{1}{B}\sum_i \nabla_a Q_{\theta_1}\left(S^i, a\right)\Big|_{a=\pi_\phi(S^i)} \nabla_\phi \pi_\phi\left(S^i\right)$
        Soft update targets: $\theta' \leftarrow \tau\theta + (1-\tau)\theta'$, $\phi' \leftarrow \tau\phi + (1-\tau)\phi'$.
      end if
    end if
    if $Done^t$ is True **break**
  end for
end for

Stage 1 (Steps 1–5): The Actor network π_ϕ and twin Critic networks Q_{θ_1} are randomly initialized to establish the policy and value estimation functions. Simultaneously, the target networks (ϕ', θ'_1, θ'_2) are initialised with weights identical to those of the main networks. An experience replay buffer D is created to store transition data, and exploration noise parameters σ_{exp} are set to facilitate action exploration. At the beginning of each episode, the simulation environment is reset, and the initial state S^1 is observed based on own ship motion and target ship information.

Stage 2 (Steps 6–10): Ships make decisions and interact with the maritime environment driven by MMG model. At each time step t, the ships select a joint action A^t based on the current policy π_ϕ (S^t) and overlay exploration noise $\varepsilon_{\text{exp}} \sim \mathcal{N}(0, \sigma_{\text{exp}})$ to prevent falling into local optima. After executing the action, the system transitions to the next state S^{t+1} and receives a reward R^t. The transition tuple $(S^t, A^t, R^t, S^{t+1}, Done^t)$ is subsequently stored in the replay buffer D to accumulate experience for subsequent learning.

Stage 3 (Steps 11–25): When the experience replay buffer has accumulated sufficient data, a mini-batch B of transition data is randomly sampled from D. To overcome the problem of value overestimation, the algorithm introduces a target-policy-smoothing mechanism and utilises the minimum value of the twin target critics to compute the target value y^{i}. Subsequently, the online Critic networks are updated by minimising the loss function. Following this, the online Actor network is updated using the policy gradient only when the delayed update condition t mod $d_{\text{freq}} = 0$ is satisfied. Finally, a soft update is performed on all target networks to ensure the stability of the training process.

4. Case Study

4.1. Two-Ship Encounter Scenario

4.1.1. Two-Ship Encounter Scenario Setting

To verify the feasibility and effectiveness of the model, a simulation environment is constructed based on the PyCharm 2024.2.2 platform. A series of encounter scenarios are formulated, starting with the validation of the collision avoidance capability in two-ship encounter situations. The initial ship parameters are shown in Table 4.

Table 4. The initial ships parameters of two-ship encounter scenario.

Scenario	No.	Position/n Mile	Destination/n Mile	Heading/°	Speed/kn
Head-on	S0 S1	(0, 0) (7.5, 7.5)	(7.5, 7.5) (0, 0)	045 225	12 12
Crossing	S0 S1	(0, 0) (7.5, 0)	(7.5, 7.5) (0, 7.5)	045 315	12 12
Overtaking	S0 S1	(0, 0) (4.5, 4.5)	(11, 11) (9.5, 9.5)	045 045	12 5

4.1.2. Simulation Results of Two-Ship Encounter Scenario

- **Head-on:** In this experiment, two ships are set to sail toward each other at an initial speed of 12 kn. At the initial stage, S0 is positioned at (0, 0) with a heading of 45°, while S1 is located at (7.5, 7.5) with a heading of 225°. Figures 2a, 3a, 4a and 5a show the encounter trajectories, the distance curves between the ships, the rudder angle curves, and the speed curves. At 720 s, both ships begin to take actions of turning starboard. The maximum rudder angles during the manoeuvre are 19.76° and 20.00° for S0 and S1, respectively, and their maximum turning amplitudes are 22.04° and 21.72°. The two ships safely pass and clear each other at 1579 s, reaching a minimum relative distance of 0.76 n mile. The minimum speed during the avoidance is 11.53 kn, and the entire collision avoidance process lasts for 859 s.

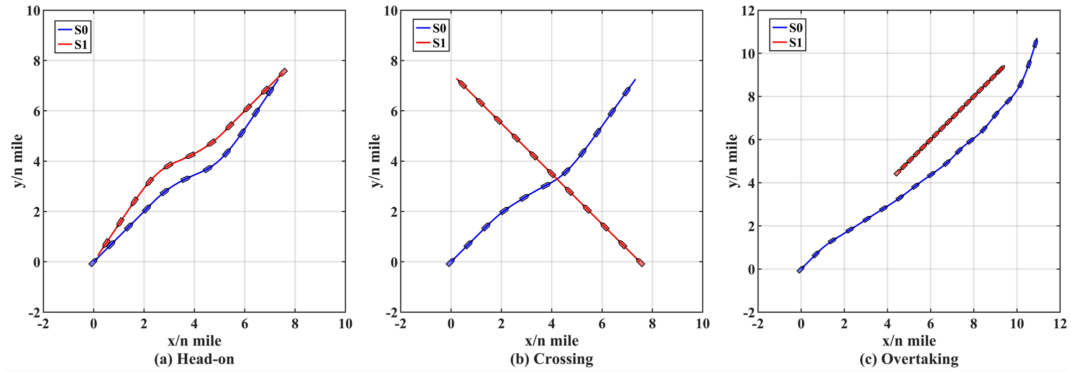

Figure 2. Encounter trajectories of each ship.

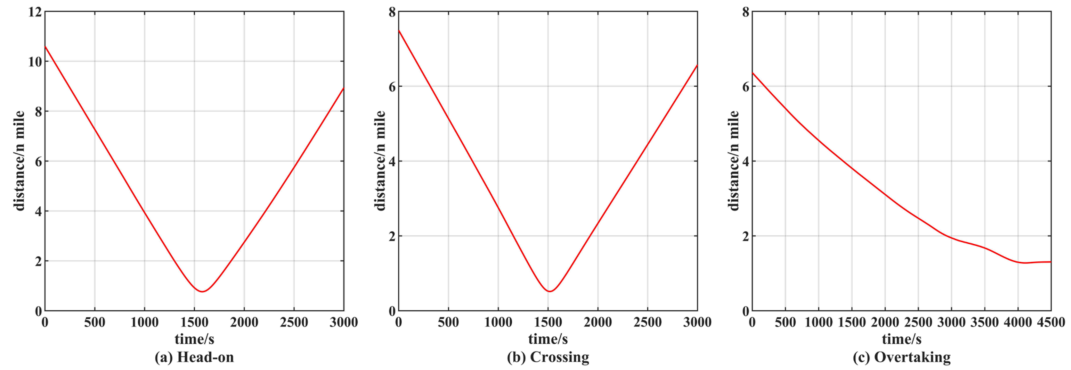

Figure 3. Distance curves between ships in different scenarios.

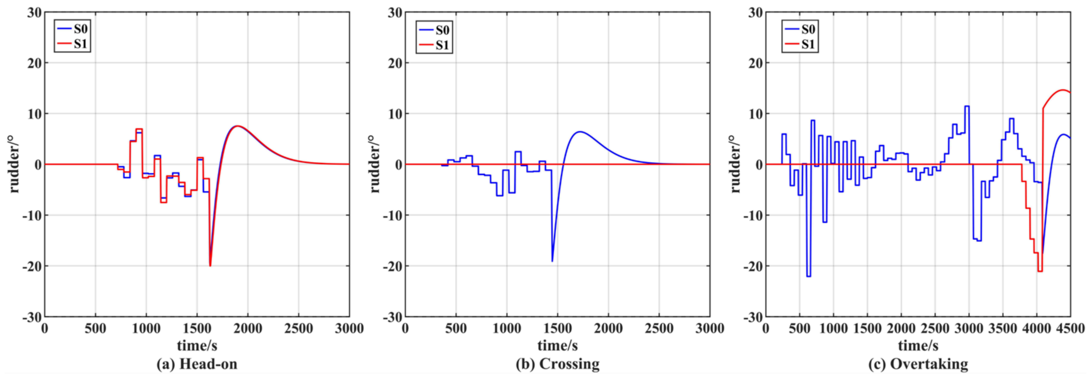

Figure 4. Rudder angle curves of each ship in different scenarios.

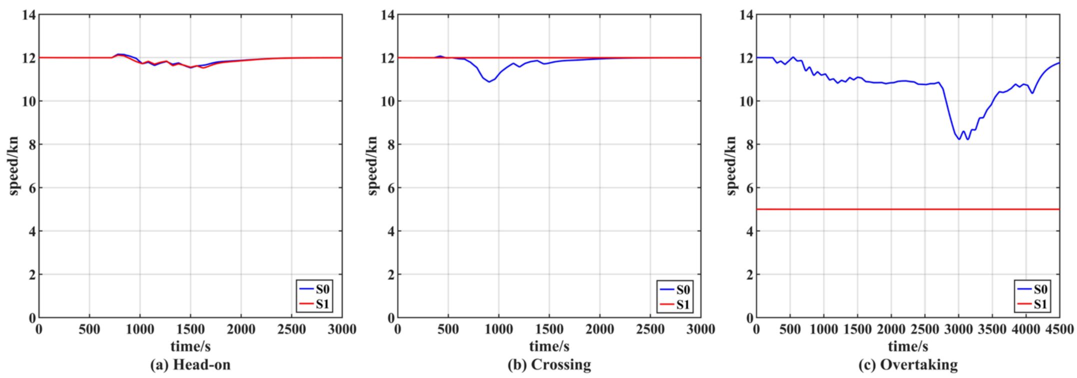

Figure 5. Speed curves of each ship in different scenarios.

- **Crossing:** In the crossing situation, S1 approaches from the starboard side of S0, with both ships maintaining an initial speed of 12 kn. The initial position of S0 is (0, 0) with a heading of 45°, while S1 starts at (7.5, 0) with a heading of 315°. Figures 2b, 3b, 4b and 5b show the encounter trajectories, the distance curves between the ships, the rudder angle curves, and the speed curves. At 420 s, the give-way ship S0 begins to turn starboard, while the stand-on ship S1 maintains its course and speed. The maximum rudder angle for S0 during the manoeuvre is 19.11°, and its maximum turning amplitude is 17.67°. The two ships safely pass and clear each other at 1517 s, reaching a minimum relative distance of 0.51 n mile. The minimum speed of S0 during the avoidance is 10.88 kn, and the entire collision avoidance process lasts for 1097 s.
- **Overtaking:** In the overtaking experiment, S0 is set to approach and overtake S1 from behind. The initial position of S0 is (0, 0) with a heading of 45° and a speed of 12 kn. S1 is initially located at (4.5, 4.5) with a heading of 45° and a speed of 5 kn. Figures 2c, 3c, 4c and 5c show the encounter trajectories, the distance curves between the ships, the rudder angle curves, and the speed curves. At 240 s, the give-way ship S0 immediately begins to take active evasion actions, while the stand-on ship S1 maintains its course and speed. The maximum absolute rudder angle for S0 during the manoeuvre is 22.07°, and its maximum turning amplitude is 29.24°. The two ships completely pass and clear each other at 4106 s, reaching a minimum distance of 1.28 n mile. The minimum speed of S0 during the avoidance is 8.22 kn, and the entire overtaking and avoidance process lasts for 5172 s.

4.2. Multi-Ship Encounter Scenario

4.2.1. Multi-Ship Encounter Scenario Setting

To further verify the practicality and reliability of the proposed algorithm, two sets of four-ship encounter scenarios are designed under different operating conditions. The initial ship parameters are summarised in Table 5. In the four-ship encounter scenario, the initial states of the ships are constructed to form a highly symmetrical and complex encounter situation. Specifically, S0 and S1 constitute a head-on situation in the longitudinal direction, while S2 and S3 also exhibit a head-on status in the lateral direction. As the planned routes of both pairs of ships intersect at the central region, a typical and complex multi-ship crossing encounter situation is formed.

Table 5. The initial ships parameters of four-ship encounter scenario.

No.	Position/n Mile	Destination/n Mile	Heading/°	Speed/kn
S0	(0, −7.5)	(0, 67.5)	000	15
S1	(0, 7.5)	(0, −67.5)	180	15
S2	(6, 0)	(−54, 0)	270	12
S3	(−6, 0)	(54, 0)	090	12

4.2.2. Simulation Results of Multi-Ship Encounter Scenario Under Normal Conditions

Figures 6–9 show the encounter trajectories, the distance curves between the ships, the heading curves, the rudder angle curves, and the speed curves. At 720 s, all ships begin to take actions of turning starboard, with turning amplitudes of S0/S1 at 30.49° and 34.75°, the heading stabilising at 358.69° and 178.46° at 2691 s and 2727 s, respectively, with the minimum speed 13.87 kn, and the whole process is 1028 s. The amplitudes of S2/S3 are 22.49° and 24.59°, starting to return to the destination direction at 1703 s and 1642 s, and finally heading to 269.19° and 89.10°. The entire process is 1057 s with the minimum speeds of 11.39 kn and 11.71 kn.

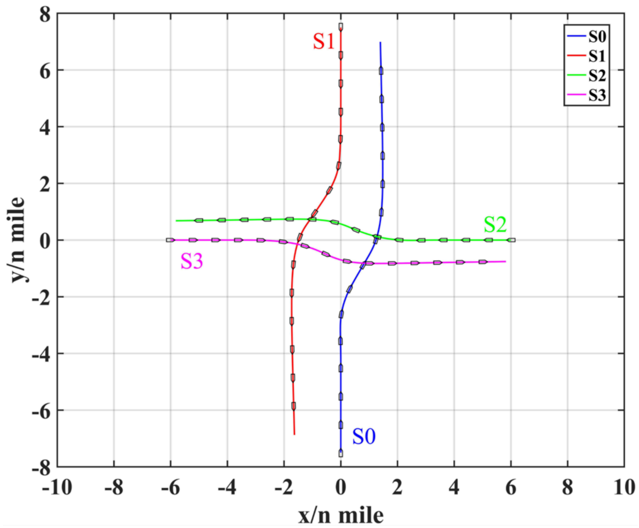

Figure 6. Encounter trajectories of each ship.

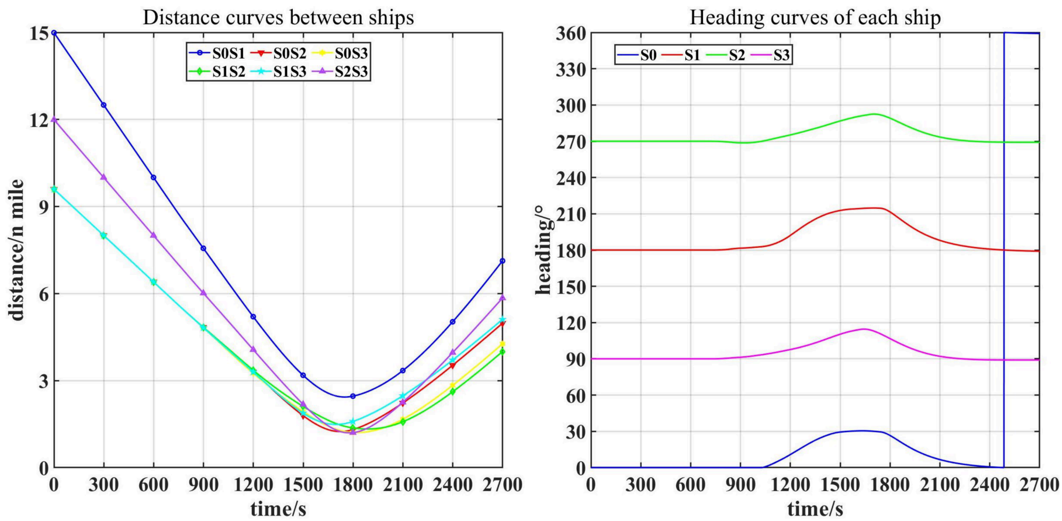

Figure 7. Curves of inter-ship distances and heading variations.

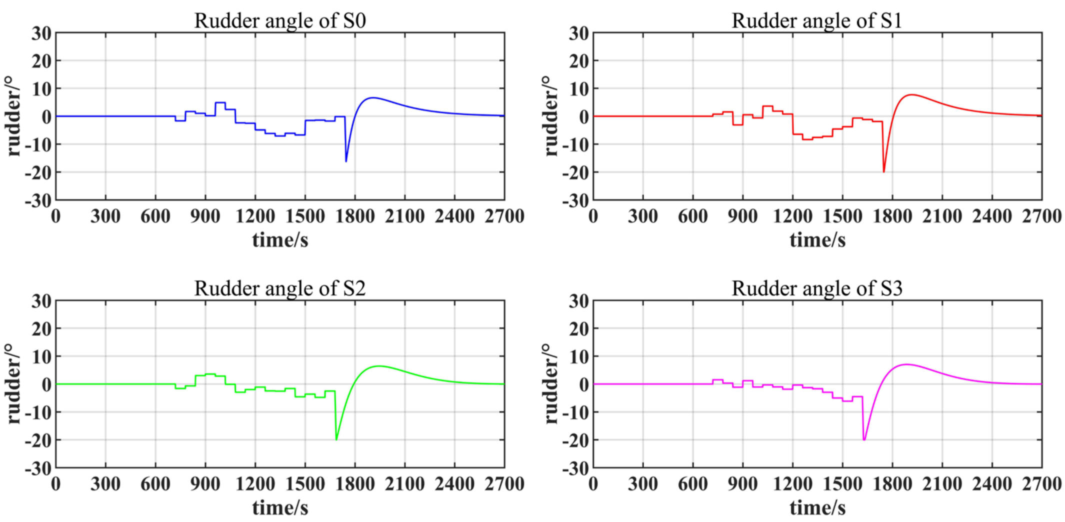

Figure 8. Rudder angle curves of each ship.

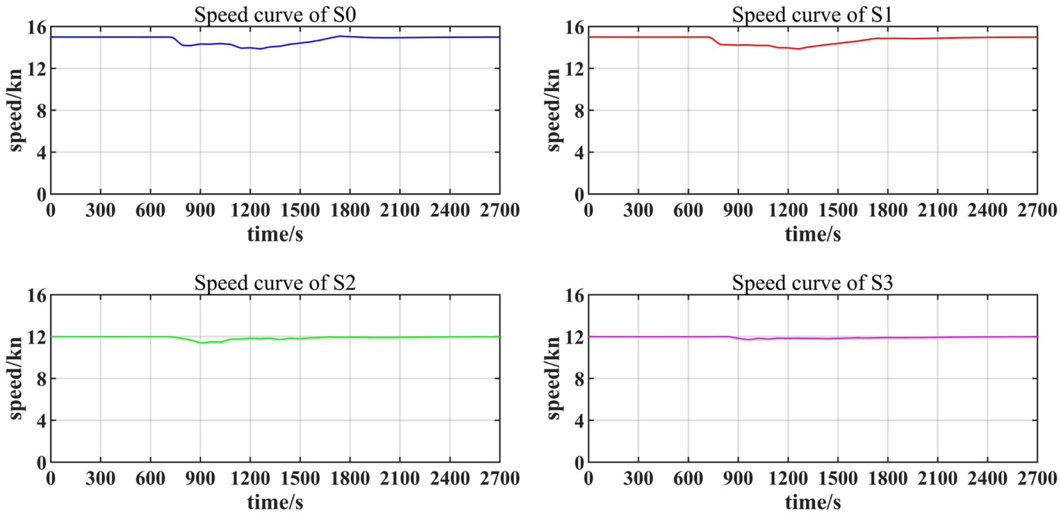

Figure 9. Speed curves of each ship.

4.2.3. Simulation Results of Multi-Ship Encounter Scenario Under Abnormal Conditions

In this simulation, S0 is set to not take actions according to game intentions. The simulation results are shown in Figures 10–13. Since S0 does not participate in the game decision, the situation is no longer symmetric. Figure 12 shows the distance curve. During the whole encounter process, the minimum distance is between S0 and S3 with a 0.74 n mile. Due to S0 keeping the heading and speed without taking action, S1 and S3 are required to take actions by turning starboard, but S2 is required to reduce the speed. S1 turns starboard at 46.86° and finally stabilises at 178.48° towards the goal point, with the process duration 2007 s and the minimum speed 9.78 kn. S2 primarily decelerates to avoid collision with the non-cooperative S0, reducing its speed from 12 kn to 3.66 kn, with the whole avoidance process lasting 1620 s and maintaining a stabilised heading of 269.93°. The turning amplitude of S3 is 24.98° towards starboard, starting its manoeuvre at 780 s and finally heading to 89.23° with a minimum speed of 8.59 kn during the entire process. These results demonstrate that the proposed method can effectively adapt to non-cooperative scenarios by redistributing collision avoidance responsibilities among cooperative ships; thereby, the whole encounter situations are maintained safety.

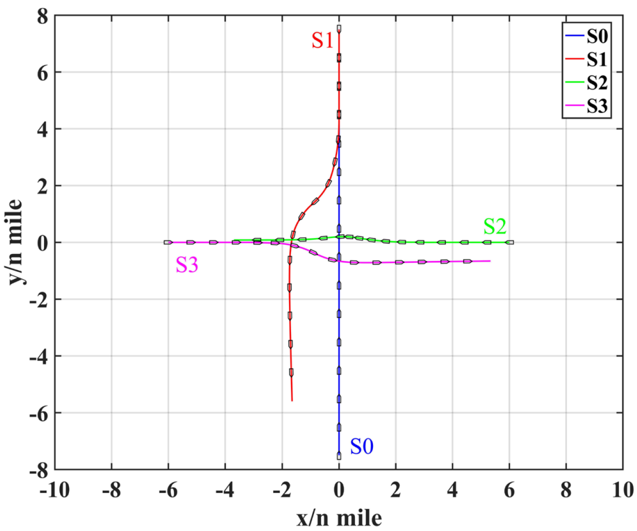

Figure 10. Encounter trajectories of each ship.

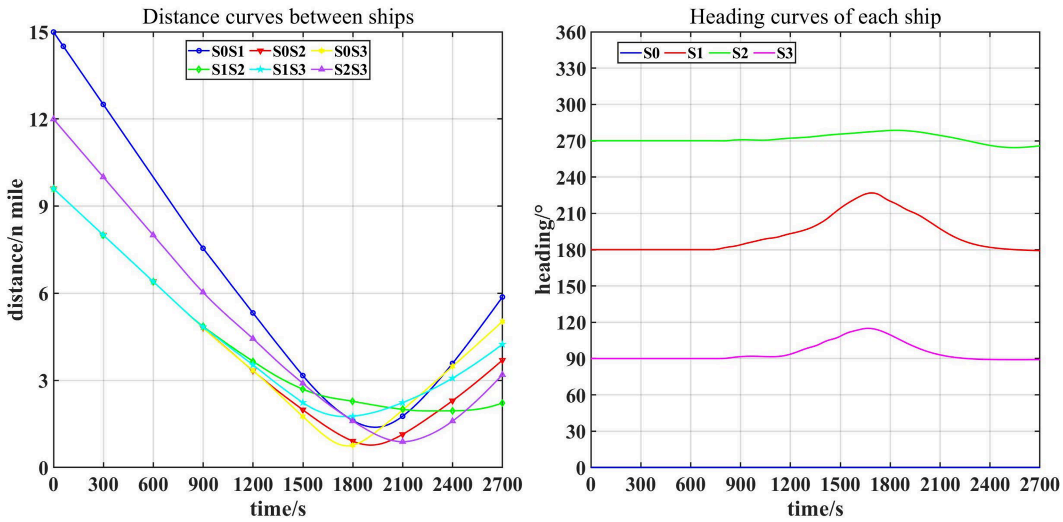

Figure 11. Curves of inter-ship distances and heading variations.

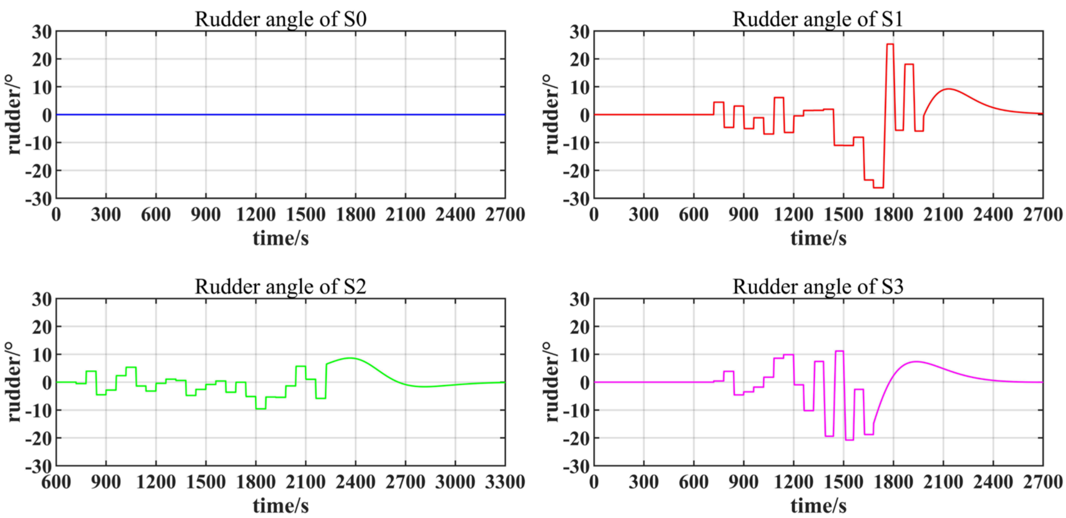

Figure 12. Rudder angle curves of each ship.

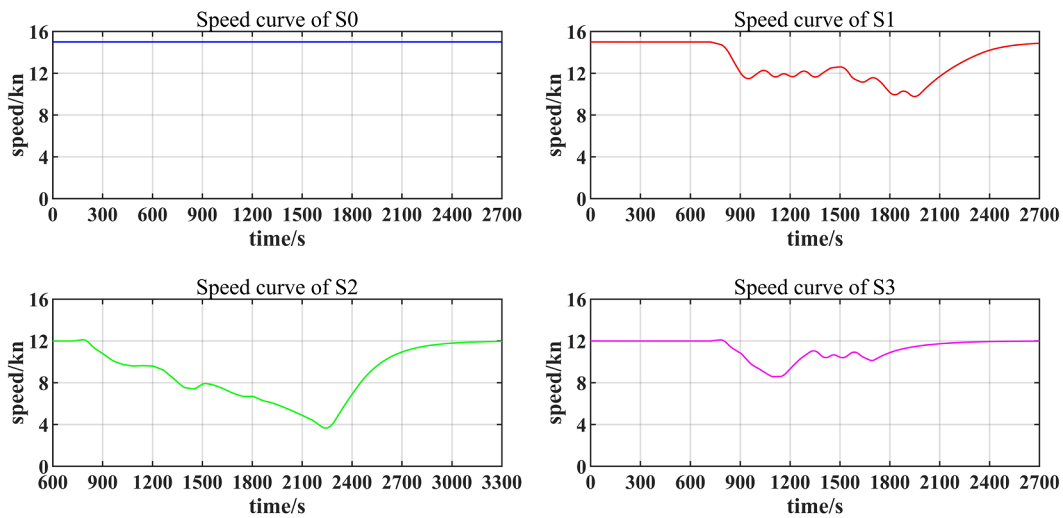

Figure 13. Speed curves of each ship.

Although the proposed framework achieves collision avoidance in the investigated scenarios, the interpretation of the obtained CPA values should consider the adopted safety thresholds. The CPA thresholds adopted in this study are primarily designed for evaluating the collision avoidance decision-making performance in simulation scenarios and do not represent universal safety criteria for practical VLCC operations. For real open-water navigation, larger CPA thresholds should be considered according to the vessel size, manoeuvring characteristics, and operational requirements.

4.3. Discussion

To validate the applicability and effectiveness of the proposed MATD3 model, a comparative study is conducted based on the experimental scenarios in Liu et al. [48]. Tables 6 and 7 summarise the performance comparison between the two methods in symmetric and asymmetric encounter situations, respectively. In the symmetric scenario, the MATD3 algorithm demonstrated superior kinetic energy maintenance compared to the algorithm proposed by Liu et al. [48]. Specifically, the minimum speed for S0/S1 is improved from 12.48 kn to 13.87 kn, while S2/S3 saw an optimisation from 10.02 knots to 11.39 knots. Furthermore, the MATD3 algorithm successfully reduces the maximum turning amplitude from 28.20° to 22.49°, directly substantiating its unique advantages in minimising the trajectory redundancy and enhancing the navigational economy. Regarding the asymmetric encounter scenario, where S0 serves as a non-cooperative moving obstacle, the MATD3 algorithm exhibits promising adaptability and differentiated decision-making behaviour. For S1, the algorithm executes a more decisive heading change with a maximum amplitude of 46.86°, compared to 28.20° in the study by Liu et al. [48]. Conversely, ship S2 prioritised safety through an assertive speed reduction to 3.66 kn, showing a more robust braking response than the 7.24 kn reported in the reference research. Most notably, S3 achieves a higher manoeuvring efficiency with a turning amplitude of only 24.98°, representing a 33.4% reduction from 37.52° observed in the PCA-Game algorithm. Overall, the comparative results indicate that the proposed MATD3 method can maintain a safe collision avoidance performance while achieving favourable speed preservation and moderate manoeuvring behaviour in the tested representative scenarios. However, it should be noted that the current comparison mainly considers the minimum speed and turning amplitude. Since the travel distance, path extension ratio, and time cost are not yet quantitatively evaluated, the present results should be interpreted as preliminary evidence of an improved manoeuvring performance rather than a comprehensive assessment of the navigational efficiency. More complete efficiency metrics will be included in future work to further evaluate the route economy and time efficiency of the proposed method.

Table 6. Performance comparison in symmetric four-ship scenario.

	Evaluation Metric	**PCA-Game Algorithm**	**MATD3 Algorithm**
S0/S1	Min Speed	12.48 kn	13.87 kn
S2/S3	Min Speed	10.02 kn	11.39 kn
	Amplitude	28.20°	22.49°

Table 7. Performance comparison in non-cooperative scenario.

	Evaluation Metric	**PCA-Game Algorithm**	**MATD3 Algorithm**
S1	Amplitude	28.20°	46.86°
S2	Min Speed	7.24 kn	3.66 kn
S3	Amplitude	37.52°	24.98°

Similar to other deep-reinforcement-learning approaches, the performance of the proposed method may be influenced by stochastic factors, such as random initialisation, exploration noise, and experience replay sampling. The current evaluation focuses on representative training results under the designed encounter scenarios. Furthermore, the performance improvement demonstrated in this study should be interpreted as the combined contribution of the MATD3-based decision-making module, high-fidelity MMG dynamics modelling, and reward design. Since the primary objective of this work is to develop an integrated collision avoidance framework rather than to investigate the superiority of a specific reinforcement learning algorithm, PCA-Game was selected as the primary baseline due to its comparable multi-ship encounter setting. Nevertheless, direct comparisons with other multi-agent reinforcement-learning algorithms, such as MADDPG, would provide further insight into the contribution of different policy optimisation mechanisms and will be explored in future studies.

5. Conclusions

This study addresses some key challenges in multi-ship collision avoidance, including oversimplified dynamic models, unstable multi-ship interaction strategies, and the presence of non-compliant ships, by developing a decision-making framework based on the multi-agent deep-reinforcement-learning (MADRL) algorithm. The main findings and conclusions are summarised as follows:

- Unlike conventional approaches that rely on constant-speed assumptions and simplified dynamics, this study incorporates a 3-DOF MMG dynamic model. By capturing the physical coupling between the ship speed and heading variation, the proposed method ensures the practical executability of collision avoidance actions and enhances the performance in the investigated multi-ship encounter scenarios.
- By jointly optimising the speed and heading variations alongside their physical coupling, the algorithm successfully generates well-regulated trajectories while effectively maintaining the ship's kinetic energy and maintaining reasonable speed profiles while allowing necessary speed adjustments during multi-ship encounter scenarios.
- The multi-agent decision-making algorithm demonstrates an effective collision avoidance capability in the investigated scenarios involving non-compliant or rule-violating target ships. In asymmetric non-cooperative scenarios, the algorithm successfully achieves safe detours and secure collision avoidance coordination.
- For future work, addressing the following issues will broaden the application scope of the algorithm and further enhance its practicality and generalisation capabilities.
- Future research will extend the current MMG model by incorporating environmental disturbances, such as wind, waves, and currents. These effects can be introduced as additional external force and moment terms in the dynamic equations, i.e., by augmenting the existing force vector with environmental components representing wind loads, wave-induced forces, and current effects.
- The current simulation assumes that the ship dynamics characteristics of all ships are uniform. Given that different ships exhibit distinct inertia and nonlinear hydrodynamic characteristics, future research will incorporate various open-source benchmark MMG ship models to systematically evaluate and enhance the generalizability and scalability of the proposed algorithm across heterogeneous fleets.
- This study primarily validates encounter scenarios in open waters. Future research should further include a larger number of diverse and randomly generated encounter scenarios with different initial positions, headings, speeds, traffic densities, and cooperative/non-cooperative ship behaviours to more comprehensively evaluate the robustness and generalisation capability of the proposed framework. In addition, restricted waters, such as inland waterways and congested port areas, should be considered, where static obstacles and bathymetric constraints impose stricter operational limitations.

- The current framework assumes that all vessels involved in the simulation are standard power-driven vessels underway with equivalent navigational status. Future research will incorporate additional semantic information, such as AIS-based vessel status, vessel type, and navigation lights/day shapes, into the state representation to enable more comprehensive COLREGs-aware decision-making in diverse maritime environments.
- Although simulation results demonstrate the framework's effectiveness, transitioning from numerical models to practical deployment remains a critical step. Future work could involve conducting hardware-in-the-loop testing and real-world sea trials on physical unmanned surface vehicles. In this context, future research will explicitly model the impact of sensor noise and perception uncertainty within the reinforcement learning loop. This will verify the algorithm's robustness against unpredictable environmental disturbances and sensor noise, ensuring the safety and reliability of autonomous decision-making in actual maritime operations.

Author Contributions: Conceptualisation, J.Z. (Junheng Zhao) and J.L.; methodology, J.Z. (Junheng Zhao); software, J.Z. (Junheng Zhao); validation, J.Z. (Junheng Zhao) and Z.H.; formal analysis, J.Z. (Junheng Zhao); investigation, J.Z. (Junheng Zhao); resources, J.L., J.Z. (Jinfen Zhang), and W.T.; data curation, J.Z. (Junheng Zhao); writing—original draft preparation, J.Z. (Junheng Zhao); writing—review and editing, J.L. and J.Z. (Jinfen Zhang); visualisation, J.Z. (Junheng Zhao); supervision, J.L. and J.Z. (Jinfen Zhang); project administration, J.L.; funding acquisition, J.L., J.Z. (Jinfen Zhang), and W.T. All authors have read and agreed to the published version of the manuscript.

Funding: The article is financially supported by the National Science Foundation of China (52372320), China Postdoctoral Science Foundation (GZC20251130; 2025M781572), Guangxi Science and Technology Program (AB23026132; GuikeLT2504240033), Key Research and Development Program of Guangxi, Zhuang Autonomous Region (Guike2023AA14003), Guangxi Natural Science Foundation (2025GXNSFHA069259; 2026GXNSFHA00640291), Guangxi Maritime Economy Talent Development Support Special Program (2025XHRC06), and Innovation Fund from State Key Laboratory of Maritime Technology and Safety under Grant (SKL202403).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. United Nations. *Conference on Trade and Development Review of Maritime Transport 2023*; United Nations: Geneva, Switzerland, 2023.
2. Okumus, D.; Gunbeyaz, S.A.; Kurt, R.E.; Turan, O. An Approach to Advance Circular Practices in the Maritime Industry through a Database as a Bridging Solution. *Sustainability* **2024**, *16*, 453. [CrossRef]
3. Allianz Commercial. *Safety and Shipping Review 2025*; Allianz Commercial: Munich, Germany, 2025.
4. Xu, H.; Guedes Soares, C. Challenges for the Development of Maritime Autonomous Surface Ships. *Auton. Transp. Res.* **2026**, *2*, 1–24. [CrossRef]
5. Feng, Y.; Wang, X.; Chen, Q.; Yang, Z.; Wang, J.; Li, H.; Xia, G.; Liu, Z. Prediction of the Severity of Marine Accidents Using Improved Machine Learning. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *188*, 103647. [CrossRef]
6. European Maritime Safety Agency. *Annual Overview of Marine Casualties and Incidents 2025*; European Maritime Safety Agency: Lisbon, Portugal, 2025.
7. Fan, S.; Yang, Z. Accident Data-Driven Human Fatigue Analysis in Maritime Transport Using Machine Learning. *Reliab. Eng. Syst. Saf.* **2024**, *241*, 109675. [CrossRef]
8. Feng, Y.; Wang, H.; Xia, G.; Cao, W.; Li, T.; Wang, X.; Liu, Z. A Machine Learning-Based Data-Driven Method for Risk Analysis of Marine Accidents. *J. Mar. Eng. Technol.* **2025**, *24*, 147–158. [CrossRef]
9. Vagale, A.; Oucheikh, R.; Bye, R.T.; Osen, O.L.; Fossen, T.I. Path Planning and Collision Avoidance for Autonomous Surface Vehicles I: A Review. *J. Mar. Sci. Technol.* **2021**, *26*, 1292–1306. [CrossRef]

10. Xiao, Z.; Hou, B.; Ning, J.; Lin, B.; Liu, Z. Collision Avoidance for Unmanned Surface Vehicles in Multi-Ship Encounters Based on Analytic Hierarchy Process–Adaptive Differential Evolution Algorithm. *J. Mar. Sci. Eng.* **2024**, *12*, 2123. [CrossRef]
11. Wang, S.; Zhang, Y.; Song, F.; Mao, W. A Collaborative Collision Avoidance Strategy for Autonomous Ships under Mixed Scenarios. *J. Navig.* **2023**, *76*, 200–224. [CrossRef]
12. Liu, J.; Yu, H.; Huang, A.; Ma, X.; Wu, B.; Sun, J.; Jia, L.; Chen, Y.; Wang, Y.; Wang, J.; et al. Concepts, Key Technologies, Applications and Development Trends in Autonomous Transportation Systems. *Auton. Transp. Res.* **2025**, *1*, 1–23. [CrossRef]
13. Liu, K.; Wu, X.; Zhou, Y.; Yuan, Z.; Yang, X.; Xin, X.; Zhuang, S. A Conflict Cluster-Based Method for Collision Avoidance Decision-Making in Multi-Ship Encounter Situations. *Ocean Eng.* **2023**, *288*, 116038. [CrossRef]
14. Pan, R.; Zhang, W.; Wang, S.; Kang, S. Deep Reinforcement Learning Model for Multi-Ship Collision Avoidance Decision Making Design Implementation and Performance Analysis. *Sci. Rep.* **2025**, *15*, 21250. [CrossRef] [PubMed]
15. Wang, Z.; Chen, P.; Chen, L.; Mou, J. Collaborative Collision Avoidance Approach for USVs Based on Multi-Agent Deep Reinforcement Learning. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 4780–4794. [CrossRef]
16. Lyu, H.; Yin, Y. COLREGS-Constrained Real-Time Path Planning for Autonomous Ships Using Modified Artificial Potential Fields. *J. Navig.* **2019**, *72*, 588–608. [CrossRef]
17. Langxiong, G.; Li, X.; Yan, T.; Song, L.; Xiao, J.; Shu, Y. Intelligent Ship Path Planning Based on Improved Artificial Potential Field in Narrow Inland Waterways. *Ocean Eng.* **2025**, *317*, 119928. [CrossRef]
18. Sang, H.; You, Y.; Sun, X.; Zhou, Y.; Liu, F. The Hybrid Path Planning Algorithm Based on Improved A* and Artificial Potential Field for Unmanned Surface Vehicle Formations. *Ocean Eng.* **2021**, *223*, 108709. [CrossRef]
19. Kuwata, Y.; Wolf, M.T.; Zarzhitsky, D.; Huntsberger, T.L. Safe Maritime Autonomous Navigation With COLREGS, Using Velocity Obstacles. *IEEE J. Ocean. Eng.* **2014**, *39*, 110–119. [CrossRef]
20. He, Y.; Zou, L.; Wu, Z.-X.; Liu, S.-Y.; Chen, W.-M.; Zou, Z.-J.; Celik, C. Integrated Path Following and Collision Avoidance Control for an Underactuated Ship Based on MFAPC. *Ocean Eng.* **2025**, *324*, 120706. [CrossRef]
21. Huang, Y.; Chen, L.; Van Gelder, P.H.A.J.M. Generalized Velocity Obstacle Algorithm for Preventing Ship Collisions at Sea. *Ocean Eng.* **2019**, *173*, 142–156. [CrossRef]
22. Wu, B.; Guo, H.; Liu, Z.; Ma, W. Optimization Method for Collision Avoidance Paths of Unmanned Vessels under Ship Position Prediction Uncertainty. *Ocean Eng.* **2026**, *346*, 123765. [CrossRef]
23. Johansen, T.A.; Perez, T.; Cristofaro, A. Ship Collision Avoidance and COLREGS Compliance Using Simulation-Based Control Behavior Selection With Predictive Hazard Assessment. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 3407–3422. [CrossRef]
24. Han, S.; Sun, J.; Ding, S.; Zhou, L. A Potential Field-Based Model Predictive Target Following Controller for Underactuated Unmanned Surface Vehicles. *IEEE Trans. Veh. Technol.* **2024**, *73*, 14510–14524. [CrossRef]
25. Tang, Y.; Chen, L.; Mou, J.; Chen, P.; Huang, Y.; Zhou, Y. Robust Model Predictive Control for Ship Collision Avoidance Under Multiple Uncertainties. *IEEE Trans. Transp. Electrif.* **2024**, *10*, 10374–10387. [CrossRef]
26. Huang, Z.; Zhang, J.; Xu, W.; Liu, Z. Anti-Pitching of High-Speed Multihull Ship Based on Fast Predictive Control and Iterative Learning Control. *Ocean Eng.* **2025**, *337*, 121927. [CrossRef]
27. Zhang, Z.; Liu, M.; Zhou, J.; Zheng, R.; Wei, P. An Improved Path Following Scheme for Low-Cost Underactuated USV with Cascaded Observers and Event-Triggered Mechanism. *Ocean Eng.* **2025**, *324*, 120703. [CrossRef]
28. Wu, Y.; Liu, J.; Li, X.; Xiao, J.; Zhang, T.; Xu, H.; Zhang, L. Towards LLM Enhanced Decision: A Survey on Reinforcement Learning Based Ship Collision Avoidance. *J. Mar. Sci. Eng.* **2025**, *13*, 2275. [CrossRef]
29. Li, L.; Ma, Y.; Wu, D. Underactuated MSV Path Following Control via Stable Adversarial Inverse Reinforcement Learning. *Ocean Eng.* **2024**, *299*, 117368. [CrossRef]
30. Niu, Y.; Zhu, F.; Wei, M.; Du, Y.; Zhai, P. A Multi-Ship Collision Avoidance Algorithm Using Data-Driven Multi-Agent Deep Reinforcement Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 2101. [CrossRef]
31. Lee, H.-T.; Kim, M.-K. Optimal Path Planning for a Ship in Coastal Waters with Deep Q Network. *Ocean Eng.* **2024**, *307*, 118193. [CrossRef]
32. Zhu, F.; Niu, Y.; Wei, M.; Du, Y.; Zhai, P. A High-Risk Test Scenario Adaptive Generation Algorithm for Ship Autonomous Collision Avoidance Decision-Making Based on Reinforcement Learning. *Ocean Eng.* **2025**, *320*, 120344. [CrossRef]
33. Xu, W.; Zhu, X.; Gao, X.; Li, X.; Cao, J.; Ren, X.; Shao, C. Manipulation-Compliant Artificial Potential Field and Deep Q-Network: Large Ships Path Planning Based on Deep Reinforcement Learning and Artificial Potential Field. *J. Mar. Sci. Eng.* **2024**, *12*, 1334. [CrossRef]
34. Qu, S.; Guan, W.; Hu, T.; Cui, Z. The Collaborative Navigation Decision-Making Method of USV by UAV Based on Improved PPO Algorithm. *Ocean Eng.* **2025**, *341*, 122381. [CrossRef]
35. Wang, J.; Lu, Z.; Hong, X.; Wu, Z.; Li, W. Navigation and Obstacle Avoidance for USV in Autonomous Buoy Inspection: A Deep Reinforcement Learning Approach. *J. Mar. Sci. Eng.* **2025**, *13*, 843. [CrossRef]
36. Xue, D.; Wu, D.; Yamashita, A.S.; Li, Z. Proximal Policy Optimization with Reciprocal Velocity Obstacle Based Collision Avoidance Path Planning for Multi-Unmanned Surface Vehicles. *Ocean Eng.* **2023**, *273*, 114005. [CrossRef]

37. Lou, M.; Yang, X.; Hu, J.; Zhu, Z.; Shen, H.; Xiang, Z.; Zhang, B. A Balanced Collision Avoidance Algorithm for USVs in Complex Environment: A Deep Reinforcement Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 21404–21415. [CrossRef]
38. Wei, Z.; Wang, Q. Autonomous Navigation and Collision Avoidance for Unmanned Surface Vehicle Based on TD3-PD Algorithm with CNN-GRU Network. *Ocean Eng.* **2025**, *341*, 122633. [CrossRef]
39. Cui, Z.; Guan, W.; Hao, S.; Qu, S.; Zhang, X. Adaptive Primal-Dual Proximal Policy Optimization Lagrange for Multiple USVs Autonomous Collision Avoidance Decision-Making Strategy. *IEEE Trans. Emerg. Top. Comput. Intell.* **2026**, *10*, 2952–2967. [CrossRef]
40. Cui, Z.; Guan, W.; Zhang, X.; Han, H. Autonomous Collision Avoidance Decision-Making Method for Multiple Marine Surface Ships Based on Deep Reinforcement Learning. *IEEE J. Ocean. Eng.* **2026**, *51*, 1198–1216. [CrossRef]
41. Lowe, R.; Wu, Y.; Tamar, A.; Harb, J.; Abbeel, P.; Mordatch, I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Nice, France, 2017; Volume 30.
42. Wang, Y.; Zhao, Y. Multiple Ships Cooperative Navigation and Collision Avoidance Using Multi-Agent Reinforcement Learning with Communication. *Ocean Eng.* **2025**, *320*, 120244. [CrossRef]
43. Wei, G.; Kuo, W. COLREGs-Compliant Multi-Ship Collision Avoidance Based on Multi-Agent Reinforcement Learning Technique. *J. Mar. Sci. Eng.* **2022**, *10*, 1431. [CrossRef]
44. Chen, C.; Ma, F.; Xu, X.; Chen, Y.; Wang, J. A Novel Ship Collision Avoidance Awareness Approach for Cooperating Ships Using Multi-Agent Deep Reinforcement Learning. *J. Mar. Sci. Eng.* **2021**, *9*, 1056. [CrossRef]
45. Fujimoto, S.; Hoof, H.; Meger, D. Addressing Function Approximation Error in Actor-Critic Methods. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1587–1596.
46. Ackermann, J.; Gabler, V.; Osa, T.; Sugiyama, M. Reducing Overestimation Bias in Multi-Agent Domains Using Double Centralized Critics. *arXiv* **2019**, arXiv:1910.01465.
47. Yasukawa, H.; Yoshimura, Y. Introduction of MMG Standard Method for Ship Maneuvering Predictions. *J. Mar. Sci. Technol.* **2015**, *20*, 37–52. [CrossRef]
48. Liu, J.; Zhang, J.; Zhang, M.; Xin, X.; Yang, Z. A Novel Collaborative Collision Avoidance Decision-Making Methodology Based on Potential Collision Areas for Intelligent Navigation. *Ocean Eng.* **2025**, *318*, 120126. [CrossRef]