Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P7222-cd

Towards robust deep reinforcement learning agent for the path following of autonomous ships amid perception sensor noise

Paul Lee

Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Norway.
E-mail: paul.lee@ntnu.no

Ekaterina Kim

Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Norway.
E-mail: ekaterina.kim@ntnu.no

In the era of maritime autonomous surface ships (MASSs), intelligent agents are projected to make safety-critical decisions without human intervention. Considering the various disturbances associated with the maritime environment, enhancing their robustness during safety-critical operations is pivotal, including those related to path following. The aim of this study is to propose a methodology that enhances the robustness of path following for a MASS amid perception sensor noise by controlling the state space parameter of a deep reinforcement learning agent. The agent is trained to follow a predefined path at various noise levels between a minimum and maximum value, and a robustness metric based on the cross-track error is defined. The case study considers a container vessel that uses light detection and ranging for the situation awareness of its surrounding environment. Simulation results suggest that when the state space parameter related to the value of the noise level is controlled, the robustness is enhanced up to 5,668% from its maximum trained value by not violating the cross-track error threshold. When the state space parameter is not controlled, an enhancement of up to 112% is noted, highlighting the effectiveness of the proposed methodology. This study contributes towards the development of agents capable of making robust decisions during safety-critical operations under uncertainty.

Keywords: Maritime autonomous surface ship, deep reinforcement learning, deep deterministic policy gradient, path following, robustness, perception senor noise, LIDAR, state space manipulation.

1. Introduction

In the new Shipping 4.0 era, the maritime industry is experiencing a paradigm shift, where conventional ships are transitioning to autonomous ships, known as maritime autonomous surface ships (MASSs), by adopting higher degrees of autonomy (Kavallieratos et al., 2020). However, autonomy comes with unique safety challenges, as safety-critical decisions are expected to be made by intelligent agents without human intervention (BahooToroody et al., 2022).

An important step towards realizing full-scale autonomy is the robustness of the agents used during safety-critical operations (Staessens et al., 2022). Specifically, the term 'robustness' refers to the degree of tolerability against disturbances without violating a set requirement (Hamon et al., 2020). Hence, robustness against the various disturbances attributed to the complex and dynamic maritime environment need to be investigated, including for path following, which is a fundamental safety-critical operation related to the safe navigation of MASSs.

2. Relevant Works

Various studies have been conducted to enhance the robustness of path following for MASSs. For instance, Fan et al. (2019) employed line-of-sight (LOS) and accommodated input saturation, model parameters uncertainties, and unknown time-varying external disturbances using radial basis function neural networks (RBFNN)-based finite-time observer. Huang and Fan (2019) employed integral LOS and compensated for the currents and model uncertainties using reduced-order linear extended state observer and RBFNN. Wen et al. (2020) employed vector field and estimated the model parameters using forgetting

factor recursive least square. Mu et al. (2020) employed fuzzy-based integral LOS and counteracted the effects of unknown dynamic and external disturbances using neural network minimum learning parameter (MLP).

Esfahani and Szlapczynski (2021) employed approximate dynamic programming based on actor-critic and time-delay control under stochastic disturbances induced by winds, waves, and currents. Mu et al. (2022) employed adaptive LOS and considered modeling error and input saturation under time-varying external disturbances using neural network MLP. Liu et al. (2022) employed model predictive control (MPC), adaptive LOS, and event-triggered mechanism (ETM) subject to external disturbances and nonlinear terms using linear extended state observer. Ren et al. (2023) employed proportional derivative-based sigmoid fuzzy function and accommodated unknown hydrodynamic coefficients and external disturbances using RBFNN. Li et al. (2024) employed model predictive static programming and ETM considering input disturbances. Song et al. (2024) employed a nonlinear MPC and estimated the interference using finite-time observer.

The review of the pertinent literature suggests that most studies focus on enhancing the robustness against environmental disturbances, whereas less focus has been put on disturbances pertaining to situation awareness, such as perception sensor noise. Hence, the aim of this study is to propose a methodology that enhances the robustness of path following for a MASS without violating the cross-track error threshold amid perception sensor noise by controlling the state space parameters of a deep reinforcement learning (DRL) agent.

The remainder of this study is as follows. Section 3 presents the proposed methodology. Section 4 presents the case study characteristics. Section 5 presents the results with pertinent discussion. Finally, Section 6 outlines the main findings, limitations, and outlook for future studies.

3. Methodology

The proposed methodology comprises six subsequent phases, as presented in Figure 1. Specifically, this methodology is adopted from the previous work of Lee et al. Lee et al. (2024), whereas a summary is given herein. In Phase 1, the main components are modeled to simulate the investigated scenarios. Specifically, the maneuverability of the ship, steering system, and navigating area is simulated using a four-degrees-of-freedom maneuvering modeling group model, a first-order linear differential equation model, and a two-dimensional binary occupancy map model, respectively. The distance measuring perception sensor used for the situation awareness is simulated using a time-of-flight equation, which is defined as:

$$d_t = \frac{cT}{2} + \epsilon_t \tag{1}$$

where d_t denotes the distance measured at each timestep $t \in \mathbb{Z}_{\geq 0}$; c, the speed of light; T, the time of flight; and ϵ_t, the Gaussian-based added noise with zero mean and variance σ^2.

In Phase 2, the Markov decision process (MDP) is formulated to model the decision-making problem of the investigated scenarios through the definition of states, actions, and rewards. The state of the agent at each t, as a set of state space parameters, is defined as:

$$S_t = [\psi_t, u_t, \dot{u}_t, v_t, \dot{v}_t, \dot{\psi}_t, \ddot{\psi}_t, U_t, \dot{U}_t, d_t, \mathcal{N}, e_{\mathrm{XT},t}, \dot{e}_{\mathrm{XT},t}, \ddot{e}_{\mathrm{XT},t}, e_{\mathrm{H},t}, \dot{e}_{\mathrm{H},t}, \ddot{e}_{\mathrm{H},t}] \tag{2}$$

where ψ_t denotes the heading angle of the ship; u_t, v_t, $\dot{\psi}_t$, $\dot{u}_t$, $\dot{v}_t$, and $\ddot{\psi}_t$, the surge, sway, yaw of the ship and their first-order time derivatives, respectively; U_t and $\dot{U}_t$, the resultant velocity of the ship and its first-order time derivative; $\mathcal{N}$, the value of σ^2; and $e_{\mathrm{XT},t}$, $\dot{e}_{\mathrm{XT},t}$, $\ddot{e}_{\mathrm{XT},t}$, $e_{\mathrm{H},t}$, $\dot{e}_{\mathrm{H},t}$, and $\ddot{e}_{\mathrm{H},t}$, the cross-track and heading errors of the ship from the path and their first- and second-order time derivatives, respectively. The action of the agent at each t is defined as:

$$A_t = [\delta_{\mathrm{C},t}] \tag{3}$$

where $\delta_{\mathrm{C},t}$, denotes the commanded rudder angle of the steering system. The reward of the agent at each t is defined as:

$$R_t = R_{1,t} + R_{2,t} + R_{3,t} + R_{4,t} \tag{4}$$

where $R_{1,t}$, $R_{2,t}$, $R_{3,t}$, and $R_{4,t}$ denote the path following, nominal navigation, actuator control,

and collision avoidance rewards, respectively, as:

$$R_{1,t} = k_1 + \frac{k_2}{e_{\mathrm{H},t}^2 + k_3 + k_4|e_{\mathrm{XT},t}|} + \frac{k_5}{e_{\mathrm{XT},t}^2 + k_6} \quad (5)$$

$$R_{2,t} = k_7 v_t^2 + k_8 \dot{\psi}_t^2 + k_9 \dot{v}_t^2 + k_{10} \ddot{\psi}_t^2 \quad (6)$$

$$R_{3,t} = k_{11} \delta_{\mathrm{C},t}^2 + k_{12} \dot{\delta}_{\mathrm{C},t}^2 \quad (7)$$

$$R_{4,t} = k_{13} d_t + k_{14} \quad (8)$$

where $k_i \in \mathbb{R}$ denotes the reward coefficient for $i \in \{1, 2, \ldots, 14\}$. Specifically, $k_1 \in \mathbb{R}_{\geq 0}$ and $k_1 = 0$ when $|e_{\mathrm{XT},t}| > B/2$, where B denotes the ship's beam, in order for the agent to receive less reward when the accuracy threshold for path following is violated.

In Phase 3, a DRL agent is trained in the formulated problem. Specifically, a deep deterministic policy gradient (DDPG)-based agent is setup that consists of 600 and 500 neurons for each of the actor and critic networks' two hidden layers, respectively. In addition, the agent is trained considering a training envelope $\sigma^2 \in [0, 25]$, where a random value with uniform distribution is investigated between the minimum and maximum value.

In Phase 4, the robustness of the agent is quantified in terms of a robustness metric, which is defined as:

$$RM = |e_{\mathrm{XT},t}|_{\max} \leq B/2 \quad (9)$$

It is worth noting that the defined robustness threshold is equal to the threshold for the path following reward, as presented in Equation 5.

In Phase 5, the state space parameters are controlled to enhance the robustness. Specifically, after the agent training, the $\mathcal{N}$ is decoupled from σ^2 to be controlled as an independent value.

Finally, in Phase 6, the robustness is verified by investigating various scenarios within and outside the training envelope, $\sigma^2 \in [0, 25]$ and $\sigma^2 > 25$, until the robustness threshold is violated.

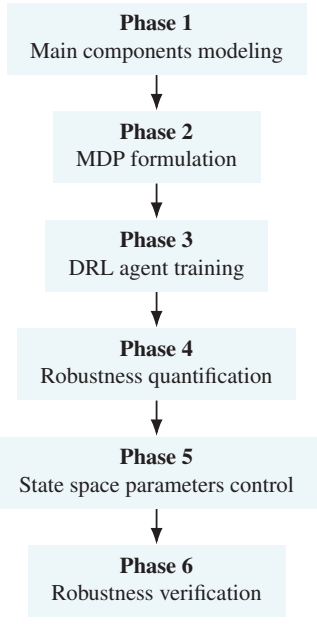

Fig. 1. Proposed methodology and its subsequent phases.

4. Case Study

The investigated case study considers a MASS that follows a global path, as presented in Figure 2. Specifically, the straight global path is generated by two waypoints located on a 5×5 km map. The same initial conditions are considered in each episode, including the initial location and nominal navigation of the MASS, but the σ^2. In addition, the episodes are diversified by randomly allocating a static obstacle on the path to avoid overfitting the agent. The MASS detects the obstacle using light detection and ranging (LIDAR), whose particulars are presented in Table 2. Finally, the main particulars of the MASS reflect the S-175 container ship, as presented in Table 1.

5. Results & Discussion

The training performance of the agent is presented in Figure 3. Considering the convergence of the return, the training is terminated at episode 5,820.

A total of six scenarios are investigated, whose robustness metrics are presented in Table 3. Specifically, robustness in scenario 3 is noted until $\sigma^2 = 53$ m^2 when $\sigma^2 = \mathcal{N}$, which is an increase

of 112% from its maximum trained value. This suggests the great generalization of DRL-based agents capable of making decisions beyond the trained values. However, robustness in scenario 4 is noted until $\sigma^2 = 1{,}442$ m^2 when $\sigma^2 \neq \mathcal{N}$, which is an increase of 5,668% from its maximum trained value. This suggests that the robustness against perception sensor noise during path following can be further enhanced, when the state space parameter $\mathcal{N}$ is controlled independently.

To highlight the effectiveness of this methodology, the agent is simulated at $\sigma^2 = 1{,}442$ m^2 and the differences between the robustness of path following when $\mathcal{N} = \sigma^2$ and $\mathcal{N} \neq \sigma^2$ are presented in Figures 4 and 5. It is noted that when $\mathcal{N}$ is not controlled the agent conducts a turning circle maneuver by commanding $|\delta_C|_{\max}$, thus failing to follow the path. However, when $\mathcal{N}$ is controlled the agent manages to follow the path without violating the robustness threshold, even when 99.7% of all LIDAR measurements have an uncertainty within ± 112.7 m.

Table 1. Main particulars of the MASS.

Particular	Symbol	Value
Length	L	175.0 m
Beam	B	25.4 m
Draft	T	8.5 m
Depth	D	11.0 m
Displaced volume	∇	21,222 m^3
Block coefficient	c_B	0.559

Table 2. Main particulars of the LIDAR.

Particular	Value
Maximum detecting range	1,341 m
Field of view	225 deg
Angular resolution	5.63 deg

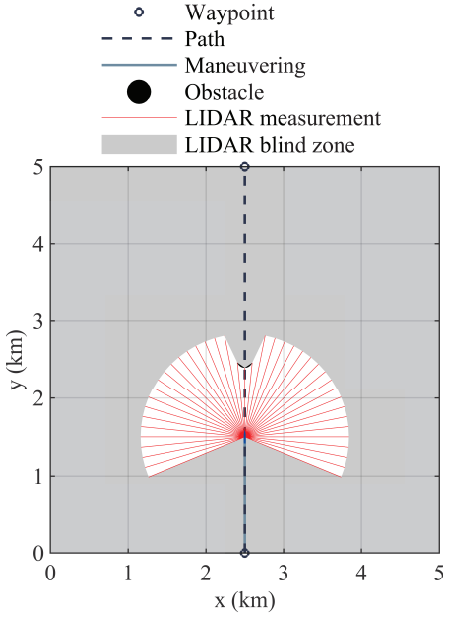

Fig. 2. Investigated case study.

Fig. 3. Training performance of the agent in terms of the return per episode. The sliding window for the moving median is 150 episodes.

Table 3. Robustness metric for path following without and with controlling the state space parameter $\mathcal{N}$.

Scenario	$\mathcal{N}$	σ^2	RM
1	0 m^2	0 m^2	0.3 m
2	25 m^2	25 m^2	2.1 m
3	53 m^2	53 m^2	11.5 m
4	0 m^2	1,442 m^2	12.7 m
5	25 m^2	784 m^2	12.7 m
6	53 m^2	191 m^2	12.7 m

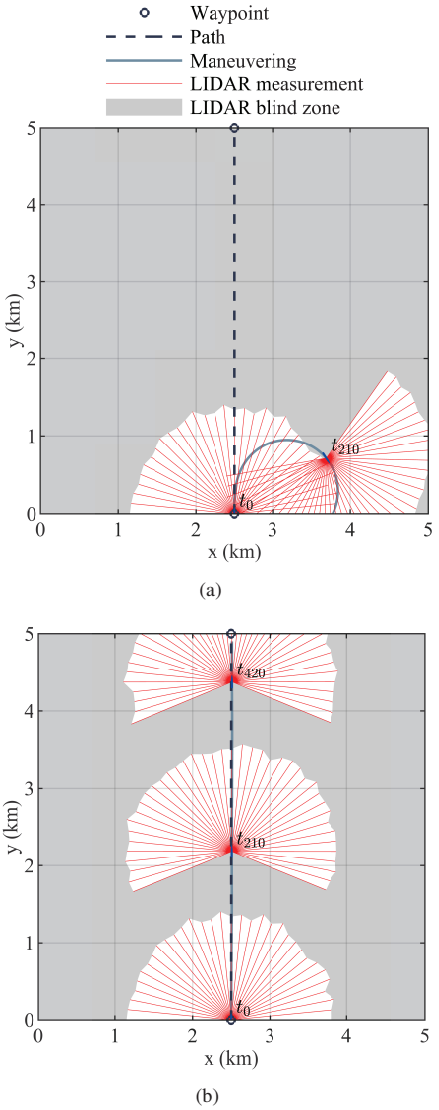

Fig. 4. Simulation result of the agent's maneuvering when (a) $\mathcal{N} = \sigma^2 = 1{,}442$ m^2 and (b) $\mathcal{N} = 0$ m^2 and $\sigma^2 = 1{,}442$ m^2.

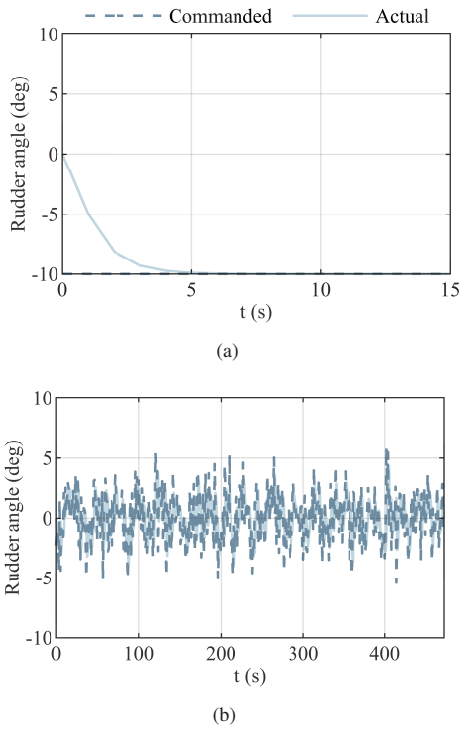

Fig. 5. Simulation result of the agent's actions in terms of δ_C when (a) $\mathcal{N} = \sigma^2 = 1{,}442$ m^2 and (b) $\mathcal{N} = 0$ m^2 and $\sigma^2 = 1{,}442$ m^2.

6. Conclusions

The robustness of intelligent agents used during the safety-critical operations of MASS is of paramount importance. The aim of this study was to propose a methodology that enhances the robustness of path following for a MASS without violating the cross-track error threshold amid perception sense noise by controlling the state space parameter of a deep reinforcement learning agent. A DDPG-based agent was trained to follow a predefined path at various noise levels between a minimum and maximum value. The case study considered a container vessel that used LIDAR for the situation awareness of its surrounding environment. The main findings of this study are as follows.

(i) Robustness of up to 112% from its maximum trained value was noted when $\mathcal{N}$ was not controlled, suggesting the generalization capabilities of DRL-based agents.

(ii) Robustness of up to 5,668% from its maximum trained value was noted when $\mathcal{N}$ was controlled, highlighting the effectiveness of the proposed methodology.

The main limitations of this study are the consideration of a simplistic noise model, the absence of other environmental disturbances, and the investigation of simple paths. Nonetheless, this study contributes towards the development of agents capable of making robust decisions during safety-critical operations under uncertainty.

Acknowledgement

The research presented in this article has been supported by the Research Council of Norway through the SFI Autoship project (project no. 309230).

References

BahooToroody, A., M. M. Abaei, O. V. Banda, P. Kujala, F. De Carlo, and R. Abbassi (2022). Prognostic health management of repairable ship systems through different autonomy degree; from current condition to fully autonomous ship. *Reliability Engineering & System Safety 221*, 108355.

Esfahani, H. N. and R. Szlapczynski (2021). Robust-adaptive dynamic programming-based time-delay control of autonomous ships under stochastic disturbances using an actor-critic learning algorithm. *Journal of Marine Science and Technology 26*(4), 1262–1279.

Fan, Y., H. Huang, and Y. Tan (2019). Robust adaptive path following control of an unmanned surface vessel subject to input saturation and uncertainties. *Applied Sciences 9*(9), 1815.

Hamon, R., H. Junklewitz, I. Sanchez, et al. (2020). Robustness and explainability of artificial intelligence. *Publications Office of the European Union 207*, 2020.

Huang, H. and Y. Fan (2019). Robust adaptive maneuvering control for an unmanned surface vessel with uncertainties. *IEEJ Transactions on Electrical and Electronic Engineering 14*(8), 1226–1235.

Kavallieratos, G., V. Diamantopoulou, and S. K. Katsikas (2020). Shipping 4.0: Security requirements for the cyber-enabled ship. *IEEE Transactions on Industrial Informatics 16*(10), 6617–6625.

Lee, P., G. Theotokatos, and E. Boulougouris (2024). Robust decision-making for the reactive collision avoidance of autonomous ships against various perception sensor noise levels. *Journal of Marine Science and Engineering 12*(4), 557.

Li, A., X. Hu, K. Dong, and B. Xiao (2024). An improved mpsp-based path-following control method for usv with input disturbances. *Optimal Control Applications and Methods*.

Liu, Z., S. Song, S. Yuan, Y. Ma, and Z. Yao (2022). Alos-based usv path-following control with obstacle avoidance strategy. *Journal of Marine Science and Engineering 10*(9), 1203.

Mu, D., G. Wang, and Y. Fan (2020). A time-varying lookahead distance of ilos path following for unmanned surface vehicle. *Journal of Electrical Engineering & Technology 15*(5), 2267–2278.

Mu, D., G. Wang, and Y. Fan (2022). Path following control strategy for underactuated unmanned surface vehicle subject to multiple constraints. *IEEJ Transactions on Electrical and Electronic Engineering 17*(2), 229–241.

Ren, Y., L. Zhang, W. Huang, and X. Chen (2023). Neural network-based adaptive sigmoid circular path-following control for underactuated unmanned surface vessels under ocean disturbances. *Journal of Marine Science and Engineering 11*(11), 2160.

Song, S., Z. Liu, S. Yuan, Z. Wang, and T. Wang (2024). A finite-time path following scheme of unmanned surface vessels with an optimization strategy. *ISA transactions 146*, 61–74.

Staessens, T., T. Lefebvre, and G. Crevecoeur (2022). Adaptive control of a mechatronic system using constrained residual reinforcement learning. *IEEE Transactions on Industrial Electronics 69*(10), 10447–10456.

Wen, Y., W. Tao, M. Zhu, J. Zhou, and C. Xiao (2020). Characteristic model-based path following controller design for the unmanned surface vessel. *Applied Ocean Research 101*, 102293.