Project 3 Report

Emmanuel Jayaraju

November 17, 2022

1 Test Results

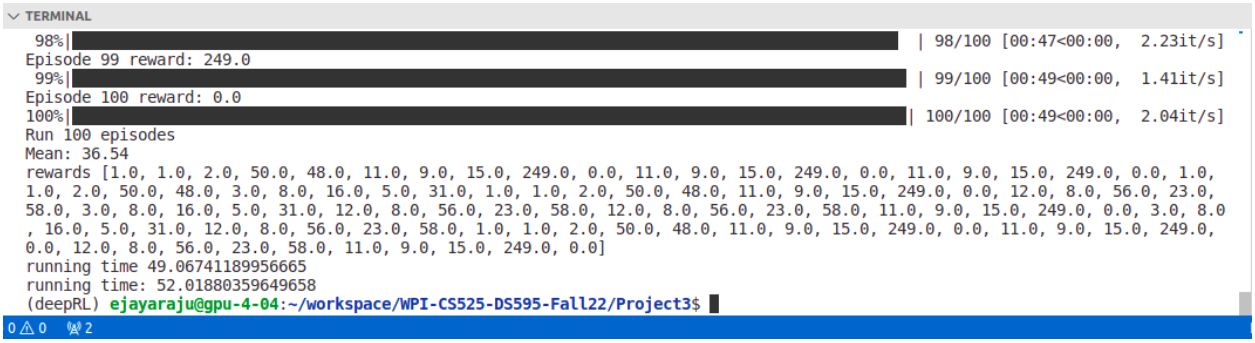

Figure 1: Test Results with best model

After days and hours of trying out various parameters, skipping frames, normalizing images, here is the conclusion of my work for this assignment.

During testing when the rewards are not clipped, the mean reward out of 100 episodes is 36.54, i.e 37 points. The algorithm used was DQN with replay buffer and 2 networks. This is the best result I could get after training for many days with different hyper-parameters as explained in the next section. The DDQN algorithm performed worse for me.

2 Experiments

2.1 Graphs

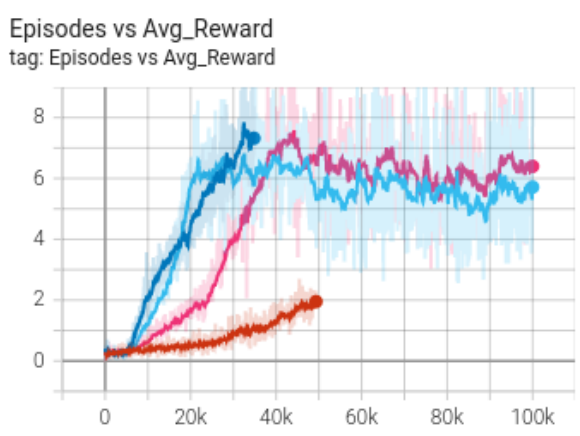

Figure 2: AvgReward vs Episodes

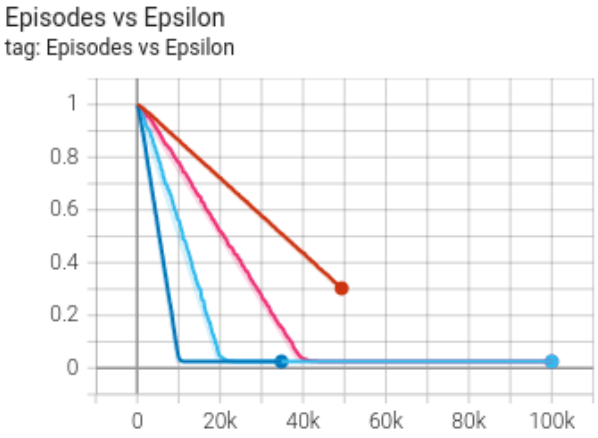

Figure 3: Epsilon vs Episodes

The above graphs are generated using the following:

	Models
RL Algorithm	DQN with * replay buffer * 2 Q networks
DL Network	CNN from DeepMind DQN Paper
Optimizer	Adam
Loss Function	Huber Loss

2.2 Hyper-parameters

The hyper parameters used for each of the experiments in the graph is listed below. The hyper parameters in the last column gave the best results for me. In my understanding, the biggest factor was finding the right amount of exploration using the epsilon to decay around 40K episodes. The graph associated with it is the one in magenta color.

Hyper-parameters	**Exp1 (Dark Red)**	**Exp2 (Dark Blue)**	**Exp3 (Light Blue)**	**Exp4 (Magenta)**
Epsilon_Decayed_at_episode	80K	10K	20K	40K
Learning Rate	1e-4	1e-5	1e-4	1e-4
Gamma (Discount Factor)	1.0	1.0	0.99	0.99
Mini-batch size	64	64	32	32
Frames Skipped	10	0	0	0

2.3 Other Algorithms

I also tried the Double DQN (DDQN) algorithm. But that gave much worse results, i.e. the average rewards during training didn't go beyond 1. In the below pictures, the best DQN result is compared with the DDQN result.

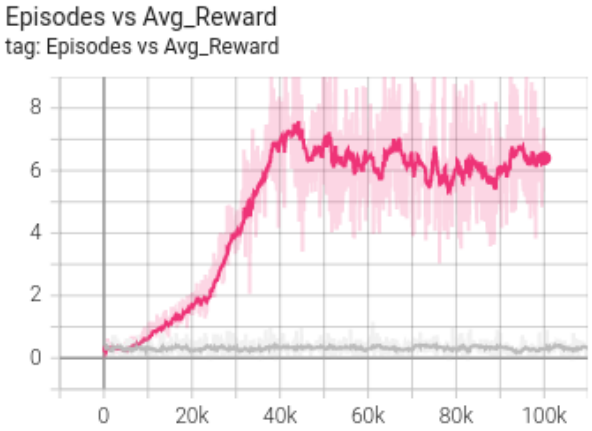

Figure 4: AvgReward vs Episodes

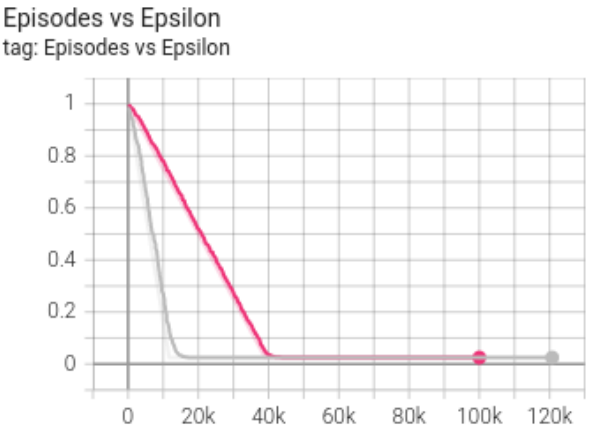

Figure 5: Epsilon vs Episodes

3 Visualization

The best result I got was from the DQN algorithm. The Learning Curve looks as shown below.

1. X-axis: number of episodes
2. Y-axis: average reward in last 30 episodes

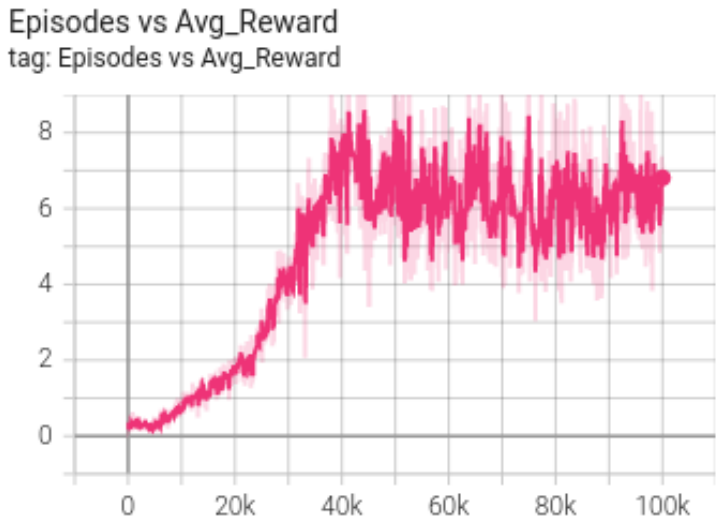

Figure 6: With my best model

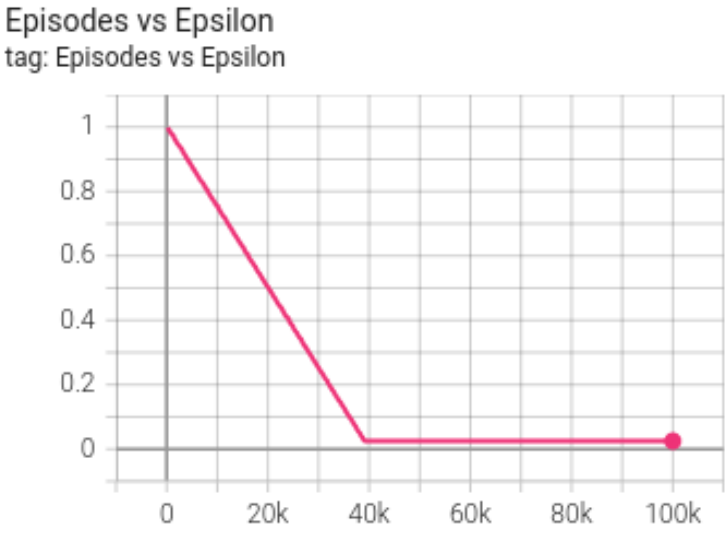

Figure 7: With my best model