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

Hao Min Chuah[1], Tatsuya Yamazaki[2]
[1]Graduate School of Science and Technology, Niigata University, Niigata, Japan
[2]Faculty of Engineering, Niigata University, Niigata, Japan

Contact email: haominchuah01@gmail.com

Chuah Hao Min

Hao Min Chuah received the bachelor's degree in engineering from the Faculty of Engineering, Niigata University, Japan, in 2024. He is currently pursuing the master's degree at the Graduate School of Science and Technology, Niigata University.

He is a student member of the Japan Society for Simulation Technology, the Institute of Electronics, Information and Communication Engineers (IEICE), and the Japan Society of Civil Engineers (JSCE).

His research interests include simulation of wave-dissipating block installation and applications of reinforcement learning in coastal engineering.

Background

Wave-dissipating block

- Crucial for dissipating wave energy and protecting coastal area, are commonly installed in harbors and seawalls
- Placement determines stability and performance

Mass (t)	Weight (kN)	Volume (m^3)	Surface Area (m^2)
3.991	39.138	1.735	9.796

Fig. 1 Shake Block 4-ton variant and specifications [1]

Breakwater structure

- Installation target for blocks

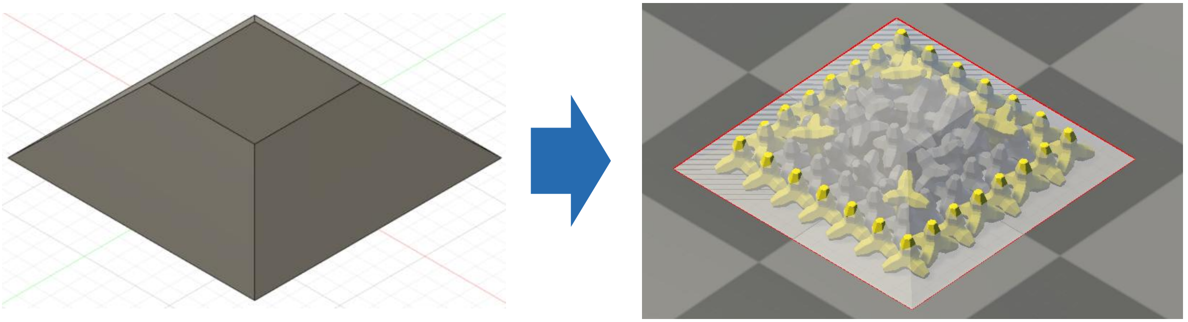


Fig. 2 Offshore breakwater model

[1] Honma Concrete Industry Co., Ltd, “Shake Block”, [Online] Available: https://www.honmacon.co.jp/02syouhin_list/01sye-kuburokku/01setumei.html, p. 3

Background

Current situation of installation work for wave-dissipating blocks

- Relies on the empirical knowledge and experience
- Trial and error on-site installation attempts
- Time-consuming and costly
- Hard to evaluate block placement

Placement of mat　　Placement of blocks　　Finished construction of breakwater

Fig. 3 Construction of breakwater [2]

[2] wakamatsukowan, "Manufacture and installation of wave-dissipating blocks.," [Online] Available: https://www.wakamatsu-kk.com/blank-3

Motivation

- Manual
- Inconsistent
- Labor-intensive
- Expensive

- Automation
- Reliable
- Scalable

- Objective solution
- Adaptive
- Long term

Artificial Intelligence (AI) has become relevant nowadays to tackle this problem

Related Works

Xu's Supervised CNN Approach [3]

- CNN trained on simulation-labelled poses

Pros

- Accurate
- Fast inference

Cons

- Dataset dependence
- No adaptability
- Short-term greedy strategy

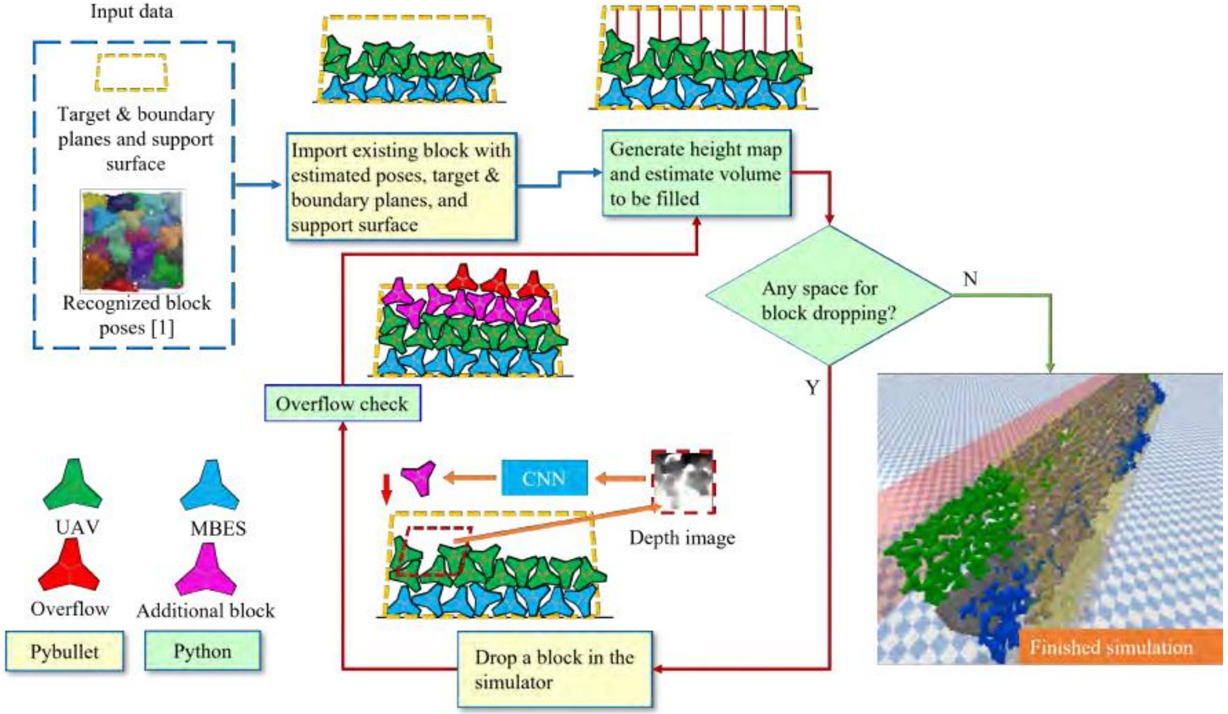


Fig. 4 Pipeline of block stacking operation

Physics-based Heuristic Approach

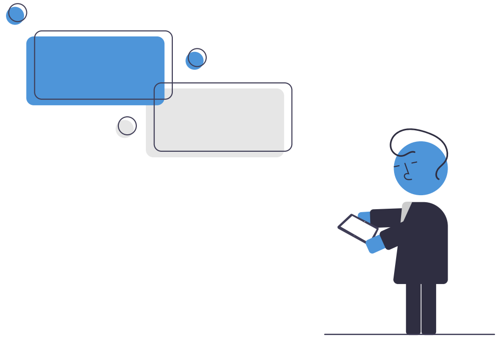
- Deterministic = No adaptability
- Break under uncertainty

[3] 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.

Research Gap

CURRENT

- Supervised CNN: Accurate but inflexible
- Heuristic: Fast but brittle



MISSING

- Adapt to any conditions
- Planning strategically over multiple placements

SOLUTION

- Reinforcement Learning (RL): Flexible
- Commonly used in robotics [4]

[4] 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.

Why Reinforcement Learning?

Adaptability

- Learn policies
- Generalize across different blocks and structure types

Strategic Optimization

- Multiple block placements
- Optimize long terms rewards

Reduced Data Dependency

- Learn from interaction
- Reduce reliance on dataset generation

3D-BW (3D Breakwater Simulator) [5]

- Developed based on Unity, a three-dimensional game development software since 2023
- Provides platform for off-site training
- Incorporates with ML-Agent, an open-source project that enables games and simulations to serve as environments for training intelligent agents

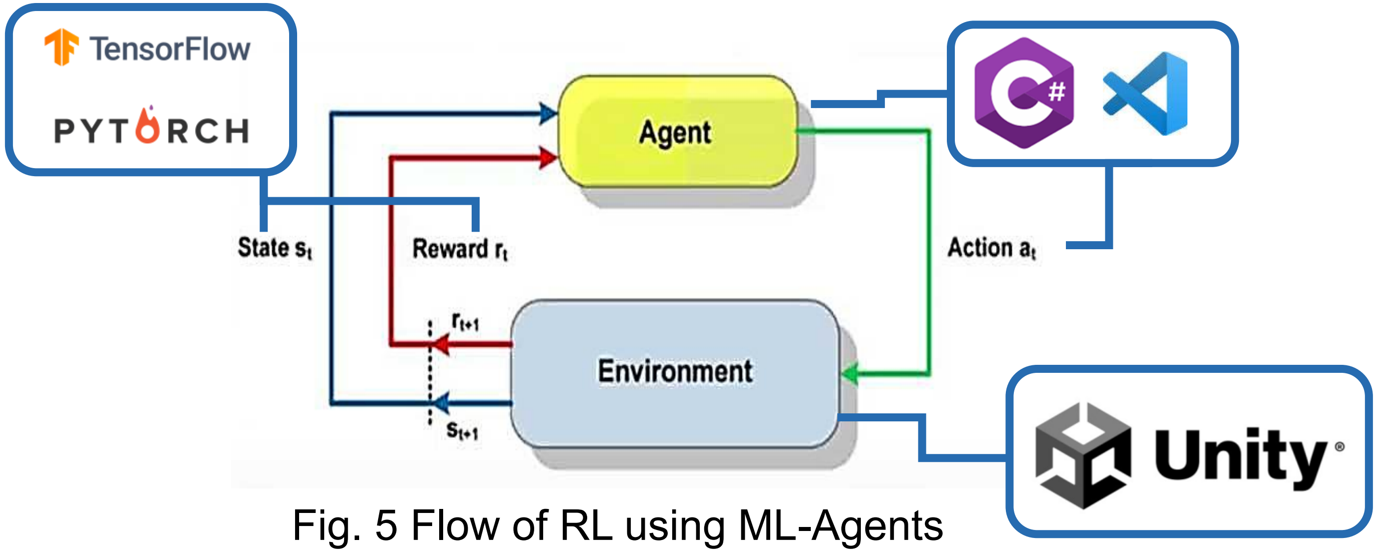


Fig. 5 Flow of RL using ML-Agents

Fig. 6 Interface of 3D-BW

[5] Chuah, H. M., Yamazaki, T., Iwasawa, R., Suto, T.: Development of Wave-Dissipating Block Installation for Inexperienced Worker Training. World Academy of Science, Engineering and Technology, International Journal of Computer and Systems Engineering vol. 18, no. 03 (2024)

Proximal Policy Optimization (PPO)

- Algorithm that updates the policy of the agent
- Policy: Strategy the agent uses to decide what action to take given its current observation or state

- The agent interacts with the environment
  - Observation: What it sees
  - Action: What it does
  - Reward: Feedback
- Advantages
  - Improves action (Learning from reward)
  - Avoids changing the policy too drastically (Stability)

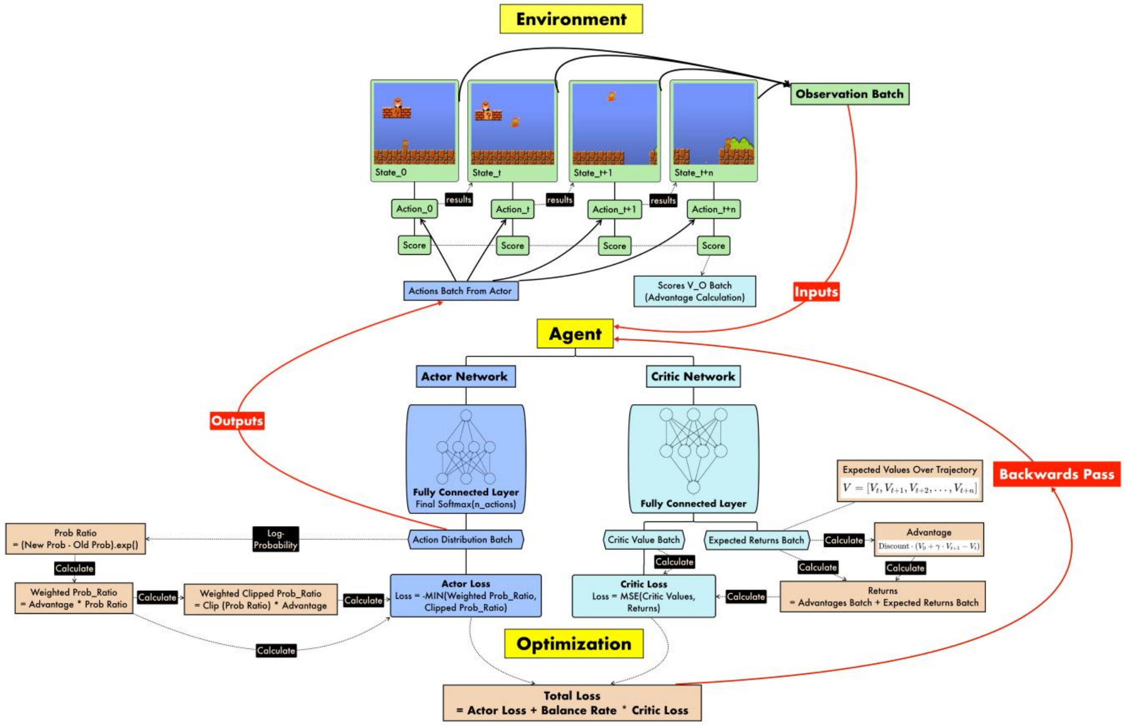


Fig. 7 PPO Flowchart [6]

Observation

$$obs_{x,z} = \frac{roofHeight_{x,z} - blockHeight_{x,z}}{\max_{i,j} roofHeight_{i,j} - blockHeight_{i,j}}$$

- Convert the breakwater into a 2D grid map
- Normalize the differences between block vertices and structure vertices in each grid cell
- Transform it into a gap map

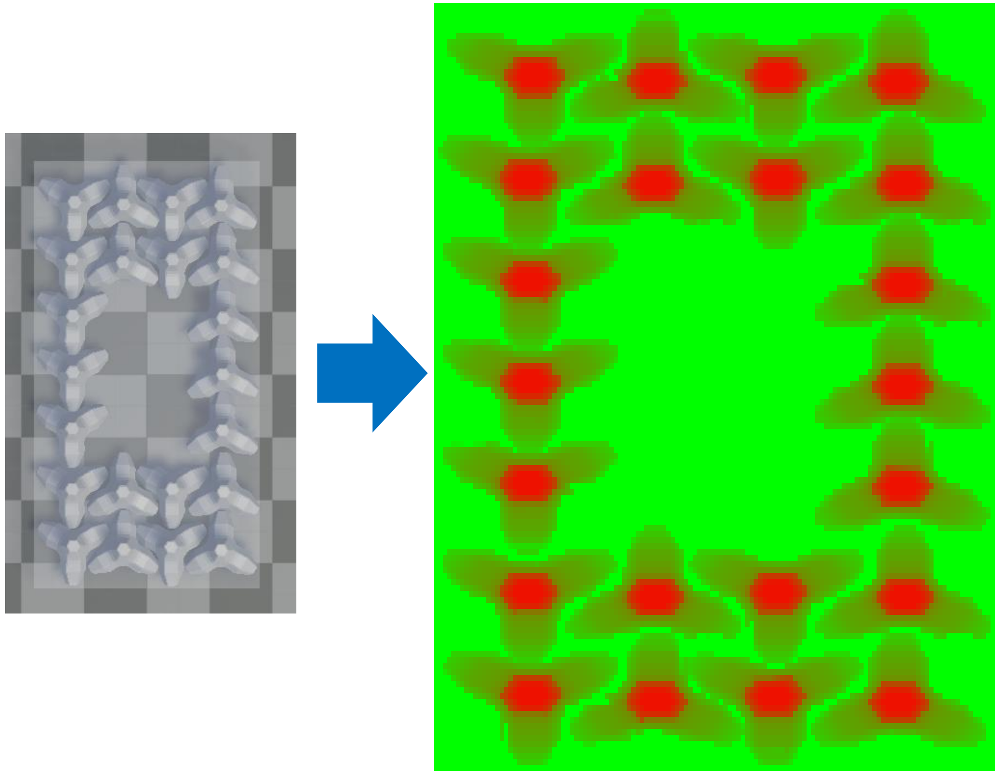
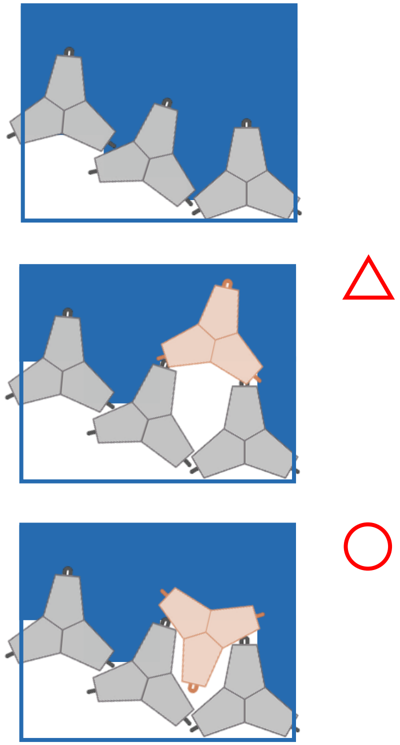


Fig. 7 Example of a grid map of 128x128

Action

- Drop from a fixed height
- Selects a discrete placement coordinate (x, z) from the grid map
- Randomized rotation angle

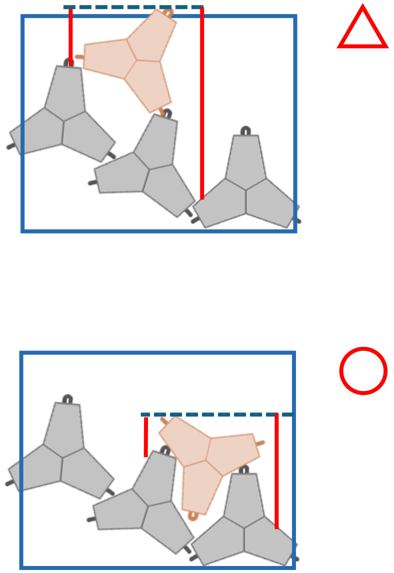
Proposed Reward Design



: Empty Spaces

Compactness

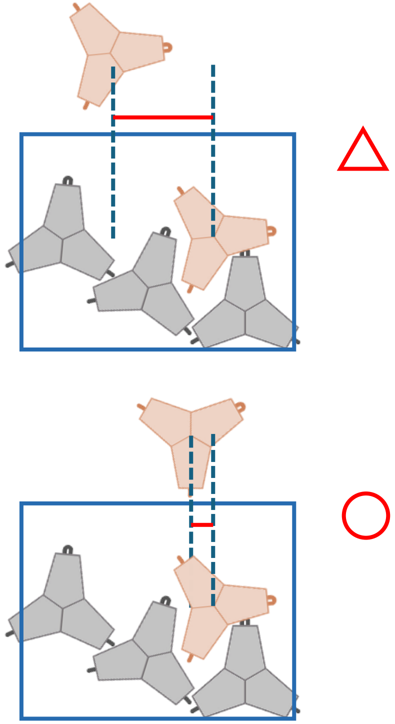
Difference between volume before and after within breakwater



: Height

Slope

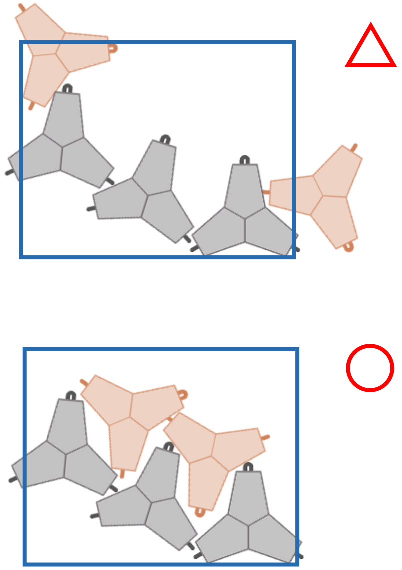
Height difference between neighbouring cells



: Distance moved

Stability

Distance moved after dropping block



Overflow

Overflowing from the top/side of breakwater

Challenges

Computational Cost

- Training takes time and resources

Reward Sensitivity

- Too strong/weak signals lead to poor polices

Slow Exploration

- May waste many episodes on random moves

<u>Possible solutions</u>

1. **Curriculum learning:** Gradually increase difficulty
2. **Behavioural Cloning:** Imitating expert demonstrations

Conclusion

- Propose an RL approach for automating and optimizing the placement of wave-dissipating blocks
- RL = Adaptability + Strategic optimization
- Careful reward design is compulsory

Future Work

- Simple experiment to test reward design
- Refine reward design
- Multiple test cases to test adaptability