Evaluating Hierarchical Deep Learning Methods For Vision Based Robot Manipulation Of Dynamic Objects In Unstructured Environments

Department for Interdisciplinary Research
College of Engineering Trivandrum

November 21, 2019

Guided by,
Linu Shine
Electronics and Comm. Engg.

Presented by,
Sreejith Krishnan R
Robotics & Automation

Motivation

Why robot manipulation?

- In order to assist in general tasks, autonomous robots should be able to interact with dynamic objects in unstructured environments
- Designing machines that can grasp and manipulate objects with anything approaching human levels of dexterity is first on the to-do list for robotics [1]

Hierarchical reinforcement learning allows to,

- Decompose complex task into hierarchy of sub-tasks
- Reuse sub-tasks in similar domain
- DNA for artificial intelligent agents

" Stop learning tasks, start learning skills " (Satinder Singh)

Literature survey I

Title	Comparing Task Simplifications to Learn Closed-Loop Object Picking Using Deep Reinforcement Learning [2] - IEEE Robotics and Automation Letters - 2019
Methodology	Uses autoencoder to reduce dimensionality of camera data which is given to 3 layer CNN to get a low dimensional encoding. This encoding is used by a 2 layer feed-forward neural network to predict the optimum action. Uses RL to train the mentioned networks
Merits	• No hand labeled data required
Demerits	• Low success rate (78%) for manipulation of objects in clutter by real robot • Non modular. Difficult to reuse model for similar task

Literature survey II

Title	Regularized Hierarchical Policies for Compositional Transfer in Robotics [3] - DeepMind - 2019
Methodology	Use hierarchical modular policies for continuous control.
Merits	• Best sample efficiency on both simulated and real robot • Uses MPO optimization algorithm which reduces the number of hyperparameters
Demerits	• High level tasks are not automatically decomposed to sub tasks • Low level policy is shared across all low level tasks making interpretability complicated • Transferring specific skills from sub-tasks policy in a predictable manner is difficult • Experiment results are obtained using model whose inputs include pose of objects in workspace

Research Gap and Objectives

Research Gap

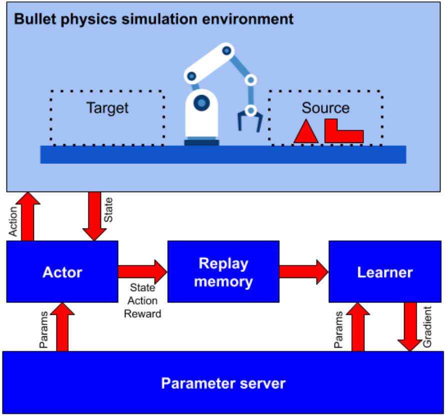
- Low success rate when transferring policies from simulation to real world
- Low sample efficiency
- Low interpretability of learned models
- Predictable transfer of learned skills are difficult
- Manual task decomposition

Objective

Compare RHPO Hierarchical Reinforcement Learning Algorithm for vision based robot manipulation with baseline RL algorithms - PPO and DDPG

Methodology I

- HRL method will be evaluated based on success rate, mean picks per hour and training time
- Simulation environment will be created on Bullet Physics Simulator
- For fast experiment feedback loop, experiments will be run in distributed fashion. Simulation environments will be tuned for maximum throughput.



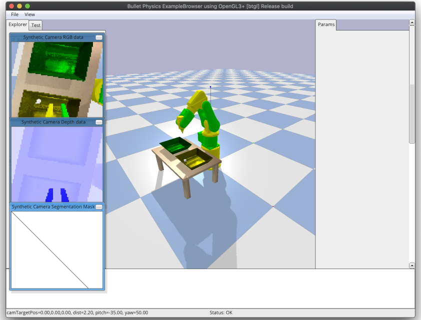
- HRL algorithm will be compared with baseline model free continuous action space PPO, DDPG and RHPO algorithms

Methodology II

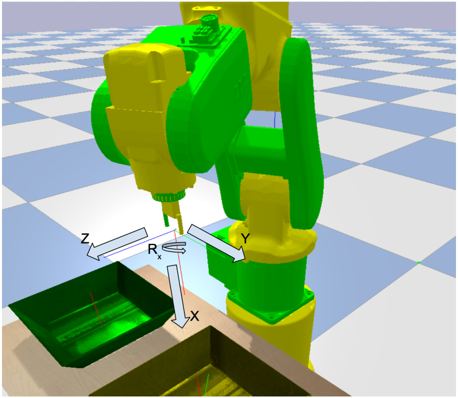
- For evaluating transferability of skills, similar tasks like table clearing and stacking of objects will be considered and reduction in training time will be measured
- Policies learned from simulator will be evaluated on ABB IRB 120 robot. Depth camera mounted on endeffector will provide visual feedback. Camera will be directly connected to computer. Program running in the computer will directly connect with via serial port in robot controller. RAPID program running on robot will read actions send from computer, execute it and send feedback to computer.

Results
ABB IRB 120 simulator

- PD control tuned for 1mm positioning accuracy which is same as ABB IRB 120 robot positioning accuracy
- 20ms mean action time without rendering and 60ms mean action time with rendering
- Multiple simulator instances can be run in parallel to scale up agent sampling throughput
- Grasp detection and collision detection algorithms



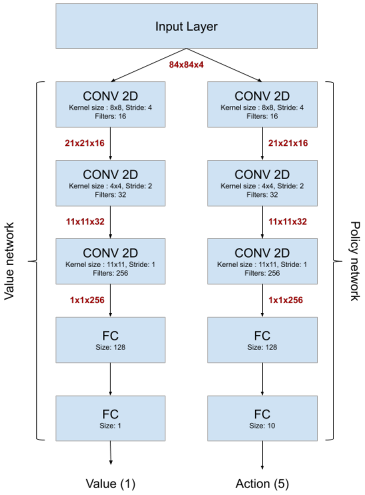
- Task is to move the object from yellow tray to green tray
- Only input to RL model is the depth image from RGB-D camera mounted on end effector. Observation space is of shape 84x84x4
- RL model can move and rotate the end effector by providing input relative to coordinate frame attached to end effector. Binary variable can be modified to open / close gripper. Action space is $[\delta x, \delta y, \delta z, \delta r_x, open]$

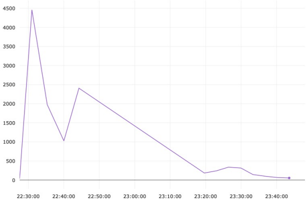


Results
Baseline PPO

- Convolution layers are used for feature extraction
- Input in RGB-D 84x84x4 matrix
- Value network predicts the how good it is to be at a particular state
- Policy network directly predicts the mean and standard deviation of a Gaussian PDF from which actions are sampled for a particular state
- Train batch size is 10240 and SGD minibatch size is 512. Number of iterations per train batch is 30
- Data from an episode is added to train batch only after the it is complete

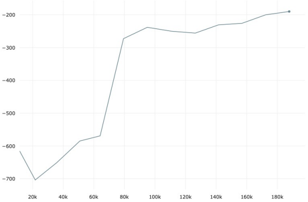
Architecture



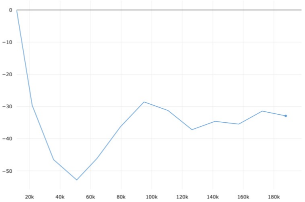


Value network loss

Policy network loss

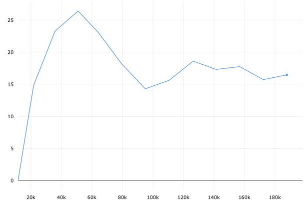


Collision penalty

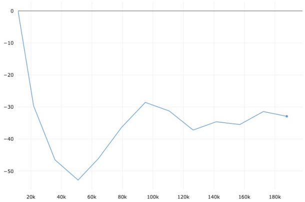


Drop penalty

- Reward shaping is critical. Small changes in reward function can change the learning process.
- Providing +ve reward when end effector moves towards target and -ve reward when end effector moves away will not work
- Initializing each episode at a random state like grasped and not grasped can improve training speed



Grasp Reward



Drop Penalty

Future scope

- Including DDPG and RHPO baseline models
- Evaluating baseline models on real robot
- Exploring methods for automatic task decomposition

References

Richard Hodson.
How robots are grasping the art of gripping.
https://www.nature.com/articles/d41586-018-05093-1.
Accessed: 2019-09-24.

M. Breyer, F. Furrer, T. Novkovic, R. Siegwart, and J. Nieto.
Comparing task simplifications to learn closed-loop object picking using deep reinforcement learning.
IEEE Robotics and Automation Letters, 4(2):1549–1556, April 2019.

Markus Wulfmeier, Abbas Abdolmaleki, Roland Hafner, Jost Tobias Springenberg, Michael Neunert, Tim Hertweck, Thomas Lampe, Noah Siegel, Nicolas Heess, and Martin A. Riedmiller.
Regularized hierarchical policies for compositional transfer in robotics.
CoRR, abs/1906.11228, 2019.