Project Report

The goal of this project was to teach a simulated agent to collect yellow bananas in a rectangular arena while avoiding blue bananas. The environment is a pre-configuted Unity game created with the Unity Machine Learning Agents Toolkit.

The Environment

The environment provides a 37-dimensional continuous observation state space employing ray-based vision. The action state space contains four discrete actions corresponding to move forward, move backward, turn right and turn left. The environment is considered solved when an agent is able to achieve a total score of 3 or more points over 100 consecutive episodes.

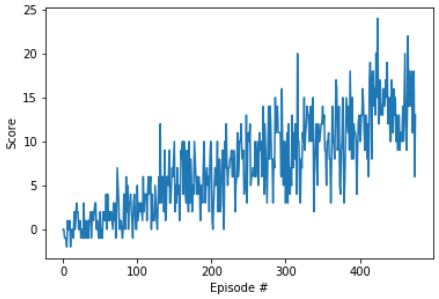
Learning Algorithm

The agent is trained using Deep-Q-Learning (see Readme for reference) with fixed Q Targets. The Q-function is approximated using fully connected feed forward neural networks, with the inputs the size of the state space, two hidden layers with 64 nodes each and a linear output layer of the same dimension as the action space. The hidden layers use ReLU activation.

To remove correlation from the experienced states and actions, the agent uses a Replay Buffer that holds the last 10000 SARS observations and samples randomly from this buffer to train the neural network in batches of length 64. The learning rate is set to 0.0005 and a learning cycle is performed every four steps. Rewards are discounted with a factor of 0.99. For exploration, the agent uses an epsilon value that starts at 0.99 and then decreases until 0.001 with a decay rate of 0.01. The DQN-code and hyper parameters are mostly based on the DRLND DQN-network exercise.

Scores

The following graph shows how the agent's score increases consistently. While there is significant variance in the scores, the agent learns continuously until he meets the specified goal after less than 500 episodes.



Future work

For future implementations, it might be interesting to see how additional tuning of the hyper parameters affects training. It might also be beneficial to add make use of more advanced Deep-Q-Learning techniques like prioritized experience replay or Double-DQN.