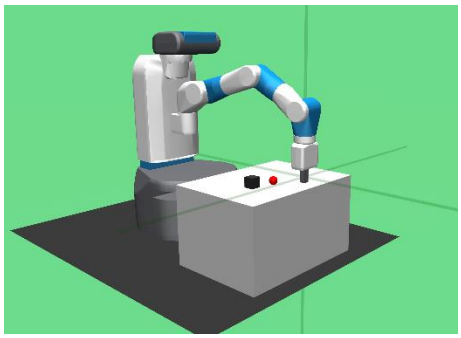
Deep Reinforcement Learning for the Control of Robotic Manipulation

Introduction:

Reinforcement Learning (RL) is a method of machine learning in which an agent learns a strategy through interactions with its environment that maximizes the rewards it receives from the environment. The agent is not given a policy but is guided only by positive and negative rewards and optimizes his behavior. In many real-world scenarios, an agent faces the challenges of sparse extrinsic rewards, learning from limited samples, and no violation of safety constraints [1].



Fetch-Push environment [7]

In our work [2], we use constrained policy optimization (CPO) [3], where the agent tries to improve its performance without exceeding a predefined cost threshold. CPO is a policy gradient method, where the objective is to optimize the expected commutated rewards. Such methods are not efficient because we can use the data sampled from a policy only to improve it, and then we cannot use them again [4].

To improve the efficiency of this method, we used model-based RL to approximate a model of the dynamic. For this thesis, we aim to use another method to improve the efficiency of our model, namely value function-based methods, like, TD3 [4]. Value function-based methods aim to minimize the Bellman error and achieve that it can use all the transition generated to improve the policy.

Another advantage of value function-based methods is that they allow us to combine RL with methods like Hindsight Replay (HER) [5] to learn from failure. The idea behind HER is simple, even though we have not succeeded at a specific goal, we have achieved a different one.

The task is to develop the method used in CPO to optimize the policy with another method from the value function-based area of research to improve the efficiency and parallel save the safety constraints. Then to combine the method with the HER algorithm to deal with sparse rewards.

Task:

The task is to develop the policy optimization method used in CPO with another method from the value function-based research area to improve efficiency and address safety constraints in parallel. The next step is to combine the resulted method with the HER algorithm to deal with sparse rewards. Finally, to evaluate the resulted method in some robotics environments like Fetch-Push [7].

References:

[1]. G. Arnold et al. "Challenges of Real-World Reinforcement Learning."
[2]. M. Zanger*, K. Daaboul* et al. "Safe Continuous Control with Constrained Model-Based Policy Optimization."
[3]. J. Achiam et al. "Constrained Policy Optimization."
[4]. R. S. Sutton et al. "Reinforcement Learning: An Introduction"
[5]. Fujimoto et al. "Addressing Function Approximation Error in Actor-Critic Methods."
[6]. M. Andrychowicz et al. "Hindsight Experience Replay."
[7]. https://gym.openai.com/envs/FetchPush-v1/