ROB 590: Heron USV Reinforcement Learning

Tung Do
tungsdo@umich.edu

Katie Wakevainen
kwakev@umich.edu

***Abstract*—In this paper, we demonstrate the first application of the Lucid Dreamer model-based reinforcement learning method to USV path planning and control. The system combines model-based predictive control, offline and online learning, along with sensor fusion to achieve robust performance under varying environmental conditions prevalent in aquatic environments. To enable better onboard perception, we trained a U-Net image segmentation model. We also propose an obstacle avoidance algorithm which demonstrated success during online simulation testing. Additionally, we created a training environment in simulation for parallelized offline reinforcement learning which mimics our real-world testing environment. Data collection and real-world experimental tests were conducted in the Marine Hydrodynamics Lab.**

I. Introduction & Motivation

Autonomous navigation in maritime environments presents significant challenges for unmanned surface vehicles (USVs). These challenges include varying weather conditions, unpredictable currents, rough seas, and strong winds. Collecting realistic training data is difficult, and noisy sensor measurements exacerbate model uncertainty, particularly with dynamic and stochastic environments. These factors lead to reduced performance and reliability.

Real-time decision-making is essential for safe and effective operation, but this is constrained by the limitations of onboard hardware, including computational resources and power efficiency. USVs must also navigate complex, unstructured environments while managing localization and mapping in the presence of environmental stochasticity. Additionally, optimizing battery usage is critical to extend mission durations, further complicating control and planning strategies.

This project aims to overcome these obstacles, enabling the successful deployment and autonomous navigation of USVs in novel environments with uncertain wave conditions. By leveraging the Lucid Dreamer model-based reinforcement learning (MBRL) approach [1], we aim to develop an autonomous navigation framework that addresses these challenges effectively. The contributions of this work are summarized as follows:

1) Trained a U-Net segmentation model for applying the Lucid Dreamer model, enhancing the USV's perception capabilities and enabling improved trajectory planning and navigation in dynamic environments.
2) Designed and tested an object avoidance algorithm in Python and C++, demonstrating simulation success and contributing to optimizing the USV's control systems for benchmarking autonomous navigation.
3) Set up an offline Lucid Dreamer training environment for the USV in IsaacSim to enable more efficient initial network weights for online training.

II. Background & Related Work

A. Model-Based Reinforcement Learning

Model-based reinforcement learning (MBRL) [1] integrates an explicit environment model to predict future states and rewards, allowing for more efficient planning than model-free methods. By simulating outcomes with the learned model, MBRL reduces the need for extensive real-world interactions, improving data efficiency and safety. However, maintaining model accuracy is challenging, and errors can degrade performance. Advances in deep learning have made MBRL more scalable, and hybrid approaches combining model-based and model-free techniques are increasingly popular. MBRL is especially valuable in robotics and healthcare, where safety and efficiency are critical. In this project, the Lucid Dreamer MBRL method is used to achieve real-time reactive control of the Heron USV in challenging aquatic environments, leveraging its ability to balance efficiency and adaptability.

B. USV Control and Path Planning

Reinforcement learning for USV control and path planning is an active area of research. One approach uses a raster map combined with a Deep Q-Learning network (DQN) to enable full-coverage path planning for a USV around static obstacles in simulation [2]. In addition to static obstacles, dynamic obstacles, and neighboring USVs are sometimes considered for the obstacle-avoidant trajectories [3].

a) MBRL: Model-based reinforcement learning is a common technique used for USV path planning with collision avoidance [3]–[7]. There are a variety of MBRL methods for ASV control that have been explored in the literature. This includes approaches that use actor-critic networks [4], Lyapunov filtered probabilistic model predictive control (LFPMPC) [5], local update spectrum probabilistic model predictive control (LUSPMPC) [7], and filtered probabilistic model predictive control (FPMPC) [6]. There is not an established method that dominates the literature at this time as each technique has its own strengths and weaknesses.

For instance, some path planning approaches for USVs focus on distributed formation control, incorporating collision avoidance as part of the reward function [8]. This approach demonstrates stability in formations but has limited testing in real-world environments. It does not account for realistic challenges such as sensor noise, communication delays, and environmental disturbances.

In contrast, data-driven MBRL for USV trajectory tracking utilizes real-time model predictive control (MPC) adjustments for trajectory tracking. This method emphasizes data efficiency and strong path adherence. However, it struggles to fully capture complex, non-linear dynamics such as hydrodynamic forces and propulsion interactions, and attempting to incorporate these complexities could hinder real-time trajectory predictions.

Lyapunov-Guided Probabilistic MBRL (LFPMPC) integrates a Lyapunov stability metric to enhance robustness in USV control [5]. While this method excels in stability, it suffers from slower initial learning, particularly in high-dimensional systems or complex environments, making it less suitable for online training when rapid adaptation is required.

The Lucid Dreamer MBRL method emphasizes online learning and adaptive control, providing fast adaptation and low computational costs [1]. It excels in rapidly adapting to changing environmental conditions, ensuring reliable performance in dynamic and unpredictable scenarios. In experimental tests it also exhibited high reliability in trajectory tracking and position-keeping, even under varying environmental disturbances, showcasing its robustness and adaptability in challenging scenarios. Lucid Dreamer's cost-effective computational performance makes it particularly suitable for deployment on platforms with limited hardware resources, enhancing its practicality for real-world applications. We demonstrate its first application to USV path planning and control in this work.

1) Vehicle Dynamics: There are varied approaches to modeling the dynamics of the USVs to enable effective control. Classically, a mathematical model of the USV is derived from its kinematics and dynamics [4]. Gaussian processes are also commonly used to model USV dynamics without prior knowledge [5]–[7]. Other approaches, including our proposed approach, require no prior knowledge of the USV dynamics and rely on a deep neural network to learn the dynamics of USV due to the use of model-based learning [3], [8].

III. Technical Approach

This project employs the Lucid Dreamer model-based reinforcement learning (MBRL) approach to enable adaptive, real-time navigation for USVs in dynamic aquatic environments. The system combines model-based predictive control, offline and online learning, along with sensor fusion to achieve robust performance under varying environmental conditions.

Applying Lucid Dreamer introduces several key innovations to the existing MBRL approaches to autonomous USV navigation and control. First, it integrates online learning with adaptive control to enable real-time performance, allowing the system to react swiftly to environmental changes. Secondly, Lucid Dreamer is optimized for low computational overhead, making it well-suited for resource-constrained platforms commonly used in USVs. Additionally, it is designed to handle environmental stochasticity and sensor degradation, ensuring robust operation even under challenging conditions.

One limitation of the Lucid Dreamer method is that real-time updates can significantly strain computational resources in highly complex environments with unpredictable, high-dimensional disturbances. To address this challenge, this project mitigates the issue by running the computational processes offboard instead of relying on Heron's Nvidia computer, ensuring smoother operation without compromising performance.

In this project, the Lucid Dreamer approach extends prior MBRL methods by combining model-based predictions with real-time adaptability to environmental changes, such as sensor degradation caused by waves. This makes it particularly well-suited for challenging aquatic environments where accurate dynamics modeling and rapid decision-making are critical.

A. Image Segmentation

Image segmentation is pivotal in this project, enabling the Heron USV to identify obstacles and navigate effectively in complex aquatic environments. Using the U-Net architecture from the open-source repository `https://github.com/spsingh37/UNet_dreamer` [9], a model was trained on a custom dataset specifically collected and labeled for Heron's operating conditions, using RoboFlow. This process involved several design stages, data collection, and troubleshooting to ensure robust performance.

B. Offline Training

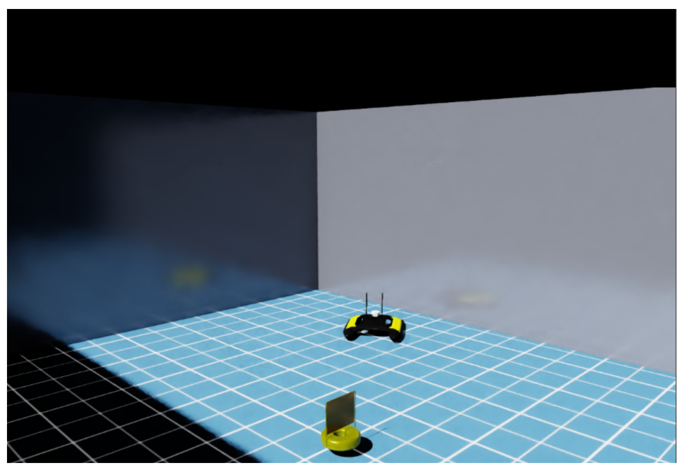

Fig. 1. Isaacsim Training Environment for the Heron USV

To enable highly parallelized offline reinforcement learning for a USV, we utilized the RANS framework which integrates with the Nvidia IsaacSim simulator [10]. The original intention was to use the simulator Gazebo for the offline policy training, but the existing libraries which integrated with Gymnasium were largely unmaintained or otherwise didn't meet our version requirements.

We chose to use RANS for a multitude of reasons. First, RANS has implemented scripts to simulate the hydrodynamics of an USV which IsaacSim doesn't offer a native package for at this time. Second, RANS had existing support for training with the Heron USV which prior work has demonstrated can transfer to real-world field tests [11]. Third, RANS is highly-parallelized compared to existing libraries for reinforcement learning with Gazebo. IsaacSim also supports higher quality visual rendering than Gazebo which is an appealing feature for future work which could utilize a camera as part of the observation space.

We created a training environment with a dynamic number of buoys to resemble the field tests as seen in Figure 1. Training was run on an Ubuntu 20.04 server with two Nvidia RTX A6000 GPUs.

IV. Experiments and Results

In this work, we utilized the Heron USV, as seen in Figure 2, to demonstrate the efficacy of Lucid Dreamer for USV path planning and control. The offline training used to initialize the network weights for the online training utilized IsaacSim and the RANS framework. In turn, our online training in simulation utilized Gazebo and ROS.

Fig. 2. The Heron USV used in this project.

A. USV Hardware Platform

The Heron USV, developed by Clearpath Robotics, is a versatile, portable unmanned surface vehicle (USV) designed for research and monitoring in aquatic environments [12]. It features a compact, catamaran-style design with deployed dimensions of 1300 mm in length, 940 mm in width, and 340 mm in height (51.2 x 37 x 13.4 inches). Its anti-fouling thrusters and 150 mm (5.9 inches) draft make it well-suited for many conditions, including shallow and hard-to-reach waters.

The Heron can carry a maximum payload of 10 kg (22 lbs) and is equipped with GPS, an Inertial Measurement Unit (IMU), and multiple communication options such as USB, TCP/IP, and RS232. These features make it ideal for environmental monitoring, data collection, and autonomous navigation applications. It is powered by a 14.4V 29Ah NiMH battery pack, allowing up to 2.5 hours of typical operation, or 10 hours in standby mode, with a recharge time of 10 hours.

Weighing 20 kg (44 lbs) without the battery and with a battery weight of 9 kg (20 lbs), the Heron is lightweight and portable, with folding pontoons that facilitate easy transport and deployment. It reaches speeds of 1.7 m/s (5.6 ft/s) and can precisely maneuver using its bi-directional jet thrusters, which allow it to turn on the spot [12].

In addition to its flexibility for research and environmental missions, the Heron supports open-source development, enabling researchers to customize it for specific tasks such as autonomous navigation.

B. Image Segmentation

The effort began with hardware preparation to equip the Heron USV to capture high-quality data for image segmentation. Custom mounts for the beacon and Intel RealSense camera were designed and 3D-printed, allowing secure attachment to the vehicle. These mounts enabled precise positioning for the RGB-D camera, which streams data through ROS. The simulation environment was updated to reflect these changes, accurately representing the Heron's setup.

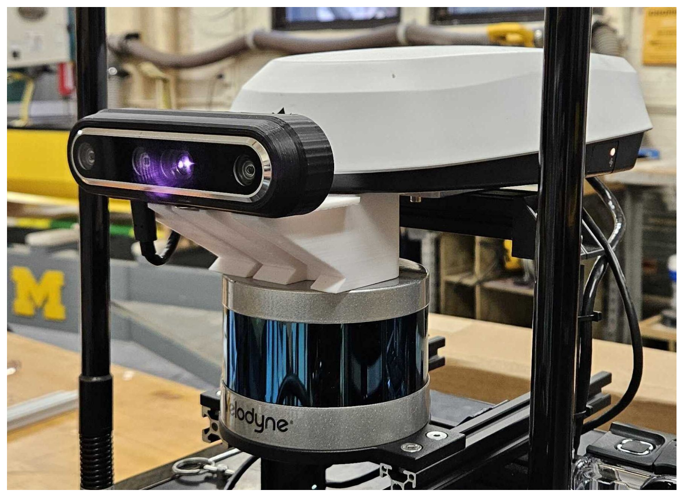

Fig. 3. Camera mounted on the Heron

Data collection occurred during field tests at the Marine Hydrodynamics Laboratory (MHL) on September 27th (pre-test) and September 30th, where the Heron was tested under three wave conditions. The captured data included multi-modal inputs from LiDAR, RGB-D camera, IMU, beacon, and control signals, providing a comprehensive segmentation and model training dataset. During this phase, the camera's frame was temporarily fixed upright to ensure stability and data quality, addressing immediate testing requirements.

The training began with initial trials on the MaSTr3125 dataset, which proved unsuitable due to differences in environmental context. Attention then shifted to the custom-labeled dataset collected from the Heron's environment. After labeling 120 of the 2,230 images, the process paused to validate the U-Net architecture by overfitting it on a small subset of 5–10 images. This step revealed a mismatch issue in the prediction output, which was resolved. Once the architecture was confirmed to work as intended, a sufficient dataset was used for training. The final model achieved a training loss of 0.0007973, demonstrating strong segmentation performance.

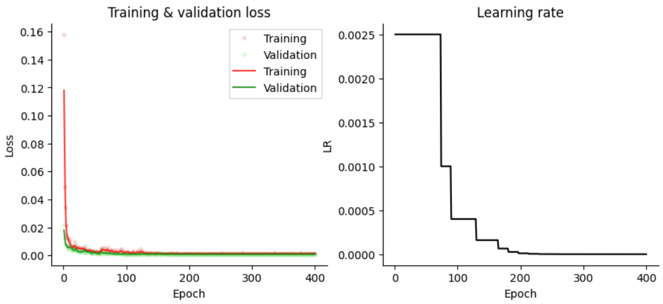

Fig. 4. U-Net Final Training Results

The U-Net training marked a significant milestone, but several ongoing efforts are focused on further improving the system. To enhance temporal resolution for more precise segmentation, the camera is being configured to capture at 30 FPS. The camera's URDF configuration is being refined in the simulation environment to ensure more accurate alignment with real-world conditions. Additionally, sensor locations and buoy dimensions are being updated to improve the accuracy of the simulation and support better integration with navigation algorithms.

With the U-Net model now operational, its segmentation outputs are ready to be integrated into the Lucid Dreamer MBRL framework. This integration is expected to significantly enhance obstacle detection, trajectory planning, and overall navigation, further advancing the Heron USV's capability to operate autonomously in dynamic and unpredictable aquatic environments.

V. Discussion

The integration of the Lucid Dreamer MBRL method with the Heron USV demonstrates promising results in enabling adaptive, efficient, and robust navigation in challenging aquatic environments. The method effectively balances computational efficiency with real-time adaptability, addressing key limitations of prior MBRL approaches.

Despite these successes, several difficulties were encountered during the project. The Heron USV occasionally experienced breakdowns, leading to delays in testing and data collection. Additionally, the vehicle's dimensions (1.3 m x 0.94 m, 29 kg) presented challenges in transport and handling, particularly when deploying and recovering the vehicle at testing sites.

Testing was constrained to the Marine Hydrodynamics Laboratory (MHL), which had a tight schedule during weekdays (8 AM–4 PM) and was unavailable on weekends. Each wave test required significant setup time (1–1.5 hours), further limiting the number of experiments conducted within the allocated time. These logistical constraints added complexity to the project timeline and necessitated efficient coordination to maximize the limited testing opportunities.

Despite these challenges, the project succeeded in demonstrating the potential of the Lucid Dreamer approach. Future iterations of this work could benefit from additional testing resources, extended access to the MHL, or alternative testing environments to reduce delays and improve overall experimental efficiency.

A. *U-Net Image Segmentation*

The U-Net segmentation model was trained on 222 training images and validated on 13 validation images. The results show strong convergence, with the training and validation loss stabilizing after approximately 100 epochs, as depicted in the loss curve in Figure 4. The final training loss is very low, suggesting that the model successfully learned the patterns within the dataset. The validation loss also closely tracks the training loss, indicating minimal overfitting and good generalization to unseen validation data.

The learning rate schedule in Figure 4 shows an effective decay strategy, where the learning rate decreases at key intervals, promoting stable training and refinement of the model. This schedule prevents large oscillations in loss and ensures smooth convergence. The final performance demonstrates that the U-Net architecture is suitable for segmenting objects in the Heron USV's environment, such as buoys, despite the relatively small dataset size.

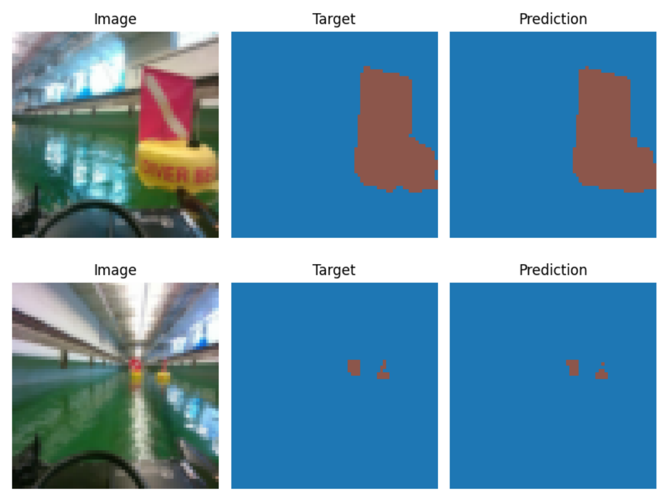

Fig. 5. Sample of U-Net Test Results

In the testing results seen in Figure 5, the predictions closely match the ground truth target labels, effectively segmenting the buoys in both calm and moderately cluttered aquatic environments. In the first example, the buoy is clearly identified, with minimal segmentation errors. In the second example, the model successfully detects smaller buoys at a greater distance, which showcases its ability to capture features at varying scales. However, some minor inaccuracies are observed, particularly with small or distant objects, where the segmentation boundaries are slightly less precise. These discrepancies could be due to the limited training data or the inherent difficulty of identifying small objects in low-resolution images.

Overall, the results highlight the effectiveness of the U-Net architecture for the segmentation task but also suggest areas for improvement. The model could benefit from additional data to enhance its ability to generalize to diverse conditions, especially for smaller or partially obscured objects. Furthermore, the resolution of input images may need to be increased to improve the segmentation quality of fine details, particularly for objects at a distance.

To improve the U-Net model's performance, data augmentation techniques such as random rotations, brightness adjustments, and cropping could be employed to simulate diverse environmental conditions and effectively increase the dataset size. Expanding the dataset with additional images capturing challenging scenarios, such as high wave disturbances, occlusions, and varied lighting, would enhance the model's generalization. Higher-resolution inputs during training and inference could improve segmentation accuracy for small and distant objects. Finally, fine-tuning hyperparameters and exploring alternative loss functions, such as Dice or IoU loss, could further optimize the model's ability to detect challenging objects.

VI. Future Work

Future efforts will focus on enhancing the system's robustness and performance under more challenging conditions. This includes improving the model to handle higher wave disturbances effectively and configuring the camera to capture at a frequency of 30 FPS for better data fidelity. Additional work will address fixing the camera's URDF configuration in the code for future needs and updating the dimensions and measurements of sensor locations to ensure greater accuracy. Extracting key data streams such as `/scan`, `/odometry/filtered`, `/imu/data`, `/cmd_vel`, `/cmd_drive`, `/tf`, and `/tf_static` will enable running Hector SLAM for detailed comparison and performance evaluation. These improvements aim to refine the system's reliability, adaptability, and precision, especially in more extreme aquatic environments.

For the offline training, there will be efforts to make the testing environment more visually realistic to real-world conditions. This would help with integrating the camera into the observation space which may improve performance. We anticipate that the IsaacSim simulator will continue to improve its offerings of provided USDs for creating environments along with accurately simulating the hydrodynamics of the USV.

Future efforts will focus on addressing these improvements to increase system robustness and adaptability, particularly in challenging aquatic environments with more extreme conditions. Additionally, further development will explore enhanced sensor fusion techniques and broader testing scenarios to validate the system's scalability and generalizability.

VII. Conclusion

This project aimed to develop an adaptive, real-time control system for autonomous maritime navigation on the Heron USV. Using Lucid Dreamer, the project addresses key challenges such as variable environmental conditions, sensor noise, and model uncertainty. Through innovations in adaptive model predictive control, real-time online learning, and sensor fusion, the system enables reliable USV navigation under unpredictable conditions while minimizing computational load. Real-world testing with wave data supports the system's stability, demonstrating effective path-following and obstacle avoidance in both calm and minor wave disturbances.

VIII. Contributions

A. Tung's Contributions

- Developed a URDF model for the Heron USV in ROS2 and Gazebo Garden to create a realistic simulation environment for testing navigation algorithms and collecting more training data.
- Designed and tested an object avoidance algorithm in Python and C++, demonstrating simulation success and contributing to optimizing the USV's control systems for benchmarking autonomous navigation.
- Prepared the Heron USV for deployment by configuring battery, electrical, and mechanical systems, maintaining its functionality for every test. Designed and 3D-printed custom sensor mounts to integrate and calibrate sensors, ensuring synchronized data acquisition.
- Conducted real-world navigation tests under varying wave and obstacle conditions to collect sensor data (LiDAR, camera, IMU, GPS, odometry) for multimodal data fusion.
- Implemented frame alignment and timestamp synchronization to improve Hector SLAM accuracy and performance analysis in RViz.
- Labeled and processed a 2,230-image dataset to train a U-Net segmentation model for applying the Lucid Dreamer model, enhancing the USV's perception capabilities and enabling improved trajectory planning and navigation in dynamic environments.
- Designed buoy and MHL model for IsaacSim simulations.

B. Katie's Contributions

- Compiled existing libraries for integrating ROS, Gazebo, and Gymnasium and their shortcomings given the project objectives.
- Documented set up processes for experimental testing with Gazebo and IsaacSim simulations.
- Created USD assets from the CAD of the entities previously used in Gazebo.
- Created USD environment to run experimental tests with the Heron USV in IsaacSim.
- Added a task for ASV gap following between dynamically generated buoys to the RANS framework.

REFERENCES

[1] E. Shrestha, H. Wan, C. Reddy, Y. Zhuang, and R. Vasudevan, "Multimodal model-based reinforcement learning for autonomous racing," in *Deployable RL: From Research to Practice @ Reinforcement Learning Conference 2024*, 2024. [Online]. Available: https://openreview.net/forum?id=wXjsMTZzxF

[2] B. Xing, X. Wang, L. Yang, Z. Liu, and Q. Wu, "An Algorithm of Complete Coverage Path Planning for Unmanned Surface Vehicle Based on Reinforcement Learning," *Journal of Marine Science and Engineering*, vol. 11, no. 3, p. 645, Mar. 2023, number: 3 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2077-1312/11/3/645

[3] C. Pan, Z. Peng, L. Liu, and D. Wang, "Data-driven distributed formation control of under-actuated unmanned surface vehicles with collision avoidance via model-based deep reinforcement learning," *Ocean Engineering*, vol. 267, p. 113166, Jan. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801822024490

[4] W. Yuan and X. Rui, "Deep reinforcement learning-based controller for dynamic positioning of an unmanned surface vehicle," *Computers and Electrical Engineering*, vol. 110, p. 108858, Sep. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790623002823

[5] L. Xia, C. Shao, H. Li, and Y. Cui, "Robust Model-based Reinforcement Learning USV System Guided by Lyapunov Neural Networks," in *2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2022, pp. 1976–1981. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10011834

[6] Y. Cui, L. Peng, and H. Li, "Filtered probabilistic model predictive control-based reinforcement learning for unmanned surface vehicles," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 6950–6961, 2022.

[7] Y. Cui, W. Shi, H. Yang, C. Shao, L. Peng, and H. Li, "Probabilistic model-based reinforcement learning unmanned surface vehicles using local update sparse spectrum approximation," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 2, pp. 1283–1293, 2024.

[8] Z. Peng, E. Liu, C. Pan, H. Wang, D. Wang, and L. Liu, "Model-based deep reinforcement learning for data-driven motion control of an under-actuated unmanned surface vehicle: Path following and trajectory tracking," *Journal of the Franklin Institute*, vol. 360, no. 6, pp. 4399–4426, Apr. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016003222007463

[9] S. P. Singh, "Unet_dreamer: U-net implementation for image segmentation," https://github.com/spsingh37/UNet_dreamer, 2023, accessed: 2024-12-06.

[10] M. El-Hariry, A. Richard, and M. Olivares-Mendez, "Rans: Highly-parallelised simulator for reinforcement learning based autonomous navigating spacecrafts," *arXiv preprint arXiv:2310.07393*, 2023.

[11] L. F. W. Batista, J. Ro, A. Richard, P. Schroepfer, S. Hutchinson, and C. Pradalier, "A deep reinforcement learning framework and methodology for reducing the sim-to-real gap in asv navigation," 2024. [Online]. Available: https://arxiv.org/abs/2407.08263

[12] C. Robotics, *Heron USV User Manual*, https://www.dndkm.org/DOEKMDocuments/GetMedia/Technology/3121-heron_usermanual.pdf, Clearpath Robotics, 2024, version 1.0. [Online]. Available: https://www.dndkm.org/DOEKMDocuments/GetMedia/Technology/3121-heron_usermanual.pdf

[13] M. El-Hariry, A. Richard, V. Muralidharan, B. C. Yalcin, M. Geist, and M. Olivares-Mendez, "Drift: Deep reinforcement learning for intelligent floating platforms trajectories," *arXiv preprint arXiv:2310.04266*, 2023.