Gokul Srinivasan 500594698

Project 3

Deep Q-Learning Network

Architecture:

The input consists of an 84x84x4 image directly fed from the game environment, on which 2D batch normalization is performed. The model has 3 layers of convolution neural networks, and each of them has a ReLu activation function to act as rectifier layer, through which the image is passed.

The Dueling DQN architecture was used, with two additional streams of computation in sequence to the two fully connected layers.

Parameters Used:

The following hyper parameters were used:

A discount factor gamma of 0.99 was used, the closer it is to 1, the better is the performance of the model. A learning rate of 5e-5 was used since the model seemed to perform better for lower value of alpha. The epsilon value was taken as 1 and a decay cycle of 5000000 was used. The train_frequency parameter is the number of steps to be played in each game for every iteration and was set to 4.

```
self.env=env
self.gama = 0.99
self.learning_rate = 5e-5
self.traning_stop = 800.0
self.output_shape = env.action_space.n
self.train_frequency = 4
self.clipping = 1
self.device=T.device('cuda')

self.epsilon = 1
self.epsilon_stop = 0.01
self.decay_cycles = 5000000
self.epsilon_slope = (self.epsilon - self.epsilon_stop)/self.decay_cycles

self.counter = 0
self.best_reward = 0.0
self.mean_reward = []
self.episode_reward = 0.0
self.game_reward = 0.0
```

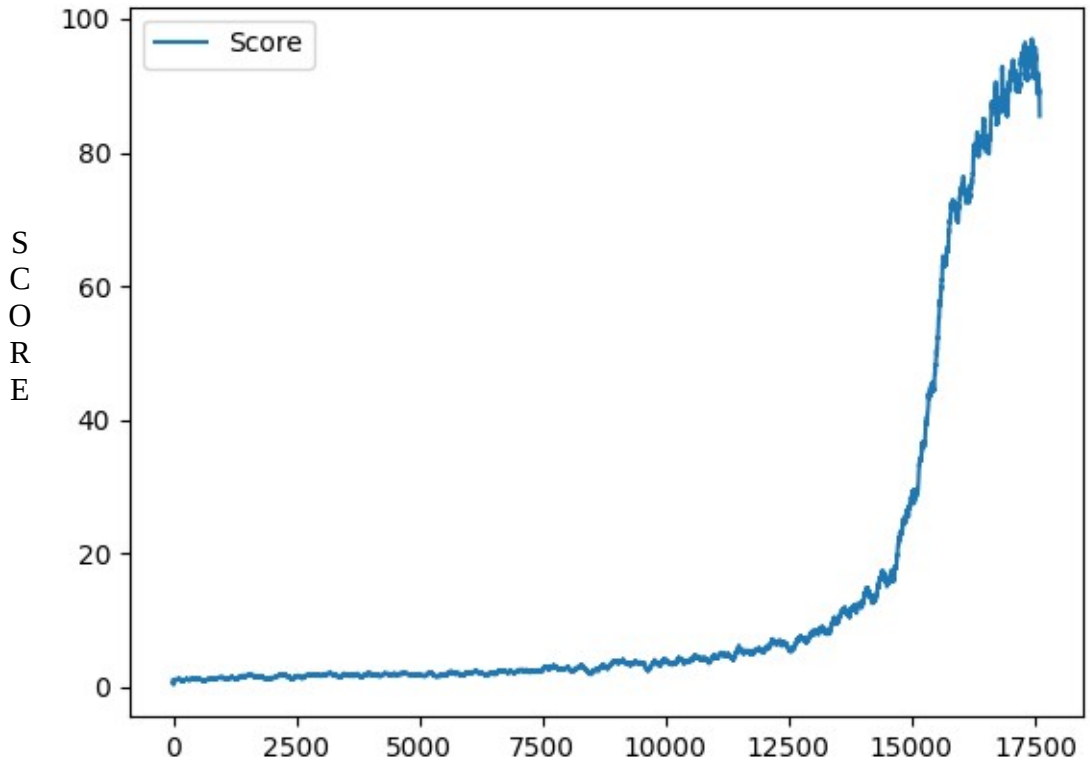
Loss Function and Optimizer:

Mean Squared error loss function and adam optimizer was used for this project.

Visualization:

Learning Curve:

The training was performed for 17600 games and the score per game is shown below. As the number of games increased, the rewards also increased well. Had the model been trained for another 3000 games, the results would have been much better. As such it took 12 hours for training 17600 games, of which the last few thousands took the major portion of the time consumed. As the number of the games increases, the model starts getting better at the game, which results in slow training speeds.



```
(cv) gs@gs:~/Desktop/WPI/Reinforcement-Learning/Projects/Project3/Gsrinivasan_Project_3$ python main.py --test_dqn
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.uint8'>. Please provide explicit dtype.
WARN: gym.spaces.Box autodetected dtype as <class 'numpy.uint8'>. Please provide explicit dtype.
loading trained model
WARN: <class 'atari_wrapper.FrameStack'> doesn't implement 'reset' method, but it implements deprecated '_reset' method.
Reward :  13.0                          Reward :  13.0
Reward :  5.0                           Reward :  5.0
Reward :  99.0                          Reward :  99.0
Reward :  5.0                           Reward :  5.0
Reward :  14.0                          Reward :  14.0
Reward :  86.0                          Reward :  110.0
Reward :  175.0                         Reward :  239.0
Reward :  40.0                          Reward :  16.0
Reward :  27.0                          Reward :  4.0
Reward :  0.0                           Reward :  4.0
Reward :  14.0                          Reward :  86.0
Reward :  0.0                           Reward :  175.0
Reward :  0.0                           Reward :  40.0
Reward :  0.0                           Reward :  27.0
Reward :  0.0                           Reward :  0.0
Reward :  110.0                         Reward :  13.0
Reward :  239.0                         Reward :  5.0
Reward :  16.0                          Reward :  99.0
Reward :  4.0                           Reward :  5.0
Reward :  4.0                           Reward :  14.0
Reward :  110.0                         Reward :  14.0
Reward :  239.0                         Reward :  0.0
Reward :  16.0                          Reward :  0.0
Reward :  4.0                           Reward :  0.0
Reward :  4.0                           Reward :  0.0
Reward :  110.0                         Reward :  86.0
Reward :  239.0                         Reward :  175.0
Reward :  16.0                          Reward :  40.0
Reward :  4.0                           Reward :  27.0
Reward :  4.0                           Reward :  0.0
Reward :  110.0                         Reward :  110.0
Reward :  239.0                         Reward :  239.0
Reward :  16.0                          Reward :  16.0
Reward :  4.0                           Reward :  4.0
Reward :  4.0                           Reward :  4.0
Reward :  13.0                          Reward :  110.0
Reward :  5.0                           Reward :  239.0
Reward :  99.0                          Reward :  16.0
Reward :  5.0                           Reward :  4.0
Reward :  14.0                          Reward :  4.0
Reward :  13.0                          Reward :  86.0
Reward :  5.0                           Reward :  175.0
Reward :  99.0                          Reward :  40.0
Reward :  5.0                           Reward :  27.0
Reward :  14.0                          Reward :  0.0
                                        Reward :  110.0
                                        Reward :  239.0
                                        Reward :  16.0
                                        Reward :  4.0
                                        Reward :  4.0
                                        Reward :  86.0
```

```
Reward :  175.0
Reward :  40.0
Reward :  27.0
Reward :  0.0
Reward :  110.0
Reward :  239.0
Reward :  16.0
Reward :  4.0
Reward :  4.0
Reward :  86.0
Reward :  175.0
Reward :  40.0
Reward :  27.0
Reward :  0.0
Reward :  110.0
Reward :  239.0
Reward :  16.0
Reward :  4.0
Reward :  4.0
Reward :  86.0
Reward :  175.0
Reward :  40.0
Reward :  27.0
Reward :  0.0
Run 100 episodes
Mean: 55.24
```