Parallelized deep reinforcement learning for robotic manipulation

Master thesis specification and schedule

Isac Arnekvist

isacar@kth.se

KTH Robotics Perception and Learning

Supervisor: Johannes A. Stork

January 18, 2017

Contents

1 Background

1.1 Objective

The interest in conducting this thesis research started with a series of articles published by researchers at Google in their research blog [1–4]. The main theme in these articles was robotic manipulation learned by gathering experience in real time in non-simulated contexts. These articles will be presented in more detail below and extended during the pilot study. In two of these articles [1, 2] tasks are learned from scratch without the need for initializing by demonstration. Although, in the article by Gu et al. [1], poses of targets and arms are known by attached equipment. It would be interesting to incorporate estimation of poses from visual feedback in this case to lessen the need for external equipment. Another central theme in these articles is the distributed collection of experience over several robots. This is done in order to decrease the time it takes to collect data and to increase variance of the data. The use cases for incorporating and extending these findings could be robotic manipulation tasks with camera as feedback where exact relative positions of objects, manipulators, and sensors need not be fixed. Also, where resources exists to use several robots for speeding up the learning process. Possible readers might be other researchers working with end-to-end machine learning for robotic manipulation. Other interested parties might also be manufacturers where repetitive tasks are a part of the production chain and variations in these make it hard for robots to be easily programmed for those tasks.

1.2 Pilot study

The following sections are the preliminary sources of information that was the initial spark for this thesis as mentioned above. How to re-implement these articles is not self-contained, so the pilot would necessarily need to also include reading into articles from the references of these. Reading of these initial articles would be needed to motivate an appropriate method, and then further research would be done with the purpose of gaining all the information needed to implement such a solution. The thesis study will be conducted at the Robotics, Perception, and Learning lab at KTH with the main interest originally being to dig into these articles and develop something further.

1.2.1 Motion planning by "Deep visual foresight"

This article [2] trains a convolutional neural network on images together with motion as inputs to predict how the image will change due to that motion. This is later used to plan movement of objects to some target pose.

1.2.2 Path Integral Guided Policy Search

In this article [4], the authors extend Guided Policy Search and demonstrate two manipulation tasks. These are initialized from demonstrations. To be able to comprehend this article, referenced articles [5–7] would have to be read as well.

1.2.3 Collective Robot Reinforcement Learning with Distributed Asynchronous Guided Policy Search

This article [3] distributes learning of door opening across several robots. The exact nature of the tasks are varied across robots to increase robustness. The learning is initialized from demonstration.

1.2.4 Parallelized training without prior demonstration

This article [1] shows several robotic manipulation tasks where learning is parallelized across platforms, and they do not require previous demonstrations. For this article, I would need to read up on an algorithm called Normalized Advantage Function (NAF) [8]. In both of the two previously mentioned articles [3, 4] pose estimation of targets and robots are done through visual feedback, while in this article [1] no sensory feedback is provided and poses are known through attached equipment. The pose estimation was done using a convolutional neural network which could be a feasible extension to this article.

1.2.5 Reinforcement learning

These articles mentioned above naturally deals with reinforcement vocabulary and assumes knowledge in this area. Therefore the pilot would include studying a book by Sutton and Barto [9]. In this book, chapters 1-3 and a section about non-linear function approximators are essential (by advice from supervisor).

1.3 Problem statement

Manipulation tasks that seem trivial to a human can be hard to learn for robots, especially from scratch without initial human demonstration due to high sample complexity. Recent research suggests ways to do this but are based on that you know the poses of the objects and the end-effector. For some scenarios these are non-trivial to find out.

Problems also arise when learning in real time by collecting experience. Robots must be able to evaluate their policies regularly at a high rate which is complicated by adding a deep convolutional neural network for pose detection. Also, learning tasks within a feasible time frame is harder when data collection and policy updates happen in real time. The approach of distributing collection of experience over several robots will be evaluated in this thesis for handling this problem.

1.4 Research question

How can deep and distributed reinforcement learning be used for learning and performing dynamic manipulation tasks with unknown poses.

1.5 Expected scientific results

If all goes well, previous results are verified in new contexts. Also they are extended to also handle unknown target and manipulator poses.

2 Method

2.1 Examination method

Preliminary method is using the mentioned distributed version of NAF and extend it with pose estimates from a convolutional neural network. This network is pretrained as in [3] by randomly placing objects and the end-effector and this way generating training data. Several robots will be used to parallelize the training process. The preliminary manipulation task is pushing of objects to some random target position.

2.2 Conditions

There will be need for several robot setups, each including a robot, computer, and camera. These will have to be able to communicate with a separate computer responsible for training the policies/neural networks. In the ideal case, this computer is supplied and has a graphics card compatible with modern neural network libraries.

2.3 Limitations

A proof of concept should be done with a corresponding report (the thesis). There are no requirements for implementation of code that should be delivered as libraries etc. The main contribution is the thesis. All code used for conducting the experiments will be openly published on GitHub.

3 Schedule

3.1 Weekly plan

V.3 Finalize this document

V.4-7 Pilot study and write down related background sections

V.8 Set up robots, method section will be written in parallel

V.9 End-effector and object pose estimation

V.10-11 Implement reinforcement learning algorithms

V.12-13 Tweak and fix bugs in order to accomplish task

V.14 Record and write down results

V.15 Finish the remainder of the thesis (Conclusions/Future work), hand in for review

V.16-17 Review and adjustment process with supervisor

V.18 All reviews from supervisor and corresponding adjustments done. Ready for presentation/public discussion and approval from examiner

V.20 Oral presentation

V.22 Finishing touches, hand in final report to supervisor and examiner

References

[1] Shixiang Gu, Ethan Holly, Timothy Lillicrap, and Sergey Levine. Deep reinforcement learning for robotic manipulation. *arXiv preprint arXiv:1610.00633*, 2016.

[2] Chelsea Finn and Sergey Levine. Deep visual foresight for planning robot motion. *arXiv preprint arXiv:1610.00696*, 2016.

[3] Ali Yahya, Adrian Li, Mrinal Kalakrishnan, Yevgen Chebotar, and Sergey Levine. Collective robot reinforcement learning with distributed asynchronous guided policy search. *arXiv preprint arXiv:1610.00673*, 2016.

[4] Yevgen Chebotar, Mrinal Kalakrishnan, Ali Yahya, Adrian Li, Stefan Schaal, and Sergey Levine. Path integral guided policy search. *arXiv preprint arXiv:1610.00529*, 2016.

[5] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.

[6] Evangelos Theodorou, Jonas Buchli, and Stefan Schaal. A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, 11(Nov):3137–3181, 2010.

[7] William H Montgomery and Sergey Levine. Guided policy search via approximate mirror descent. In *Advances in Neural Information Processing Systems*, pages 4008–4016, 2016.

[8] Shixiang Gu, Timothy Lillicrap, Ilya Sutskever, and Sergey Levine. Continuous deep q-learning with model-based acceleration. *arXiv preprint arXiv:1603.00748*, 2016.

[9] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.