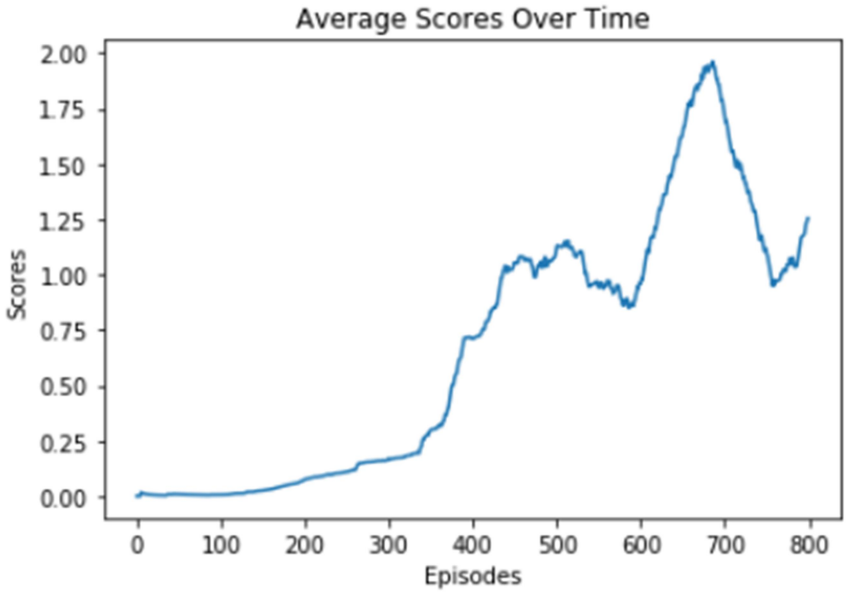
Report

```
Max Average Score: 1.961000029221177
```



The number of episodes needed to solve the environment was about 360 episodes, and reached a max average score of 1.96.

The Learning Algorithm & Neural Network

The learning Algorithm chosen was DDPG since it is a Continuous Control algorithm. The neural network architecture chosen was an Actor/ Critic neural network architecture. The Actor network has three layers with relu activations except for the last layer was TanH. There is 24 units for the first layer, 400 units for the second, and 300 units for the last layer. Next, the Critic network was also a three layer network, but the first layer is used to only encode the state, while the second takes in the encoded state and the action taken. The activations where all Relu except for the last layer which was Linear. The hidden units for Critic network is 400 units for the first hidden layer, 300 for the second hidden layer, and one node for the output layer.

Hyperparameters

The hyperparameters are as follows: The BUFFER_SIZE was 5e5, BATCH_SIZE is 256, GAMMA was set to 0.99, TAU was 1e-3, Learning rate for the Actor was 1e-4, Learning rate for the Critic was 1e-3, and the update multiplier was set to 100.

Ideas for Future Work

To extend the efficiency and maximize this project I would do two things. First, I would limit the size of the networks, by limiting the layers to only two, and limiting the hidden neurons to 200 or less. This would hopefully stabilize training since the original network could be too large and was memorizing rather than generalizing. Finally, I would implement “Importance Sampling” since the reward signal is sparse. By implementing Importance Sampling the buffer would hold more relevant data pointing to the desired goal, and theoretically maximizing performance.