Report

Model

A simple 3 layer deeply connected neural network is used

```
## Hyperparameter
BUFFER_SIZE = int(1e5)  # replay buffer size
BATCH_SIZE = 64         # minibatch size
GAMMA = 0.99            # discount factor
TAU = 1e-3              # for soft update of target parameters
LR = 5e-4               # learning rate
UPDATE_EVERY = 10       # how often to update the network
```

The default for UPDATE_EVERY was 4, and increase up to 10. to make sure the agent can increase its score

Architecture - Learning algorithm

Using Replay buffer architecture to store experience tuple, learn better when do multiple passes over same experience

Using Adam as a optimization technique to update the weight

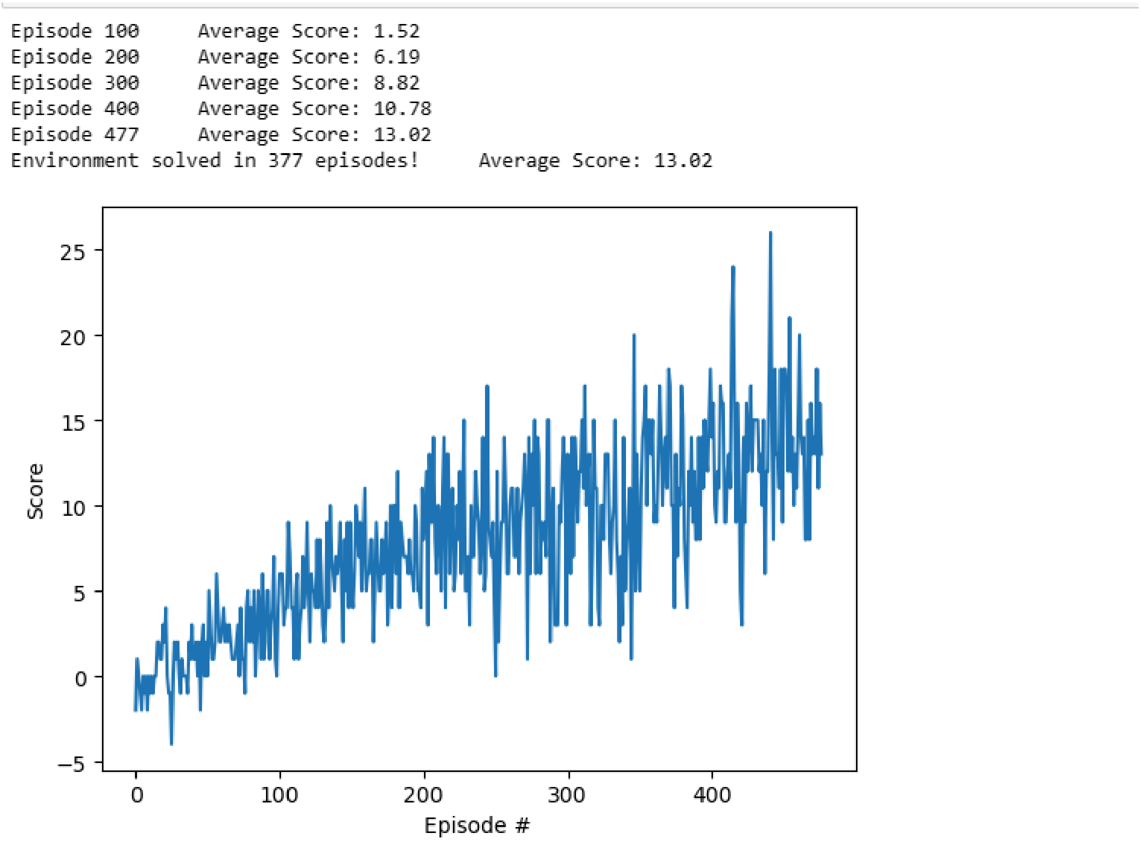
Epsilon-greedy action selection was used to balance exploration (new action with different outcome) and exploitation (leverage known action that yield high reward). A high epsilon means high exploration and vice versa

Plot of reward

Plot of Rewards

The agent will stop once reach 13 average score

The agent also doesn't need to train for 2000 episodes to achieve the following result



```
Episode 100	Average Score: 1.52
Episode 200	Average Score: 6.19
Episode 300	Average Score: 8.82
Episode 400	Average Score: 10.78
Episode 477	Average Score: 13.02
Environment solved in 377 episodes!	Average Score: 13.02
```

Ideas for future work

Using Learning from Pixels: to learn its velocity, along with ray-based perception of objects around its forward direction

To improve the agent, further technique can be used such as Double DQN, Dueling DQN, Rainbow

Model: can use a more complicated model (CNN) to improve the model