Introduction of Reinforcement Learning into Automatic Stacking of Wave-dissipating Blocks

Hao Min Chuah
Graduate School of Science and Technology
Niigata University
Niigata, Japan
e-mail: haominchuah01@gmail.com

Tatsuya Yamazaki
Faculty of Engineering
Niigata University
Niigata, Japan
e-mail: yamazaki.tatsuya@ie.niigata-u.ac.jp

***Abstract*—Accurate and strategic placement of wave-dissipating blocks is essential for effective coastal protection structures. Current supervised learning-based approaches have achieved precise single-block placement. However, they inherently suffer from significant limitations, such as a lack of adaptability to environmental and structural changes, an inability to optimize sequences of multiple-blocks, and a heavy reliance on extensive pre-generated labeled data. This paper identifies the key limitations inherent in supervised Convolutional Neural Network methods and proposes a novel reinforcement learning (RL)-based approach to address these issues. By illustrating how RL naturally provides adaptability, strategic multi-block placement, and reduced reliance on labeled data, this early-stage idea is expected to contribute to the integration of simulation methodologies and machine learning approaches.**

Keywords-wave-dissipating blocks; reinforcement learning; simulation; automatic stacking.

I. INTRODUCTION

Wave-dissipating blocks play a pivotal role in coastal engineering, protecting infrastructure by effectively dissipating wave energy. The optimal placement of these blocks significantly influences the overall stability, compactness, and performance of breakwater structures. However, the installation of wave-dissipating blocks still heavily depends on the empirical knowledge and experience of skilled workers. To overcome the limitations, Xu [1] achieved accurate single-block placements using supervised Convolutional Neural Network (CNN) methods. Albeit, his methods may suffer from inflexibility in adapting to structural changes and an inability to perform long-term optimization. In this paper, we explore an automatic stacking method for wave-dissipating blocks using Reinforcement Learning (RL) in our self-developed 3D-BW (3-Dimensional BreakWater Simulator) [2]. This RL-based method offers enhanced flexibility and adaptability, enabling the learning agent to optimize long-term structural integrity and dynamically adapt to changes in block types, structure geometry, and target goals.

In Section 2, we review related works, particularly focusing on supervised learning-based approaches for block placement and their limitations. In Section 3, we present our proposed methodology based on reinforcement learning, detailing the motivation, agent design, and inherent challenges. Finally, Section 4 concludes the paper by summarizing key contributions and outlining future directions for integrating reinforcement learning into coastal block placement simulations.

II. RELATED WORKS

Accurate placement of wave-dissipating blocks has been studied using several computational approaches, with supervised learning being one of the most explored methods. Xu [1] achieved accurate single-block placements using a supervised Convolutional Neural Network (CNN) trained on labeled data generated from a physics-based simulator.

A. Xu's Supervised CNN Approach

The process consisted of three phases.

1) Data Generation & Pose Labeling: A sliding window extracted 512 × 512 depth patches from the structure's surface. For each patch, 1000 simulated drops were tested at varying poses. Two criteria were evaluated:

- Compactness: measurement by comparing the placed block's position against the target height map, calculating insufficient volume (gap filling).
- Stability: horizontal displacement determination after settling (displacement ≤ 0.2m).

Then, the best performing pose becomes the ground-truth label for CNN training.

2) CNN Training: The network learned to predict optimal translation and rotation from depth patches, minimizing supervised loss between predictions and labels.

3) Real-Time Inference: The CNN predicted poses for each patch and placed blocks iteratively until the structure was completed.

This approach achieved high local accuracy and fast inference, making it suitable for controlled, static construction conditions. However, it has limitations:

- Dataset Dependence: Requires extensive pre-generated labeled data for each structural configuration.
- Lack of Adaptability: Cannot generalize new block types or changing conditions without retraining.
- Greedy Placement Strategy: Optimizes only immediate placement, ignoring long-term structural goals.

B. Other Relevant Approaches

Beyond CNN-based placement methods, other computational strategies have been explored for similar optimization problems. Physics-based heuristic approaches use deterministic rules to maximize local compactness and stability. However, their performance tends to degrade in dynamic or unpredictable environments. In the field of robotics, reinforcement learning has been successfully applied to adaptive planning tasks, such as object manipulation, grasping, and stacking under uncertainty [5][6]. Hybrid methods that combine CNN-based perception for accurate pose estimation with RL-based decision-making have also been proposed [7], enabling both precision and adaptability, although these approaches face scalability challenges when applied to large, irregular structures.

C. Summary of Achievements and Research Gap

Supervised CNN methods, such as that of Xu [1], have demonstrated high placement accuracy for static, controlled conditions, but their lack of adaptability and inability to perform strategic optimization over multiple steps remain significant drawbacks. In contrast, RL-based methods learn directly from interactions with the environment, removing the dependence on fixed datasets, and can evaluate the consequences of each placement in the context of a long-term construction sequence. They are also inherently more flexible, adapting to changes in block geometry, environmental constraints, and overall structural goals without the need for complete retraining. Nevertheless, Xu's dataset suffers from limited diversity, being tailored to specific block types and structural configurations. Incorporating data augmentation techniques, such as introducing synthetic noise, randomizing textures, and perturbing poses could improve the robustness of both supervised and RL-based methods, and in the RL case, could be integrated into pretraining phases, such as behavioral cloning to accelerate convergence.

III. PROPOSED METHODOLOGY: REINFORCEMENT LEARNING (RL)-BASED BLOCK PLACEMENT

To address the limitations identified in [1], we introduce a RL-based approach using Unity ML-Agents [3] and Proximal Policy Optimization (PPO). The method leverages the interaction-based learning paradigm inherent to RL to dynamically adapt and optimize the strategic placement of wave-dissipating blocks.

A. Motivation of Utilizing RL

RL allows a learning agent to iteratively obtain an optimal policy by interacting directly with its environment, receiving feedback through reward signals, and adapting actions accordingly. Unlike supervised methods that depend on extensive pre-labeled data, RL's ability to continuously refine its strategies based on outcomes makes it uniquely suited to scenarios that involve complex and dynamic decision-making, such as block stacking in coastal engineering. The primary reasons for employing RL in this research include:

- Adaptability: the RL agent dynamically adapts to structural or block-type changes.
- Strategic Long-Term Optimization: RL considers the implications of each block placement in a multi-block scenario, addressing global objectives, such as porosity reduction and structural stability.
- Reduced Data Dependency: the agent learns from interaction outcomes rather than extensive simulations, reducing the need for dataset preparation.

B. RL Agent Design

The RL agent operates within our self-developed 3D-BW environment, performing iterative block placements by observing the current structural state using data representations, such as gap maps and depth maps derived from discretized grid cells. Figure 1 illustrates a bird's-eye view visualization of a gap map. Based on these observations, the agent selects a discrete placement coordinate (x, z) and a rotation angle, then drops a block from a predetermined height.

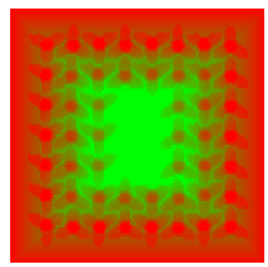

Figure 1. Visualization of gap map of breakwater structure from a bird-eye view

The agent aims to optimize multiple explicit and adaptable objectives, including compactness, stability, overflow penalty, and porosity. After multiple block placements, the Proximal Policy Optimization algorithm [4] updates the policy parameters based on the observed rewards and outcomes. Through iterative learning, the agent is expected to gradually improve its strategic placement capabilities.

C. Limitations

While RL offers significant potential advantages, several challenges must be acknowledged:

- Increased Training Complexity: potentially requires substantial computational resources and training time.
- Reward Function Sensitivity: strong dependency on effective reward design, potentially challenging to tune accurately.
- Exploration Efficiency: initial random placements may cause slow training convergence, necessitating strategies like curriculum learning or behavioral cloning to mitigate this issue.

IV. CONCLUSION

This idea contribution proposes an RL approach as an innovative, adaptive, and strategic method for optimizing the placement of wave-dissipating blocks. By leveraging Unity ML-agents, physics-based simulations, and well-designed reward functions, RL demonstrates significant potential to overcome the inherent limitations of supervised learning methods. Although challenges remain regarding computational resources and careful reward design, these issues may be mitigated by incorporating techniques, such as behavioral cloning. This approach lays the groundwork for more autonomous and efficient block placement strategies in future coastal engineering applications.

ACKNOWLEDGMENT

We extend our sincere gratitude to Honma Concrete Industry Co., Ltd. for generously providing the three-dimensional data essential to our research. We also wish to express our appreciation to Onogumi Co., Ltd. for their valuable guidance and insightful advice throughout the course of this research.

REFERENCES

[1] Y. Xu, A study on stacked object recognition and stacking operation planning combining 3D point cloud representation, deep-learning and physics engine, M.Eng. thesis, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan, 2020, doi:10.14943/doctoral.k15552.

[2] H. M. Chuah, T. Yamazaki, R. Iwasawa, and T. Suto, "Development of wave-dissipating block installation simulation for inexperienced worker training," World Academy of Science, Engineering and Technology, International Journal of Computer and Systems Engineering, vol. 18, no. 9, pp. 611-618, 2024.

[3] Unity Technologies, "Unity ML-Agents Toolkit," GitHub repository, [Online]. Available: https://github.com/Unity-Technologies/ml-agents. [retrieved: July, 2025].

[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[5] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3389-3396, 2017.

[6] D. Silver et al., "Deterministic policy gradient algorithms," in *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pp. 387-395, 2014.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.