Literature Review

Bardia Mojra

February 24, 2021

Robotic Vision Lab

The University of Texas at Arlington

Reinforcement Learning of Active Vision for Manipulating Objects under Occlusions

1 Metadata

- Authors: Ricson Cheng, Arpit Agarwal, Katerina Fragkiadaki
- Code: `https://github.com/ricsonc/ActiveVisionManipulation`
- Paper: `https://arxiv.org/pdf/1811.08067.pdf`

2 Introduction

In this paper, the authors propose a novel reinforcement learning method for monocular RGB grasping systems where camera pose is controlled through visual feedback to reduce occlusions. The task of simple object grasping has been fairly achieved but in the real world, there are often occlusions with other objects that needs . First, the authors pose the question of learning manipulation policies under occlusions and propose agents capable of hand-eye movement coordination with various distractors present in the scene. Secondly, they introduce a **modular actor-critic network architecture** based on [1] for active perception and action in Mujoco simulation environment, [2]. This paper examines various reinforcement learning modalities and highlights the importance of environment difficulty (distractors) in **curriculum learning** methods.

3 Problem Statement

The authors focus on the task of pushing an object to target locations in environments with distractors where an actor-critic architecture is deployed for eye-hand coordination. Hand-eye or camera-griper coordination is base on [3] and by integrating it, the authors, aim to achieve state estimation easier (train faster) by reducing an **information gap** or deficiency caused by static camera. First, they trained a vanilla CNN for the task of pushing an object with minimal distractors present, which failed. Then they integrated an object detector module in the actor-critic architecture which enabled effective learning. Secondly, they trained state-of-the-art reinforcement learning models with simple distractors present and occasionally occluded the target object, which also failed to learn. Then, they initialized the actor-critic network weights from policies learned in environments without distractors. This reinforces what is hypothesized by *curriculum learning.*

4 Method

The authors represent the mentioned problem in form a multi-goal Partially Observable Markov Decision Process (POMDP) which is a constraint satisfaction problem formulation. The reinforcement learning environment is modelled by the observations ($\mathcal{O}$), states ($\mathcal{S}$), goals ($\mathcal{G}$), gripper actions ($\mathcal{A}^{\mathcal{G}}$) and camera actions ($\mathcal{A}^{\mathcal{C}}$) spaces.

The critic network (a CNN) take RGB images as input and encodes low dimensional embeddings using what-where decomposition which is presents object appearances f_t. Moreover, a faster-RCNN [4] is used for object detection, followed by PnP for target object pose estimation, $\hat{o}_t$.

The authors use HER's [1] object-centric representation as it is shown to result in faster learning based on empirical data. HER also introduces the powerful idea of learning from failed experiences are encoded and stored in an **experience buffer** to draw heuristics from in possible future states. This allows the agent to extract much greater level of information from previously seen data.

5 Experiments

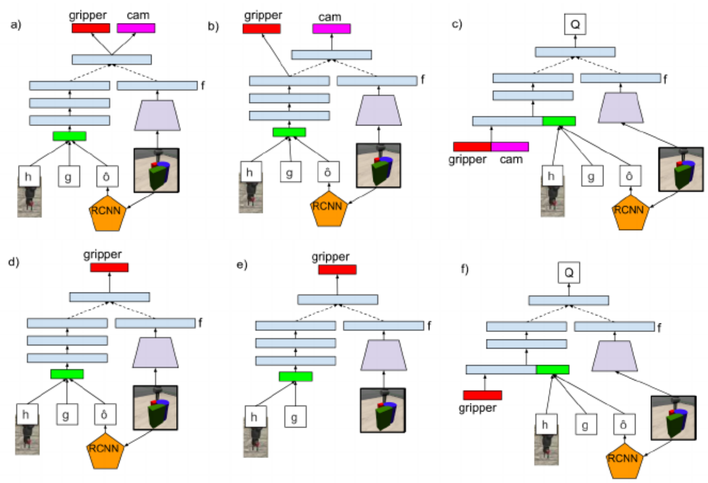

Figure 1: **Actors and critic architectures for active** *(top row)* **and static** *(bottom row)* **camera agents**. a) Actor network for *cam-active-full*. The same state representation is fed as input to hand and eye actor networks. b) Actor network for *cam-active-abstr*. The gripper actor responsible for pushing the object does not receive direct information from the RGB input, but only the 3D location of the object. c) Critic network for all active camera agents. d) Actor network for *cam-static*, the agent can only control its gripper, not its camera. e) Actor network for *cam-static-image*. The pretrained object detector subnetwork is omitted. f) Critic network for *cam-static*. Dotted lines denote elementwise addition. Trapezoids denote convolutional neural sub-networks, and blue rectangles denote fully connected layers.

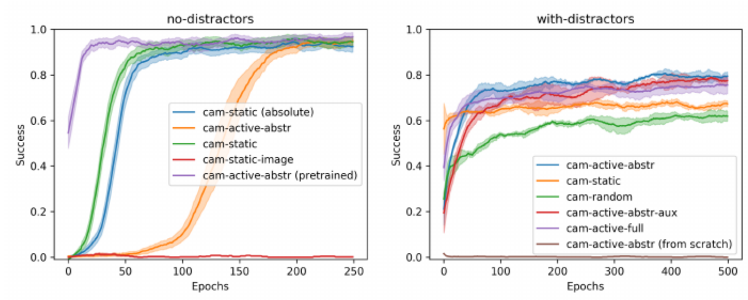

Figure 3: *Left:* **Environments without distractors.** Hand-eye policies train slower (*cam-active-abstr*), yet all architectures achieve good asymptotic performance. Hand-eye policies can be effectively pretrained from hand only policies (*cam-active-abstr* (pretrained)). Object-centric state encoding is beneficial (*cam-static* outperforms *cam-static (absolute)*). Finally, ignoring the location of the object of interest provided by the detector, and rather using only frame-centric appearance encoding does not result in successful behaviour (*cam-static-image*). *Right:* **Environments with distractors.** Active vision helps to handle occlusions from distractors (*cam-active-abstr* outperforms *cam-static*). State abstraction helps for the hand actor policy (*cam-active-abstr* outperforms *cam-active-full*). Training directly in the environment with distractors, without pretraining on the easier environment does not result in successful behaviours (*cam-active-abstr* (from scratch)). Auxiliary visibility reward is not helpful (*cam-active-abstr-aux*)necessary. A learned camera policy is superior to a random camera policy (*cam-random*). Shaded area shows 1 standard error on the mean fraction of episodes which ended with success during training. We took the mean and computed the error over 20 episodes in each of 5 training runs using different seeds.

References

[1] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba, “Hindsight experience replay,” *arXiv preprint arXiv:1707.01495*, 2017.

[2] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, IEEE, 2012.

[3] S. Soatto, “Actionable information in vision,” in *Machine learning for computer vision*, pp. 17–48, Springer, 2013.

[4] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *arXiv preprint arXiv:1506.01497*, 2015.