Project 3 (Collaboration and Competition) Report

Learning Algorithm

The algorithm chosen is the Multi-Agent Deep Deterministic Policy Gradient (MADDPG), as outlined in this [paper](#), which re-purposes the DDPG algorithm by also considering the policies learnt by the competing / cooperating agent.

Only slight modifications need to be applied to the original DDPG algorithm to get the Multi-Agent version one to work. In particular, both agents share a global replay buffer memory, while retaining their own actor and critic networks.

Model Architecture

Both Actor and Critic share a similar network architecture:

| Layers	# Nodes
Input	24 (Position, Velocity of ball and racket)
Fully Connected 1	256
Fully Connected 2	256 (+ 2 for Critic)
Output	2

I used ReLUs for all the fully connected layers and applied batch normalization to the first fully connected layer.

Hyperparameters

```
BUFFER_SIZE = int(1e6)  # replay buffer size
BATCH_SIZE = 256 # mini-batch size
GAMMA = 0.99 # discount factor
```

```
TAU = 1e-3 # for soft update of target parameters
LR_ACTOR = 2e-4 # learning rate of the actor
LR_CRITIC = 2e-4 # learning rate of the critic
WEIGHT_DECAY = 0.000 # L2 weight decay
```

Plot of Rewards
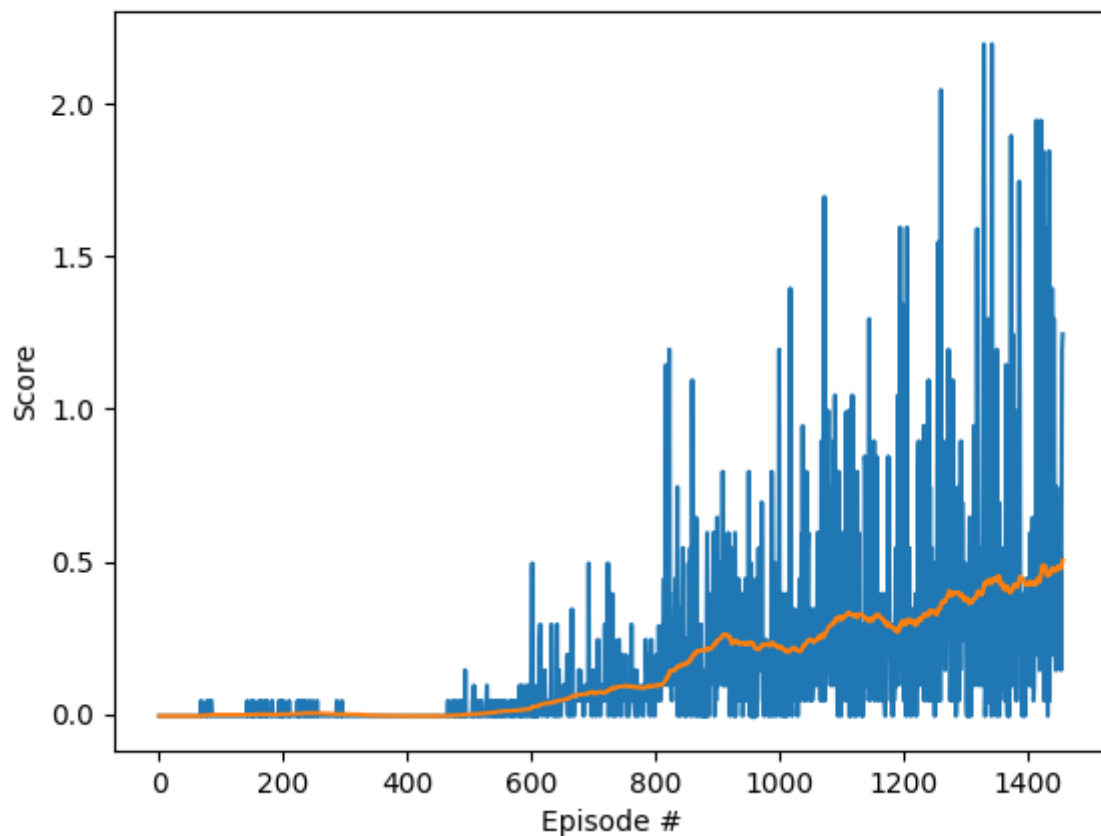
Ideas for Future Work

This project only deals with collaboration. I would also like to apply the same algorithm to the Soccer environment and see how the agents perform.