

Curriculum learning for robot manipulation tasks through environment shifts

Bachelor Thesis

Advisor: Erdi Sayar (erdi.sayar@tum.de)

Supervisor: Prof. Alois Knoll (erdi.sayar@tum.de)

Introduction and Problem Description

Reinforcement Learning (RL) help an agent learn optimal behavior through trial-and-error interactions with an environment. One of the challenges in RL is to design appropriate reward function which involves multiple factors and engineering labors. Therefore, binary reward is used for developing RL algorithms. Recently, Hindsight Experience Replay (HER) has been proposed to allow off-policy RL algorithms to perform effective learning in solving goal-based tasks with sparse/binary rewards, such as the manipulation of robotic arms [1]. HER takes advantage of failed trajectories by replacing desired goals with the achieved goals. However, it cannot solve the tasks If desired goal is far away from the initial states. Curriculum-based RL approaches decompose complex tasks into sequences of gradually more difficult tasks, by relying on heuristics that guide the agent to explore the environment more efficiently. One possible alternative is to generate curricula in an open-ended way, i.e., by creating novel and increasingly more complex tasks without bounds.[2]

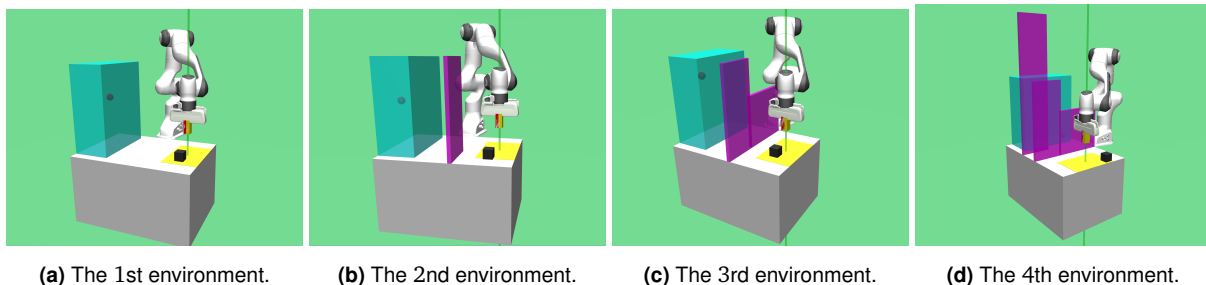


Figure 1: Environment evolution in the PickAndPlace task.

Task Description

In this thesis, your task will be learning state-of-the-art knowledge of reinforcement learning and HER, then develop curriculum algorithms. To be specific:

- You will first learn basic knowledge of reinforcement learning.
- You will reproduce the results from HER and other related research results. By doing this, you will have a deep understanding of HER and the state-of-the-art research results.
- You will generate different environment open-ended way using evolutionary algorithm.

Requirements

- High self-motivation;
- Experiences or knowledge from related courses
- Python programming experiences

References

- [1] Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O., and Zaremba, W. "Hindsight experience replay". In: *Advances in neural information processing systems* 30 (2017).
- [2] Wang, R., Lehman, J., Clune, J., and Stanley, K. O. "Paired open-ended trailblazer (poet): Endlessly generating increasingly complex and diverse learning environments and their solutions". In: *arXiv preprint arXiv:1901.01753* (2019).