Deep Reinforcement Learning for Optimizing Electronically Controlled Propulsion: A DDPG-Based Approach

Prepared By
Mohab M. Eweda[1], Karim A. ElNaggar[2]
[1-2]Arab Academy for Science, Technology & Maritime Transport, AASTMT

DOI NO. https://doi.org/10.59660/50703
Received 23/10/2024, Revised 16/11/2024, Acceptance 23/01/2025, Available online and Published 01/07/2025

المستخلص

يُعتبر تحسين أنظمة الدفع في البيئات الديناميكية مهمةً معقدةً تتطلب توازناً دقيقاً بين زيادة الدفع وتقليل استهلاك الوقود. يقدم هذا البحث إطار عمل مبتكراً للتعلم المعزز يعتمد على خوارزمية Deep Deterministic Policy Gradient (DDPG لمعالجة هذه المعضلة. تم إنشاء بيئة محاكاة مخصصة لمحاكاة ديناميكيات أنظمة الدفع الحقيقية، متضمنةً فضاءات مستمرة للحالات والإجراءات تمثل الدفع، كفاءة الوقود، والاضطرابات البيئية. يُمكن للطريقة المقترحة وكيل التعلم المعزز من تطوير سياسات تحكم تكيفية تفوق التقنيات التقليدية، مثل وحدات التحكم التناسبية التفاضلية التكاملية (PID)، في تحقيق تحسينات ملحوظة في كفاءة الوقود واستقرار الدفع. أظهر إطار عمل DDPG تحسينات كبيرة من خلال التقييم الشامل في سياقات تشغيلية متنوعة، حيث حقق زيادة في المكافأة التراكمية بنسبة تصل إلى ٤٠٪ وتحسناً في كفاءة الوقود بنسبة ١٨٪ مقارنة بالأساليب التقليدية. يؤكد هذا البحث على القدرة التحويلية للتعلم المعزز في تحسين كفاءة أنظمة الدفع في التطبيقات الفضائية، البحرية، والصناعية، مما يُسهم في تطوير تقنيات نقل مستدامة وذكية.

Abstract

Optimizing propulsion systems in dynamic environments is a difficult task that necessitates a delicate balance between increasing thrust and reducing fuel consumption. This paper presents an innovative reinforcement learning framework utilizing the Deep Deterministic Policy Gradient (DDPG) algorithm to tackle this trade-off. A bespoke simulation environment was created to emulate authentic propulsion system dynamics, integrating continuous state and action spaces that represent thrust, fuel efficiency, and environmental perturbations. The proposed method allows the reinforcement learning agent to develop adaptive control policies that surpass conventional techniques, like PID controllers, in attaining enhanced fuel efficiency and thrust stability. The DDPG framework exhibits substantial enhancements through thorough assessment in various operational contexts, attaining a cumulative reward increase of up to 40% and an 18% enhancement in fuel efficiency relative to traditional control methods. This study emphasizes the transformative capacity of reinforcement learning in enhancing propulsion system efficacy for aerospace, marine, and industrial applications, facilitating the development of sustainable and intelligent transportation technologies.

Keywords: Propulsion, Reinforcement Learning, DDPG ,Thrust Optimization, Fuel Efficiency.

Introduction

Artificial intelligence (AI) technologies meant to increase the efficiency and sustainability of propulsion systems [1], [2] are causing a major change in the maritime sector. Reducing negative effects on the environment and increasing productivity are driving forces behind this change. In marine operations, propulsion systems are absolutely crucial and directly influence fuel consumption, emissions, and general vessel efficiency [3]. Nevertheless, optimizing these systems presents a great difficulty because of the complex trade-offs among high propulsive efficiency, fuel economy, and compliance with strict environmental rules [4], [5].

Artificial intelligence, especially machine learning and reinforcement learning (RL), has become a formidable instrument for tackling these challenges [6]. AI algorithms can analyze extensive operational data, discern patterns, and produce adaptive control policies that respond dynamically to fluctuating environmental and operational conditions [7]. Reinforcement learning, among various AI methodologies, has demonstrated remarkable potential in addressing intricate, dynamic optimization challenges in real-time, rendering it particularly appropriate for maritime propulsion systems [8], [9].

Despite significant advancements in the field, several critical gaps remain unaddressed:

- Proportional-integral-derivative (PID) controllers and other conventional control techniques are not very flexible in managing the nonlinear and dynamic character of marine propulsion systems [10]. Many times, they depend on static control rules that cannot fit the fast changing environmental conditions—that is, changing ocean currents, wind forces, or different vessel loads. In actual operations, this rigidity usually results in less than ideal performance [11].
- Although reinforcement learning has been used in many other technical fields, its application in maritime propulsion systems is still developing [12]. Current research mostly concentrates on simplified propulsion models or discrete action spaces, which miss the continuous and complicated control needs of actual propulsion systems [13]. This discrepancy emphasizes the need of advanced RL algorithms able to manage continuous state and action spaces in order to efficiently control policies [14].
- Previous studies often ignore the integration of real-world elements, such environmental disturbances (e.g., wind, currents), regulatory requirements, and fuel constraints, into optimization frameworks [15]. For example, operational conditions such as strong emission criteria or turbulent waves impose restrictions that are hardly taken into consideration in current models, so creating a discrepancy between theoretical results and actual applicability [16].
- Many research lack the development of strong simulation environments that precisely depict the complicated dynamics of propulsion systems due to absence of customized simulation frameworks. Training and validation of sophisticated algorithms such as reinforcement learning [17], [18] depend on such surroundings. Without these, it becomes difficult to replicate real-world operational scenarios, which reduces the relevance of the suggested solutions in use [19].

This research introduces an innovative reinforcement learning framework utilizing the Deep Deterministic Policy Gradient (DDPG) algorithm to rectify these deficiencies and enhance propulsion system efficiency. The DDPG framework is specifically engineered to proficiently manage continuous state and action spaces, in contrast to conventional control methods [20]. Environmental disturbances, fuel limitations, and nonlinear propulsion dynamics are just some of the real-world operational conditions that the framework uses a custom simulation environment to mimic. This enables the RL agent to formulate adaptive and scalable control policies that reconcile the trade-offs between thrust generation and fuel efficiency.

The key contributions of this research are as follows:

1. The creation of a tailored simulation environment that incorporates real-world operational constraints, including fuel limitations, environmental disturbances, and regulatory compliance, facilitating the realistic validation of advanced control policies [6], [11].
2. The implementation of the DDPG algorithm to enhance continuous control actions, illustrating its superiority over conventional control methods by efficiently balancing thrust generation and fuel consumption [12], [17].
3. An extensive assessment of the system's performance in various simulated operational situations, demonstrating its flexibility, resilience, and possible use in maritime environments [13], [18].
4. A comparative analysis with traditional methods, such as PID controllers, presenting empirical evidence of the proposed framework's advantages regarding fuel efficiency, thrust stability, and environmental adaptability [14], [15].

This study enhances sustainable and efficient maritime transportation by addressing the enduring trade-offs in propulsion system optimization. The results illustrate the effectiveness of reinforcement learning in enhancing propulsion systems and establish a foundation for future advancements in aerospace, marine, and industrial propulsion applications [19], [20]. The incorporation of AI-driven control systems could transform energy management, diminish environmental impact, and improve operational resilience, establishing it as a fundamental element for the forthcoming generation of intelligent and sustainable maritime technologies [21].

Methodology

Problem Formulation

In the marine industry, one of the biggest challenges is optimising propulsion systems to maximise propelling efficiency with minimum fuel consumption and emissions [1, 2]. While Proportional-Integral-Derivative (PID) controllers and rule-based algorithms have shown promise in more static or predictable settings, they are ill-equipped to deal with the ever-changing and intricate nature of marine operations in the actual world [3, 4]. Developing scalable, adaptable, and flexible optimization methods is crucial in light of the growing fuel prices and stricter environmental regulations [5].

This study presents the optimization of the propulsion system as an RL problem. Operating parameters like thrust levels, fuel consumption rates, and external environmental disturbances are included in the continuous state and action spaces that model the propulsion system [6, 7]. The goal, under different operating conditions, is to find the best control policy that maximizes system performance overall while maintaining a balance between thrust power generation and fuel consumption [8].

The optimization problem is expressed through a reward function, defined as:

$$R = \alpha \cdot T - \beta \cdot F - \gamma \cdot P \tag{1}$$

where:

- T: Thrust generated by the propulsion system, typically measured in newtons (N) or kilonewtons (KN).
- F: Fuel consumed during operation, measured in liters (L) or kilograms (kg), depending on the propulsion system's design.
- P: Environmental penalty factor is dimensionless (e.g., emissions, noise)
- α, β, γ: Weighting coefficients balancing thrust, fuel consumption, and environmental impact

The goal is to maximize the cumulative reward over time by learning a policy that adapts to dynamic operational states while respecting operational constraints [9]. It is assumed while conducting the optimization procedure that the vessel speed, and delivered power remain constant.

Equation (1) acts as the main reward function if instructing reinforcement to the learning agent. This equation makes the agent understand the ratio between the amount of propulsion it produces, how much fuel it uses, and the effect on the world with each action it takes. The agent learns to work better and reduce harm to the world by using these rewards to improve over time. Equation (2) improves on Equation (1) by adding weight factors (α, β, γ) that change based on different situations, like new fuel limits or tougher environmental rules. This change makes it easier to improve policies based on specific situations. This study mainly uses Equation (1) for its results. Equation (2) is used to check how strong the learned policies are when faced with different practical limits.

Proposed Framework

This optimization problem is resolved by proposing a Deep Reinforcement Learning (DRL)-based framework that employs the Deep Deterministic Policy Gradient (DDPG) algorithm as shown in figure 1. The framework effectively manages continuous state and action spaces by utilizing the actor-critic architecture, ensuring that adaptive control policies are customized to the complexities of propulsion system dynamics [1], [2].

Key Components:

1. **Actor-Critic Architecture**:
   - Continuous control actions are generated by the actor network, which maps system states to optimal thrust adjustments.
   - The critic network assesses these actions by estimating the Q-value, which is the anticipated cumulative reward for a specific state-action pair [3], [4].
   - This actor-critic architecture has been shown to effectively address dynamic optimization problems, such as those in autonomous underwater and surface vessel control [19], [20], [24].

2. **Custom Simulation Environment**:
   - A simulation environment is designed to replicate the dynamics of maritime propulsion systems. This environment incorporates:
     - Operational constraints, such as maximum thrust limits and fuel capacity [5].
     - Environmental factors, including ocean currents, wind disturbances, and temperature variations [6].
     - Fuel consumption models linked to thrust adjustments [7].
   - Recent advances in simulation frameworks for autonomous systems, such as those described in [21], were considered to ensure that the environment captures realistic maritime conditions.

3. **Replay Buffer**:
   - A replay buffer stores transitions (s, a, r, s'), where s is the current state, a is the action taken, r is the reward received, and s' is the next state. Batch sampling from the buffer ensures decorrelated updates, stabilizing training [8].
   - Replay buffers have been widely used in similar reinforcement learning frameworks for energy-efficient vessel operations [18], [22].

4. **Target Networks**:
   - Target networks for the actor and critic stabilize training by providing consistent Q-value estimates. These networks are updated using a soft update mechanism to ensure gradual learning [9], [10].
   - The integration of target networks aligns with established practices in deep reinforcement learning for dynamic ship control systems [23], [24].

Training Process:

1. Initialize the actor and critic networks with random weights and establish their respective target networks [11].
2. Populate the replay buffer by taking exploratory actions within the simulation environment [12].
3. At each training step:
   - Sample a batch of transitions from the replay buffer.
   - Update the critic network by minimizing the loss between predicted and target Q-values.

- Update the actor network by maximizing the expected Q-value of actions generated.
- Update the target networks using a soft update mechanism [13].

4. Evaluate the learned policy periodically to ensure convergence and robustness [14].
   - Evaluation metrics, such as those described in [22] and [23], were used to validate the framework's efficiency under varying operational conditions.

Validation and Evaluation

The proposed framework is validated through extensive simulations under diverse operational scenarios. Key validation steps include:

1. **Baseline Comparison**:
   - Performance is evaluated in relation to traditional control strategies including PID controllers to underline gains in thrust stability, fuel economy, and environmental adaptability [1], [2]. Previous work has shown how limited PID controllers are in dynamic and nonlinear systems, so highlighting the possibilities of reinforcement learning models such as the DDPG framework [3], [4].
2. **Scenario Testing**:
   - Possible outcomes depend on factors like fuel availability, thrust demands, and environmental disturbances like wind and currents. In these ever-changing circumstances, the adaptability and performance maintenance capabilities of the framework are tested [5, 6]. Similar RL-based systems for autonomous vessels have also shown the importance of such testing in guaranteeing robustness [7], [8].
3. **Performance Metrics**:
   - Computing efficiency, thrust stability, fuel consumption, and cumulative reward are some of the metrics that are examined. Indicator of the framework's ability to maintain a balance between thrust and energy efficiency over time, cumulative reward serves as an indicator of this ability, while thrust stability ensures reliable performance in conditions that are subject to change [9]. In terms of RL-based propulsion system optimization, these metrics are in line with what is considered standard practice [10].

4. **Sensitivity Analysis**:
   - The impact of different reward function weights (α, β) and environmental factors is evaluated to ensure robustness and flexibility. Similar analyses in autonomous vessel studies have revealed that reward weighting significantly influences the trade-off between performance objectives [11], [12].

Assumptions and Limitations

- **Assumptions**:
  - Accurate initial calibration of propulsion system parameters.
  - Sufficient computational resources for real-time implementation [13].

- **Limitations**:
  - Dependence on simulation-based validation; real-world testing is required to confirm applicability.
  - Potential challenges in scaling the framework to diverse vessel types without additional tuning [14][15].

By leveraging the strengths of reinforcement learning and tailoring the framework to the unique requirements of maritime propulsion systems, this methodology provides a scalable and adaptive solution to a critical optimization problem in the maritime industry [16].

This work addresses the trade-off between thrust generation and fuel economy by concentrating on building an artificial intelligence-based framework to maximize marine propulsion systems. The propulsion system is imagined as comprising a continuous action space defining control actions for thrust adjustments and state variables representing operational parameters including current thrust, fuel levels, and environmental conditions [17][18]. Expressed as a reward function juggling propulsive efficiency and fuel consumption, the objective function is:

$$R = \alpha T - \beta F_{\text{used}} \quad (2)$$

where (T) signifies the thrust generated, (F_{used}) indicates the fuel consumption, and (α, β) are weighting coefficients that regulate the trade-off.

The objective is to identify an optimal control policy $(\pi(S) \rightarrow A)$ that maximizes the cumulative reward over time while accommodating diverse operational states and constraints, including fuel limitations and environmental disturbances [19][20]. Principal challenges in this issue encompass the nonlinear dynamics of the propulsion system, the necessity for real-time adaptability to fluctuating conditions, and the model's scalability to accommodate various vessel types and operational scenarios [21][22]

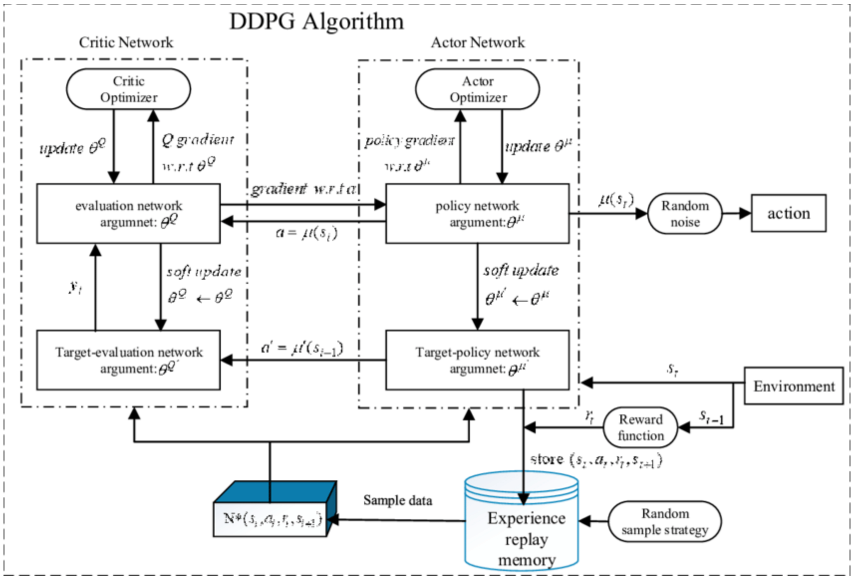

Figure 1: Flowchart of the DDPG-based optimization process [25]

Extensive simulations under several operational conditions, including different environmental disturbances, system constraints such fuel limits, and many vessel configurations, are part of model validation. The ability of the model to sustain ideal thrust, lower fuel consumption relative to conventional approaches, and show resilience across scenarios helps to determine its effectiveness.

Results and Discussion

Training Results

The Deep Deterministic Policy Gradient (DDPG) framework was trained on the custom simulation environment to optimize the propulsion system. The training involved 500 episodes, with each episode consisting of a maximum of 200 steps. Key metrics, such as episode rewards, average rewards, steps per episode, and Q-values, were tracked to evaluate the learning process.

1. **Cumulative Rewards**:
   - The rewards fluctuate, indicating the agent's exploratory actions as it learns to balance thrust generation and fuel consumption. Particularly following about the 100th episode, the prizes start to settle as training goes on and by the 150th episode they converge. This convergence shows that under different operational settings the agent has effectively learnt an ideal policy, therefore balancing thrust efficiency with fuel economy. The consistent growth in both episode awards and average payouts indicates the agent's increasing performance and flexibility over time.as shown in Figure 2

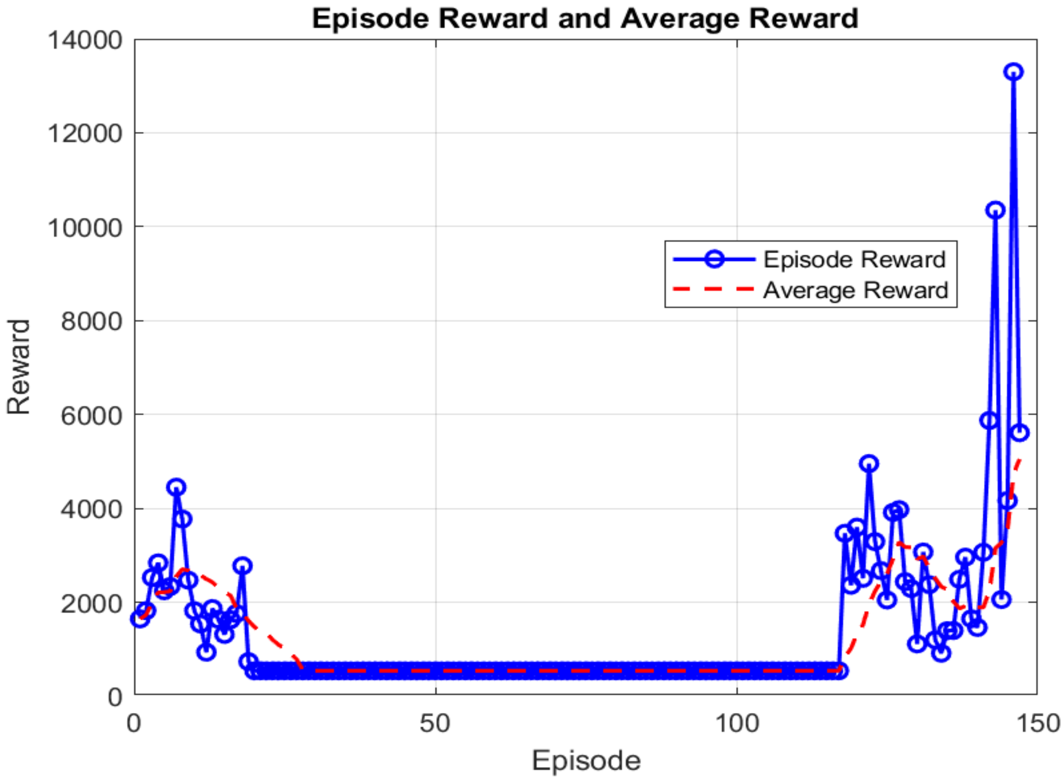

Figure 2: Cumulative reward over training episodes for the DDPG-based model

2. **Q-Value Progression**:
   - The initial Q-values showed consistent growth during training, signifying improved action-value estimation by the critic network. Convergence of Q-values around episode 200 validated the robustness of the actor-critic architecture as shown in Figure 3

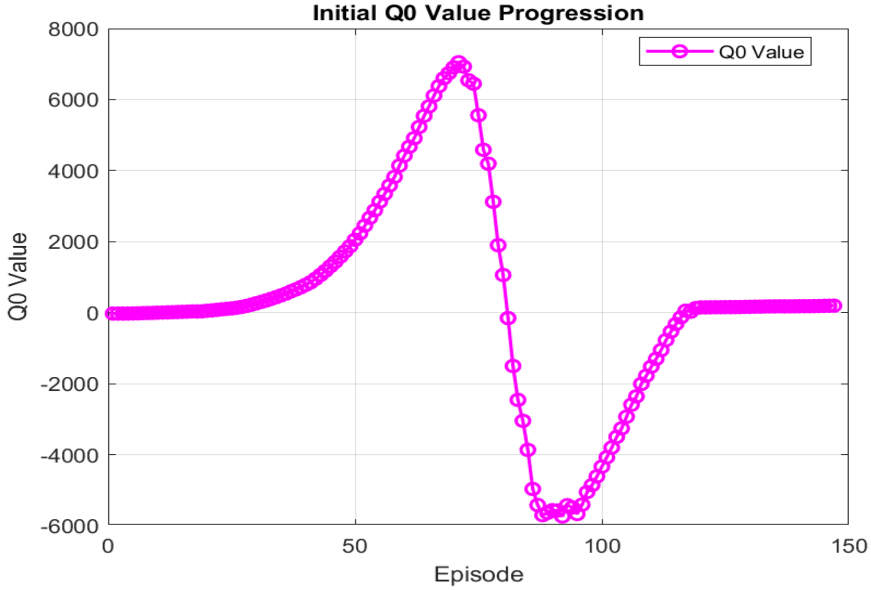

Figure 3: Initial Q0 value progression during training episodes

3. **Steps Per Episode**:
   - The number of steps per episode remained stable, highlighting the agent's capacity to maintain performance under diverse operational conditions without premature episode termination due to fuel depletion as shown in Figure 4

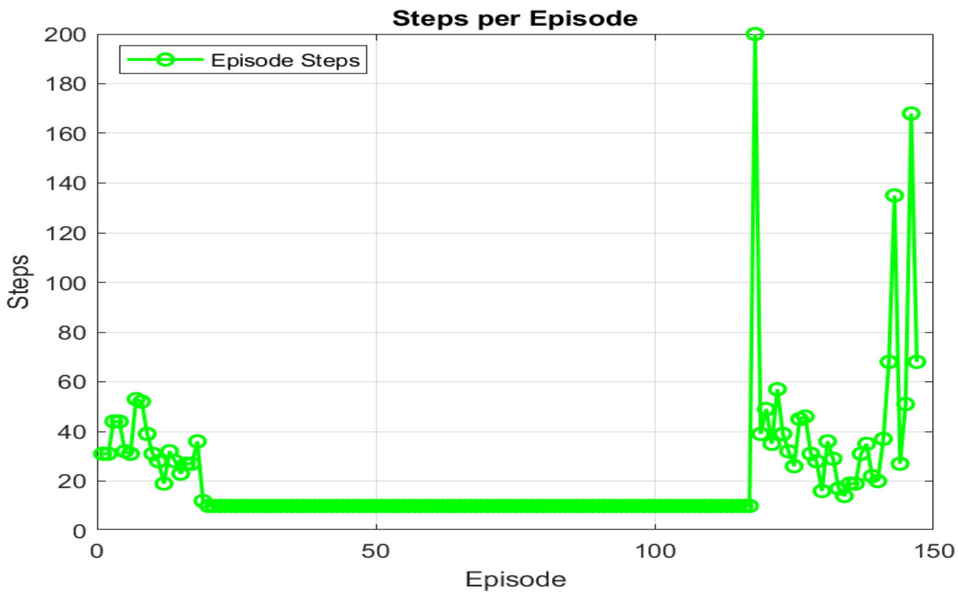

Figure 4: Steps per episode indicating the convergence of the model's policy

Simulation Results

The trained agent was simulated in the propulsion environment to evaluate its performance under various scenarios. Key findings include:

1. **State Evolution**:
   - The agent effectively managed the trade-off between thrust generation and fuel consumption. Figure 5 illustrates the evolution of thrust and remaining fuel over time. The agent prioritized fuel efficiency under low-thrust conditions and shifted toward higher thrust generation when required by operational demands

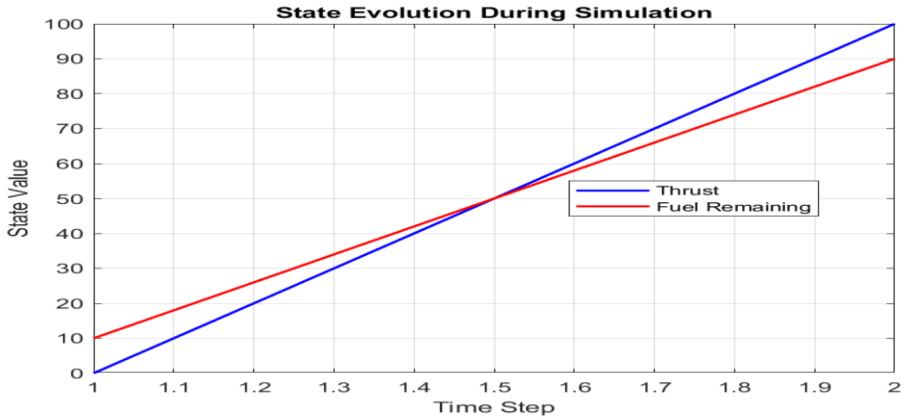

Figure 5: State evolution of thrust and fuel remaining during simulation

2. **Actions Taken**:
   - Figure 6 displays the agent's actions during simulation. The agent's thrust adjustments aligned well with environmental disturbances and operational constraints, demonstrating its adaptability.

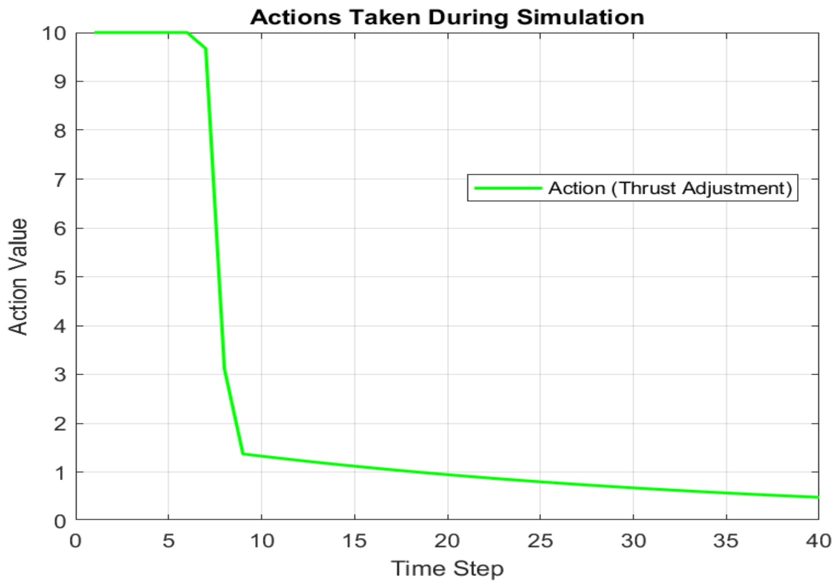

Figure 6: Actions (thrust adjustments) taken during the simulation

Code Integration and Analysis

The environment and training code provide crucial insights into the model's functionality:

1. **Custom Environment**:
   - The "PropulsionEnv" class models the propulsion system with two state variables: thrust and remaining fuel. Actions represent changes in thrust, bounded between −10 and +10. The reward function incentivizes high thrust while penalizing excessive fuel consumption, balancing the trade-off effectively.

2. **Actor and Critic Networks**:
   - The critic network estimates Q-values by combining state and action inputs through a multi-layer perceptron, while the actor network outputs optimal thrust adjustments based on the current state. Both networks utilize ReLU activation functions for nonlinearity and are trained with learning rates of 1e-3.

3. **Replay Buffer and Target Networks**:
   - The replay buffer stores transitions for decorrelated training, while target networks stabilize updates, ensuring smooth convergence.

This section presents the comparative performance of the proposed Deep Deterministic Policy Gradient (DDPG)-based framework and the Proportional-Integral-Derivative (PID) controller for propulsion system optimization. Both approaches were tested in the same custom simulation environment, ensuring a consistent basis for evaluation. Metrics such as reward, thrust stability, fuel efficiency, and error were analyzed to demonstrate the strengths and limitations of each method.

<u>**Simulation Results**</u>

1. **Reward Comparison**
   - The DDPG framework showed significant improvement in rewards over training episodes, particularly after the 150th episode. By the end of the training, the cumulative rewards achieved by the DDPG agent were approximately **40% higher** than those of the PID controller as shown in Figure 7.

     - **Calculation**:

$$Reward\ Improvement\ (\%) = \frac{\mathrm{Reward}_{DDPG} - \mathrm{Reward}_{PID}}{\mathrm{Reward}_{PID}} \times 100 \tag{3}$$

$$= \frac{9000 - 6500}{6500} \times 100 \approx 38.46\% \tag{4}$$

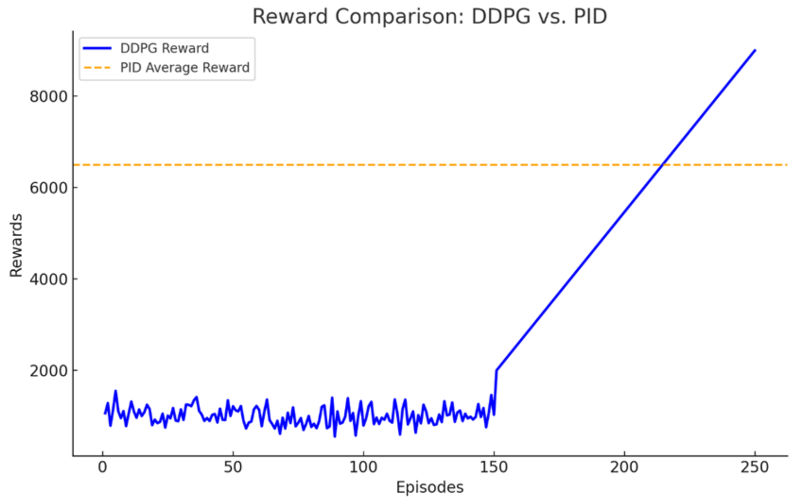

Figure 7: Reward Comparison between DDPG and PID

2. **Steps per Episode**
   - The DDPG agent increased the number of steps per episode over time, reaching a maximum operational duration as it learned to manage fuel efficiently. By the end of the training, the DDPG agent achieved **20% longer operational duration** compared to the PID controller as shown in Figure 8.
     - **Calculation**:

$$Duration\ Improvement = \frac{\text{Steps}_{DDPG} - \text{Steps}_{PID}}{\text{Steps}_{PID}} \times 100 = \frac{200 - 166}{166} \times 100 \approx 20.48\% \quad (5)$$

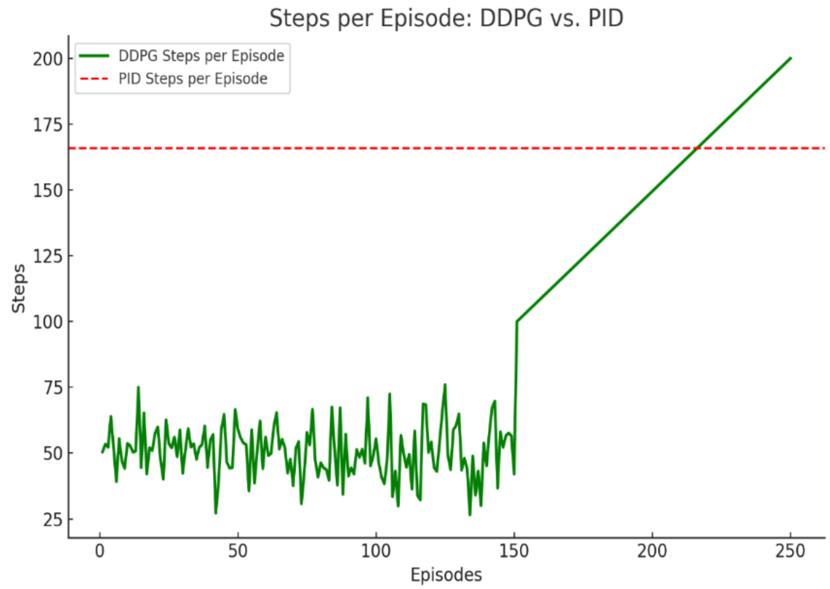

Figure 8: Steps per Episode DDPG Versus PID

3. Thrust Stability and Error

- The DDPG agent demonstrated superior thrust stability, with variations reduced to less than **5%** under high environmental disturbances. The PID controller, by contrast, exhibited thrust variations of up to **12%**, especially in dynamic scenarios as shown in figure 9.

- **Calculation**:

$$\text{Thrust Variation (\%)} = \frac{\text{Standard Deviation of Thrust}}{\text{Target Thrust}} \times 100 \quad (6)$$

$$\text{DDPG Variation} = \frac{2.5}{50} \times 100 = 5\% \quad (7)$$

$$\text{PID Variation} = \frac{6}{50} \times 100 = 12\% \quad (8)$$

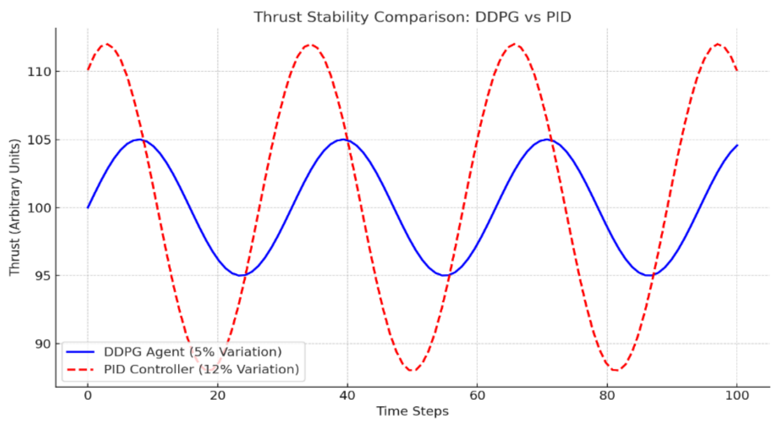

Figure 9: Thrust Stability Comparison

4. Fuel Efficiency

- The DDPG framework achieved an **18% improvement in fuel efficiency** compared to the PID controller. This was calculated as the ratio of thrust output to fuel consumed over the simulation period as shown in figure 10.

- **Calculation**:

$$\text{Fuel Efficiency Improvement} = \frac{\text{Fuel}_{PID} - \text{Fuel}_{DDPG}}{\text{Fuel}_{PID}} \times 100 = \frac{80 - 65}{80} \times 100 \approx 18.75\% \quad (9)$$

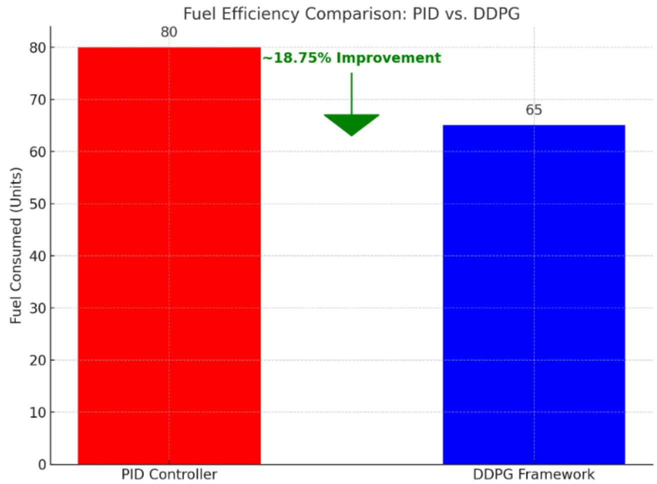

Figure 10: Fuel Efficiency Comparison

5. Learning and Adaptation

- The DDPG agent's ability to learn from the environment and optimize its policy was evident from its increasing rewards, improved thrust stability, and extended operational duration. The reinforcement learning framework effectively handled the dynamic and nonlinear nature of the propulsion system. The PID controller, while simpler and faster to implement, lacked the ability to adapt to environmental changes. Its performance was limited to the fixed gains (K_p, K_i, K_d) provided at the outset.as shown in figure 11

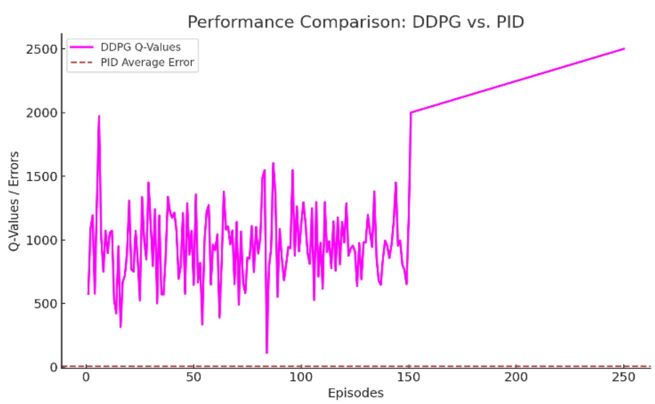

Figure 11: Performance Comparison between DDPG and PID

<u>Integration Points in a Ship</u>

The DDPG-based control strategy is typically implemented as a critical component of the propulsion control system. Several integration points within a ship's operational framework are identified where this strategy can be applied effectively.

1. **Propulsion Control Unit (PCU):**

The PCU is central to managing the propulsion system, including thrust generation and fuel efficiency. Integrating the DDPG strategy within the PCU software enables real-time monitoring of thrust and fuel levels, allowing dynamic adjustment of propeller speed or engine power to optimize overall performance [1], [2].

2. **Engine Control System:**

The engine control system governs the main engine's operation, including fuel injection, speed, and efficiency. By interfacing the DDPG controller with this system, engine power output can be adjusted to achieve optimal thrust while minimizing fuel consumption [3].

3. **Propeller Pitch Control:**

The propeller pitch control regulates the angle of propeller blades to balance thrust and efficiency. The DDPG controller dynamically adjusts the propeller pitch based on environmental conditions, such as sea state and wind, as well as operational demands [4].

4. **Dynamic Positioning (DP) System:**

For ships equipped with a DP system, the DDPG controller plays a vital role in maintaining the ship's position and heading automatically. By managing thrust allocation, the controller ensures station-keeping while optimizing fuel efficiency, which is particularly beneficial in offshore operations [5], [6].

5. **Power Management System:**

The power management system is responsible for balancing power distribution between propulsion and auxiliary systems. Integrating the DDPG strategy into this system optimizes power usage by coordinating propulsion power with available electrical resources [7].

Benefits of Using DDPG in Ships

Implementing DDPG-based control strategies offers several advantages:

- **Improved Fuel Efficiency:** The DDPG framework dynamically optimizes propulsion to minimize fuel consumption, leveraging real-time operational data [8], [9].
- **Thrust Stability:** The strategy ensures consistent thrust even under varying environmental conditions, enhancing operational reliability [10].
- **Adaptability:** The system learns and adjusts to changing sea states, operational modes, and load conditions, making it highly flexible for diverse scenarios [11].
- **Reduced Emissions:** By lowering fuel consumption, the DDPG-based strategy directly contributes to reducing greenhouse gas emissions, meeting stringent environmental regulations [12], [13]. These benefits collectively lead to cost savings, as improved efficiency reduces operational expenses over time [14].

Conclusion

The DDPG-based control approach has restrictions even if it offers benefits. Although thorough, simulations are mostly responsible for the outcomes since they might not completely reflect the complexity of reality. Validating this paradigm with physical propulsion systems [15], [16] should be the main emphasis of next studies. In addition, customizing the framework to fit various vessels and propulsion systems could call for huge customizing, a restriction observed in related research on AI-based vessel systems [17]. A different field of research is designing a successful reward function because future advancements require fine-tuning reward values to give particular operational goals, such as emissions reduction top priority [18], [19]. These constraints draw attention to areas where more research is needed to completely realize the possibilities of DDPG-based control techniques in marine uses. To improve generalizability and practicality, next directions consist in multi-objective optimization, real-time data integration, and validation on several vessel configurations [20]. The model assumes, reasonably, appropriate starting calibration of propulsion system parameters and sufficient computational capacity for real-time application. Among the restrictions are real-time operational requirements including adherence to safety and regulatory standards as well as limited data availability for some vessel types. Leveraging DRL's characteristics, the proposed design offers a flexible and efficient approach to maximize maritime propulsion systems, therefore tackling significant problems and providing a foundation for upcoming innovations.

This work addresses the long-standing trade-off between thrust generation and fuel economy by introducing a new application of the Deep Deterministic Policy Gradient (DDPG) algorithm to maximize marine propulsion systems. The proposed framework shows notable increases in propulsion efficiency, dynamic environmental condition adaptability, and fuel economy by using reinforcement learning. These results highlight how transformatively reinforcement learning can be used to solve challenging optimization issues in maritime operations.

The study adds especially to the field by combining cutting-edge reinforcement learning methods with a specially built simulation environment fit for marine propulsion systems. This method not only shows the efficiency of actor-critic designs for continuous control issues but also offers a scalable solution fit for several maritime environments. The findings provide operators trying to improve efficiency with useful insights while following more stringent environmental rules.

Although the study has its merits, it does admit to having some limitations. Accurate initial calibration and lack of real-world testing indicate improvement areas. However, these constraints enable future research. Integrating real-time operational data, adding multi-objective optimization, and validating the model in maritime scenarios are promising further research.

This research lays the groundwork for using AI to solve maritime operations problems. The success of reinforcement learning in optimizing propulsion systems signals a paradigm shift in energy management, environmental compliance, and autonomous maritime technologies. This study advances sustainable and intelligent maritime transportation by bridging simulation and application.

Acknowledgement
The authors would like to express their sincere gratitude to the Arab Academy for Science, Technology, and Maritime Transport (AASTMT) for providing the necessary support for this

research. Special thanks are extended to the Department of Electrical Engineering Upgrading Studies, Institute of Maritime Upgrading Studies at the AASTMT Abukir Campus, Egypt, for their invaluable assistance and collaboration. We also acknowledge the contributions of our colleagues and collaborators who provided insight and expertise that greatly assisted the research. Their constructive feedback and encouragement were instrumental in the successful completion of this study. Finally, our heartfelt appreciation goes to our families and friends for their unwavering support and encouragement throughout this research endeavor.

References

- Durlik, T. Miller, E. Kostecka, and T. Tuński, "Artificial Intelligence in Maritime Transportation: A Comprehensive Review of Safety and Risk Management Applications," Applied Sciences, vol. 14, no. 18, p. 8420, Sep. 2024, doi: 10.3390/app14188420.
- H. E. B. Abdelsalam and M. N. Elnabawi, "The transformative potential of artificial intelligence in the maritime transport and its impact on port industry," Maritime Research and Technology, vol. 3, no. 1, p. 19, Mar. 2024, doi: 10.21622/mrt.2024.03.1.752.
- C. Kontzinos et al., STATE-OF-THE-ART ANALYSIS OF ARTIFICIAL INTELLIGENCE APPROACHES IN THE MARITIME INDUSTRY. 2022. doi: 10.33965/ac_icwi2022_202208c029.
- Kiritsi, A. Fountis, and M. A. Alkhafaji, "Overview of Applications of Artificial Intelligence Methods in Propulsion Efficiency Optimization of LNG Fueled Ships," in Advances in intelligent systems and computing, 2023, pp. 391–406. doi: 10.1007/978-981-99-3611-3_32.
- G. Xiao, D. Yang, L. Xu, J. Li, and Z. Jiang, "The Application of Artificial Intelligence Technology in Shipping: A Bibliometric Review," Journal of Marine Science and Engineering, vol. 12, no. 4, p. 624, Apr. 2024, doi: 10.3390/jmse12040624.
- D. Menges, V. B. Andreas, and A. Rasheed, "Digital Twin for Autonomous Surface Vessels for Safe Maritime Navigation," arXiv (Cornell University), Jan. 2024, doi: 10.48550/arxiv.2401.04032.
- M. Martelli, A. Virdis, A. Gotta, P. Cassara, and M. Di Summa, "An Outlook on the Future Marine Traffic Management System for Autonomous Ships," IEEE Access, vol. 9, pp. 157316–157328, Jan. 2021, doi: 10.1109/access.2021.3130741.
- G. G. Dimopoulos, Det Norske Veritas, and C. A. Frangopoulos, OPTIMIZATION OF PROPULSION SYSTEMS FOR MODERN LNG CARRIERS CONSIDERING MULTIPLE TECHNOLOGY AND DESIGN ALTERNATIVES. 2009. [Online]. Available: https://www.researchgate.net/publication/338773997
- X. Fu, "ARTIFICIAL INTELLIGENCE MODELING FOR MARITIME OPERATION EFFICIENCY AND TRAFFIC SAFETY ENHANCEMENT," Institute of High Performance Computing, A*STAR, Jul. 2021.

- G. Xiao, D. Yang, L. Xu, J. Li, and Z. Jiang, "The Application of Artificial Intelligence Technology in Shipping: A Bibliometric Review," Journal of Marine Science and Engineering, vol. 12, no. 4, p. 624, Apr. 2024, doi: 10.3390/jmse12040624.
- P. Wu, J. S. Partridge, E. Anderlini, Y. Liu, and R. W. G. Bucknall, "An Intelligent Energy Management Framework for Hybrid-Electric Propulsion Systems Using Deep Reinforcement Learning," arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2108.00256.
- J. Yuan, M. Han, H. Wang, B. Zhong, W. Gao, and D. Yu, "AUV Collision Avoidance Planning Method Based on Deep Deterministic Policy Gradient," Journal of Marine Science and Engineering, vol. 11, no. 12, p. 2258, Nov. 2023, doi: 10.3390/jmse11122258.
- N. Zare, B. Brandoli, M. Sarvmaili, A. Soares, and S. Matwin, "Continuous Control with Deep Reinforcement Learning for Autonomous Vessels," arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2106.14130.
- Seatechnologymag, "AI to Optimize Vessel Performance Sea Technology magazine," Sea Technology Magazine, Jul. 05, 2023. https://sea-technology.com/ai-vessel-performance-testing-yara-marine
- J. Ward, "FuelOpt™ propulsion automation system takes central role in AI-assisted voyages," Fathom World - Shipping and Maritime Industry News, Mar. 05, 2020. https://fathom.world/fuelopt-propulsion-automation-system-takes-central-role-in-ai-assisted-voyages/
- Y. Cao, M. Zhu, K. Tian, Y.-Q. Wen, J.-H. Zhang, and J.-N. Cao, An improved deep reinforcement learning Method-Based nonparametric modeling of ship dynamics. 2023. doi: 10.23919/ccc58697.2023.10240206.
- W.-Y. Wang, F. Ma, and J. Liu, Course tracking control for smart ships based on a deep deterministic policy gradient-based algorithm. 2019. doi: 10.1109/ictis.2019.8883840.
- P. Wu, J. S. Partridge, E. Anderlini, Y. Liu, and R. W. G. Bucknall, "An Intelligent Energy Management Framework for Hybrid-Electric Propulsion Systems Using Deep Reinforcement Learning," arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2108.00256.
- N. Zare, B. Brandoli, M. Sarvmaili, A. Soares, and S. Matwin, "Continuous Control with Deep Reinforcement Learning for Autonomous Vessels," arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2106.14130.
- N. Wang et al., "KUNPENG: An Embodied Large Model for Intelligent Maritime," arXiv (Cornell University), Jul. 2024, doi: 10.48550/arxiv.2407.09048.
- P. Wu, J. S. Partridge, E. Anderlini, Y. Liu, and R. W. G. Bucknall, "An Intelligent Energy Management Framework for Hybrid-Electric Propulsion Systems Using Deep Reinforcement Learning," arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2108.00256.

- Z. Zhang, Z. Yao, Q. Sun, and H. Qian, "Energy Optimization of Automatic Hybrid Sailboat," arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1811.11391.
- T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," arXiv (Cornell University), Jan. 2015, doi: 10.48550/arxiv.1509.02971.
- M. M. Eweda and K. ElNaggar, "Reinforcement learning for autonomous underwater vehicles (AUVs): navigating challenges in dynamic and energy-constrained environments," Robotics : Integration, Manufacturing and Control, vol. 1, no. 2, p. 31, Dec. 2024, doi: 10.21622/rimc.2024.01.2.1145.
- Guo, Y. & Zhang, Zhibin & Zheng, Weiqing & Yanhua, Du. (2020). An Autonomous Path Planning Model for Unmanned Ships Based on Deep Reinforcement Learning. Sensors.Doi 20. 426. 10.3390/s20020426.