Report

Udacity Deep Reinforcement Learning

Navigation Project report

Overview

The goal of this project was to coach an agent to navigate an easy world environment, collecting yellow bananas while avoiding blue bananas.

The project environment is made in Unity and may be a version of the Banana Collector, customised for the Udacity Deep Reinforcement Learning nanodegree. The task is episodic and there are 37 different state space dimensions with 4 different possible actions. Rewards of +1 are accumulated when a yellow banana is collected and -1 when a blue banana is picked up. There are not any rewards or penalties for moving. The task is taken into account solved when a score of 13 or more is achieved over 100 consecutive episodes.

The model used successfully solved the matter after 422 episodes, which was exceptionally surprising given the extended target within the project was to possess it achieve this with 1300 episodes

Learning algorithm

The solution implemented may be a simple Deep Q-Learning (DQN) algorithm supported the classic Deep Mind paper from Nature. Hassabis et al. Human-level control through deep reinforcement learning. Nature February 2015. It included experience replay and glued Q targets, both of which are considered standard during a DQN

Model architecture

A range of neural models were tried exploring wide, deep and shallow configurations. Overall, the simpler models performed also or better than deeper ones and wide models performed worse than narrow ones. All models started with a 37 x 1 input vector from the environment, constructed two or more fully connected hidden layers and ended with a totally connected layer outputting 4 outputs, one for every action.

Hyper parameters

Replay buffer size 100,000

Report

Discount factor 0.99 (gamma)

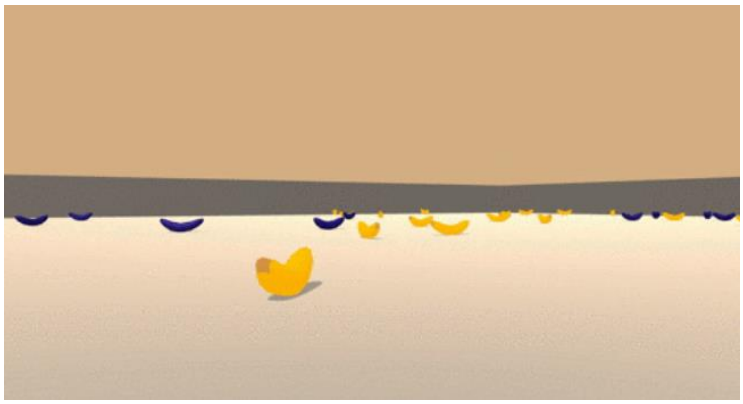
Soft update factor 0.001 (tau)

Learning rate 0.0004 (alpha)

Network update step interval 4

Results

I was surprised by the results, specifically, how easy it had been to realize the goal score with almost no modifications. The consistent solving of the matter when training, and therefore the low number of episodes required, wasn't expected.



We had expected to spend significant time tuning hyperdata's and running on GPU instances, however, our MacBook Pro CPU completed the training in under 5 minutes.