Graduate School of Fundamental Science and Engineering
Waseda University

博　士　論　文　概　要
Doctoral Dissertation Synopsis

論　文　題　目
Dissertation Title

Deep Reinforcement Learning Adapted to Real-World Training Data Limitations

実環境データの制限に対応した深層強化学習

申　請　者
(Applicant Name)
André Yuji YASUTOMI
保富　アンドレ 裕二

Department of Intermedia Studies, Research on Intelligence Dynamics and Representation Systems

May, 2023

Deep reinforcement learning (DRL) is a rapidly growing area of research in the field of artificial intelligence, and it has shown exceptional success in solving complex tasks such as robotic manipulation. DRL involves training an agent to learn how to perform a task through trial and error by receiving feedback in the form of rewards or penalties. Despite its promising results, DRL faces significant challenges in its application to real-world scenarios. One of the primary challenges is the limited access to real-world data, as collecting real-world data is time-consuming, arduous, and potentially dangerous.

Given this limitation, the objective of this thesis is to propose methods for adapting DRL for real-world data limitations. The challenge with adapting DRL to real-world data limitation can be further divided into three challenges: (1) transferability to the real world, (2) sample efficiency of the DRL policy, and (3) generalization capability of the DRL policy.

The transferability challenge refers to the fact that the limited amount of data available restricts the capability of DRL policies to reflect real-world conditions, which can lead to underperformance in actual scenarios. To address this lack of data, simulation is often utilized. However, accurately simulating physical conditions, such as deformation and friction related to robot-environment interactions, as well as sensor measurements such as image, force, and torque, is difficult in simulations. Therefore, a method is required to improve the transferability of DRL policies to the real environment.

The sample efficiency challenge refers to the fact that with limited data, DRL algorithms that are sample-efficient are required. In other words, these algorithms should enable learning with less data. One possible approach for improving the sample efficiency of a DRL algorithm is to use curriculum learning (CL). However, conventional CL approaches change the environment to progressively enable the policy to learn from simple and difficult concepts, and real-world environments do not often provide the unimpeded possibility to change the environment. Therefore, there is an urgent need for an alternative method that can improve sample efficiency without changing the environment.

Finally, the generalization capability challenge refers to the fact that a DRL policy trained for one environment may not be used in another environment with limited data, and if the conditions change compared to the conditions in which the DRL policy was trained, the DRL policy could underperform. Therefore, a DRL policy that can generalize to different environments and environmental conditions is required.

To overcome the aforementioned challenges, we propose: (1) An offline training

framework that uses maps (in this study, hole maps) of the environment to train the DRL policy to be transferable to the real environment, (2) two action space curriculum learning methods that optimizes the usage of a small dataset via the progression of the action space during training instead of the environment, and (3) a visual attention deep reinforcement learning policy that improves the DRL policy's generalization to different visual conditions.

We evaluated the proposed approaches with an industrial robot controlled by the trained DRL policies to execute the anchor bolt insertion task. This task is a peg-in-hole task executed extensively in construction to insert anchor bolts into holes in concrete. It includes two main subtasks: hole search and peg insertion. Since hammering is used for peg insertion into the holes in concrete, this research focuses on the hole search task. The hole search task is executed in an experimental setup with 13 holes pre-opened in concrete; the performance of the approaches is compared based on the success rate, which should be high to be comparable to human performance, and through the policy training and task completion time, which should be short to fulfill the construction lead time requirements.

This thesis comprises seven chapters. Chapter 1 introduces the study's background, research objective, and an overview of our proposed methods.

Chapter 2 reviews the existing research in the related fields. We first review DRL approaches and compare them to imitation learning approaches that are also used for robotic task execution. Then, we review sim-to-real approaches that are conventionally used to address the transferability challenge. Next, we review curriculum learning approaches, including related methods for improving the effectiveness of policy actions. Furthermore, we survey spatial attention research, which is related to our approach to incorporate visual information for improving task execution performance. Then, we review classical peg-in-hole approaches, followed by analyzing the most recent data-driven approaches. Finally, we position our study's contribution in relation to existing works.

In Chapter 3, we introduce the anchor bolt insertion task, clarify the issues related to the automation of this task, and introduce the experimental setup required to evaluate the proposed approaches. In addition, we propose a DRL policy trained online in the experimental setup to perform the hole search for the anchor bolt insertion task. We use this policy as the baseline to compare to the proposed approaches to adapt to real-world data limitations. The policy is input with force, moment, and robot displacement, and it outputs discrete actions (specifically, search direction and step size) to guide the robot to perform the search. To move between search positions, a hopping motion is used to avoid the high and variable friction of

the concrete surface. At the end of this chapter, we demonstrate that this baseline DRL policy guides the robot to find the holes successfully; however, its success rate is low, and its completion time and training time are long. Additionally, the training is conducted with the real robot, which degraded the robot and the environment.

In Chapter 4, we propose an offline training framework that includes (1) acquiring a discrete hole map by attempting to insert a peg at multiple positions around the hole and storing the data after each attempt and (2) using this hole map to train the policy. By using this framework, the DRL policy can be trained offline, allowing a faster and less hardware-degrading training. Moreover, the framework allows for a smooth transfer of the DRL for usage in the real world because the training is conducted with data from the real environment.

In Chapter 5, we propose two action space curriculum learning approaches to improve sample efficiency. These approaches are designed to optimize the usage of small datasets by progressing the action space size instead of the environment. The first approach starts with a small and simple action space that provides a high success rate, and then, it gradually increases the action space size and complexity. The second approach masks part of the action space for training and selects the part of the action space to be used depending on the competency progress of each part. In this way, it prioritizes the action space that learns faster to improve action usage learning. Although both approaches improve the DRL policy performance with the same amount of data, the second approach achieves a better performance improvement, can be trained faster, and does not require manual dataset handling, making it the most suitable approach for automating real-world tasks.

In Chapter 6, we propose an end-to-end model that includes a DRL policy integrated with a visual spatial attention mechanism for improving generalization to different image inputs. The spatial attention mechanism generates points of high relevance (i.e., attention points) from images, even when the lighting conditions drastically change from the data used for training. The DRL policy uses these attention points and the proprioceptive input used in Chapter 3's DRL policy to generate actions for the hole search. The end-to-end training enables the attention mechanism to generate task-specific attention points and the DRL policy to generate attention point-specific robot actions. We demonstrate that this model considerably improves the DRL policy performance even under challenging lighting conditions.

Finally, Chapter 7 concludes this dissertation by summarizing the achievements of this study in terms of adapting a DRL policy to real-world data limitations, providing possible applications of the proposed approach to other similar tasks, reviewing the remaining issues, and proposing future research directions.

List of research achievements for application of Doctor of Engineering, Waseda University

Full Name : 保富　アンドレ 裕二　seal or signature

Date Submitted(yyyy/mm/dd): 2023/07/02

種類別 (By Type)	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む） (theme, journal name, date & year of publication, name of authors inc. yourself)
発表	○ A. Y. Yasutomi and T. Ogata, "Automatic Action Space Curriculum Learning with Dynamic Per-Step Masking," IEEE International Conference on Automation Science and Engineering, "Accepted".
論文	○ A. Y. Yasutomi, H. Mori and T. Ogata, "Visual Spatial Attention and Proprioceptive Data-Driven Reinforcement Learning for Robust Peg-in-Hole Task Under Variable Conditions," in IEEE Robotics and Automation Letters, 2023, vol.8, No. 3, pp. 1834-1841.
発表	○ A. Y. Yasutomi, H. Mori and T. Ogata, "Curriculum-based Offline Network Training for Improvement of Peg-in-hole Task Performance for Holes in Concrete," 2022 IEEE/SICE International Symposium on System Integration (SII), 2022, pp. 712-717.
発表	○ A. Y. Yasutomi, H. Mori and T. Ogata, "A Peg-in-hole Task Strategy for Holes in Concrete," 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 2205-2211.