Deep RL Arm Manipulation

Stephan Becker

Abstract—This paper aims to give the results for a Deep Reinforcement Learning project, where a robotic arm is trained to touch a small tube. The objectives of the project where to

- Have any part of the robot arm touch the object of interest, with at least a 90% accuracy for a minimum of 100 runs.
- Have only the gripper base of the robot arm touch the object, with at least a 80% accuracy for a minimum of 100 runs.

Both objectives could be achieved by tuning the rewards to the respective tasks.

Index Terms—Robot, IEEEtran, Mobile Robotics, DeepRL.

1 Reward Functions

The robotic arm was controlled by updating the joint positions. Each joint's position could be either increased or decreased by the agent.

The reward for winning an episode (REWARD_WIN) was set to +10, and to -10 for losing (REWARD_LOSS).

The full winning reward was issued when the arm successfully touched the tube for task 1 or when the gripper based touched the tube for task 2. Hitting the ground was penalized by REWARD_LOSS, as was running out of time (exceeding 100 frames). In all those cases the episode was terminated.

To guide the arm toward the goal intermediate rewards were used. These intermediate rewards were based on the (smoothed) progress the gripper was making towards the goal. Failing to make sufficient progress or moving in the wrong direction resulted in a small penalty, while progress towards the goal resulted in a small (based on the magnitude of the progress) reward.

2 Hyperparameters

Trainig was completed on a AWS p2 instance. The input for the agent was a 128x128x3 image.

The agent used a LSTM network of size 256 and a replay memory of 20000. The network was trained with RMSprop, a learning rate of 0.1 and a batch size of 256. The choice of parameters was mainly based on the default parameters used in the original implementation by Dustin Franklin, with some increases to input size, replay memory and batch size to better utilize the capabilities of the p2 instance.

3 Results

The RL agent was able to quickly (within 50 episodes) achieve the goal accuracy of 90% on the first task of being able to touch the tube with any part of the arm. The agent would then just repeat the winning trajectory over and over again, finishing each episode quickly and with high accuracy.

For the second task the agent needed considerably more training time to converge on the 80% overall accuracy;

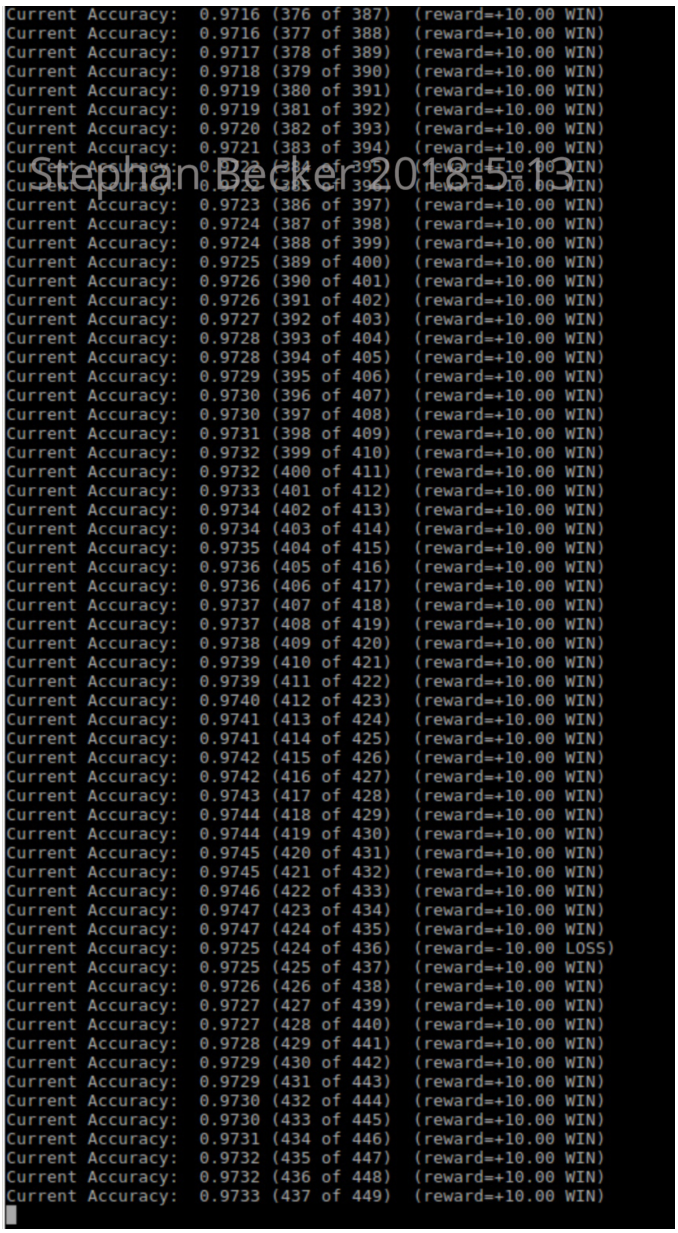

Fig. 1. Task 1 accuracy

eventually the agent was able to consistently touch the tube with the gripper's base, hitting the ground very rarely (approx. 1 in 10 approaches). The agent had more trouble finding an optimal trajectory than in the first task. Even with many episodes of training the arm would sometimes just move a small distance forward and back again repeatedly for multiple frames.

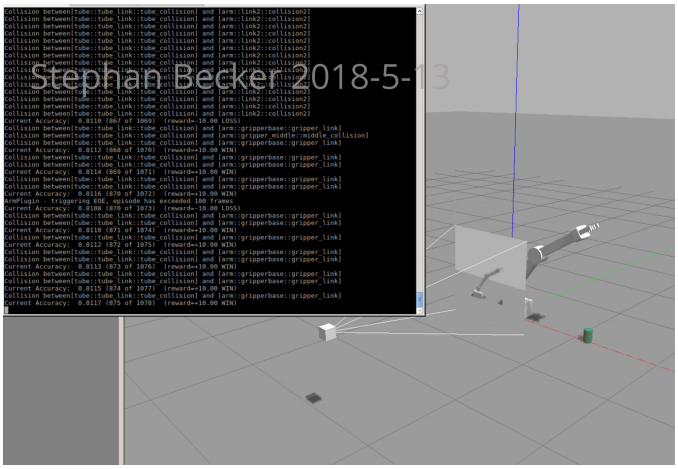

Fig. 2. Task 2 accuracy

4 Future Work

Especially for the second task the convergence too a long time; it would be worthwhile to further experiment with the reward system to get the agent to learn more quickly. Building on that the additional challenges (random placement of the tube, additional DoF for the arm) should be a future project to be investigated.