A Reinforcement Learning Based Approach for Robotic Grasping Tasks

Asad Ali Shahid
Department of Mechanical Engineering
Politecnico di Milano
Milano, Italy
asadali.shahid@mail.polimi.it

Loris Roveda
Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA)
Scuola Universitaria Professionale della Svizzera Italiana (SUPSI)
Università della Svizzera Italiana (USI) IDSIA-SUPSI Lugano, Switzerland
loris.roveda@idsia.ch ORCID: 0000-0002-4427-536X

Dario Piga
Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA)
Scuola Universitaria Professionale della Svizzera Italiana (SUPSI)
Università della Svizzera Italiana (USI) IDSIA-SUPSI Lugano, Switzerland
dario.piga@supsi.ch

Francesco Braghin
Department of Mechanical Engineering
Politecnico di Milano
Milano, Italy
francesco.braghin@mail.polimi.it

***Abstract*—Robots are nowadays increasingly required to deal with (partially) unknown tasks and situations. The robot has, therefore, to adapt its behavior to the specific working conditions. Reinforcement learning (RL) holds the promise of autonomously learning new control policies through trial-and-error. However, RL approaches are prone to learning with high samples, particularly for vision-based continuous control tasks. In this paper, a learning based method is presented that uses simulation data and kinematic state information to learn an object manipulation task through RL. Unlike vision, the input modality based on kinematic information allows learning with less samples and facilitates transfer to the real-world without additional training requirements. The control policy is parameterized by a neural network and learned using modern Proximal Policy Optimization (PPO) algorithm. A dense reward function has been designed for the task to enable efficient learning of an agent. The proposed approach is trained entirely in simulation (exploiting the MuJoCo environment) from scratch without any prior demonstrations of the task. A grasping task involving a Franka Emika Panda manipulator has been considered as the reference task to be learned. The proposed approach has been demonstrated to be generalizable across multiple object geometries and initial robot/parts configurations. Furthermore, it is shown that the learned policy can transfer the past learning experience and quickly adapt to new variations of the environment by fine tuning through on-policy RL, having the robot able to re-execute the target task in modified setting.**

***Index Terms*—reinforcement learning, intelligent robotics, object manipulation, proximal policy optimization**

I. Paper Contribution

Training in simulation is a feasible approach for learning manipulation tasks with deep reinforcement learning [2]. Simulation enables access to the full state of the system, e.g., part position, allowing faster training of RL policies. In this paper, a model-free RL method is proposed that leverages simulation data and proprioceptive state information to learn a grasping task. Specifically, Proximal Policy Optimization (PPO) is elected to train the robot controller, due to its recent success on difficult control problems, e.g., manipulation of a Rubik's cube with robot hand [1]. Therefore, the control actions (i.e., the desired joint velocities and gripper's joint position) are learned in continuous spaces on the basis of a defined reward function, allowing to take into account the success of the task and its performance. The learned behavior can then be transferred to the real robot, to execute the target task. Furthermore, an adaptation procedure is presented that improves the sample complexity by initializing the target task's policy with the base policy, thus, reducing the required amount of training data. The proposed approach has been demonstrated to be generalizable across multiple object geometries and initial robot/parts configurations. In fact, despite the task has been trained only on a single cube with no prior task information, the learning is able to generalize across different geometric shapes and sizes, minor changes in the position of the part to be manipulated, and across multiple initial configurations of the robot.

II. Task Description

The aim of the paper is to implement the described approach to autonomously learn a grasping task. The task consists of a robot interacting with a cube of nominal size 6 cm placed on a table. The goal of the robot is to successfully position its grip site around the cube, grasp it and then finally lift it off the contact surface. In order to evaluate the proposed approach on a grasping task learning, Franka Emika Panda, a 7-DoF torque-controlled robot is used as a robotic platform.

III. Results

PPO is applied to the task of lifting the cube above a certain height. The main aim is to analyze whether the proposed method can learn to accomplish this task and how well can it generalize to new situations that were not seen during training. The policy is trained for a total of 10 million time steps where

The work has been developed within the project ASSASSINN, funded from H2020 CleanSky 2 under grant agreement n. 886977

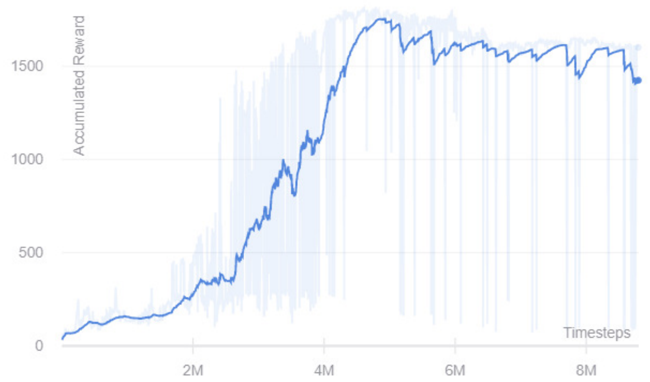

Fig. 1. PPO training results. The plot depicts the progression of mean accumulated reward for episodes.

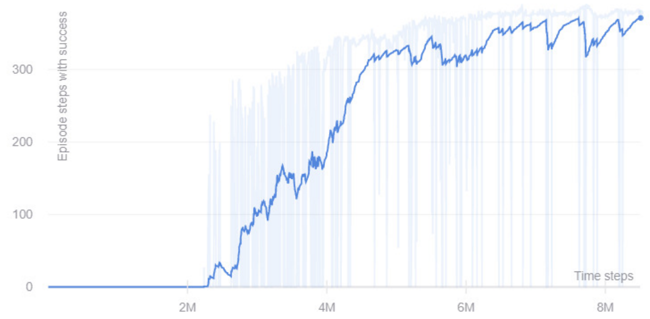

Fig. 2. PPO training results. The plot depicts the progression of mean success rate for episodes.

each episode lasts for 600 steps, giving an agent approximately 2.5 seconds to accomplish the task. At the beginning of each episode, initial configuration of the robot and the cube is reset to fixed position. The results of training are shown in Figure 1 and Figure 2, suggesting that the the agent has learned to perform the task successfully after 4 million steps.

The generalisation capabilities and robustness of the learned neural network policy have been tested, considering nominal cube and different geometries to be grasped:

- nominal cube with size 6 cm;
- smaller cube with size 4 cm;
- cylinder with size 3 cm radius and 3 cm height;
- screw-driver.

10 test trials have been run and results are summarised w.r.t. successful completion of the task. The proposed model shows the success rate of 100% in all four cases. In last two cases, the policy's ability to grasp new shapes is tested by replacing the cube with a cylinder and a screw-driver placed nearly at the cube's original position.

Table II shows the results in which the original position of robot and/or of the cube has been changed. In the first case, a small random noise has been added to the initial joint positions of the robot at the beginning of each episode. The robot can still grasp and lift the cube in most cases, whereas in 2 failed trials, the robot could not successfully position the gripper around the cube and remained near the cube for the rest of episode. For the second case, the nominal position of the cube has been modified by ±8 cm in order to test the policy's robustness to variation in part's position. For this specific case, 20 trails are performed, 10 for each direction and success rate is reported as the mean of 20 trials. The task has been completed successfully in most test runs, while sometimes the robot's actions results in failure. In these specific situations, the robot either grasped the cube at the edges resulting in unstable grasp or collided its gripper with the cube not being able to grasp successfully. In the last test, the potential for adapting the learned task's policy has been performed with significant changes in position of the cube. With simple fine-tuning procedure for a pre-trained policy, the resulting performance achieves 100% success within 30 minutes of additional training, indicating that the policy can successfully reuse prior task experience and quickly learns to perform in new task settings.

IV. Discussion

In this paper, an intelligent task learning has been formulated as a RL problem, demonstrating the possibility of learning low-level control actions purely from gathered experience in a simulated environment. Results show that it is possible to train continuous control actions based only on the state observations, and in a reasonable amount of time. Achieved results highlight that the learned policy performs well to new situations, and it can also adapt its learning to significant variations of the environment with slight amount of additional training.

References

[1] I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas et al., "Solving rubik's cube with a robot hand," arXiv preprint arXiv:1910.07113, 2019.

[2] A. Rajeswaran, V. Kumar, A. Gupta, J. Schulman, E. Todorov, and S. Levine, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," CoRR, abs/1709.10087, 2017.

TABLE I
Evaluation trials for fixed positions.

Test	Object	Success rate
1.	Nominal cube	10/10
2.	Smaller cube	10/10
3.	Cylinder	10/10
4.	Screw-driver	10/10

TABLE II
Evaluation trials for varied positions.

Test	Variable	Amount	Additional training	Success rate
1.	Robot position	max 2%	No	8/10
2.	Cube position	±8 cm	No	7/10
3.	Cube position	+10 cm	Yes	10/10