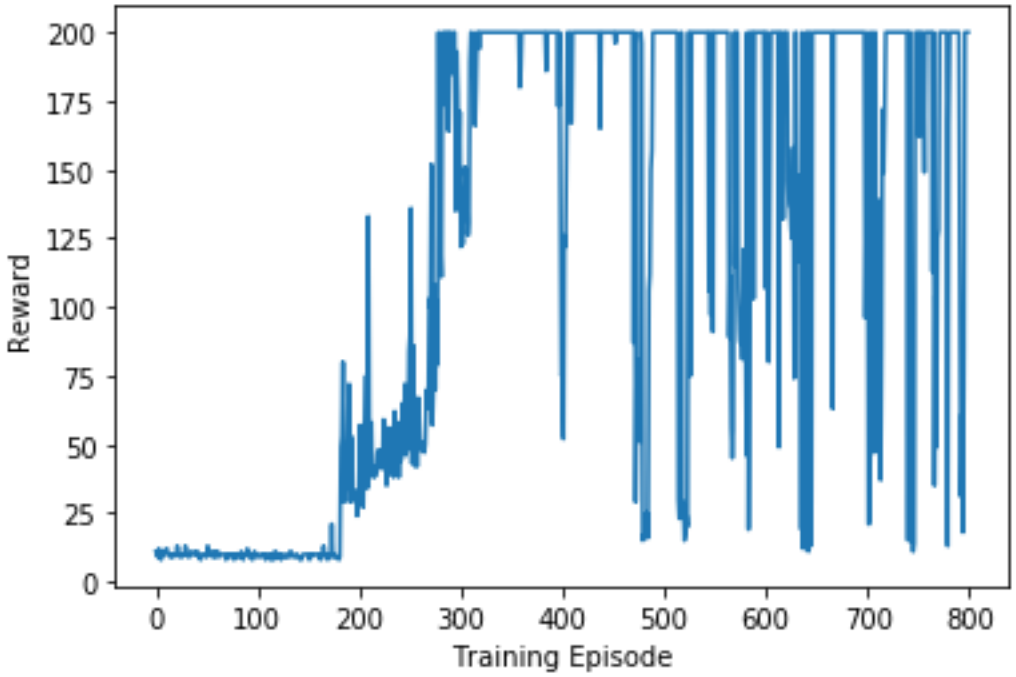
Report for final project

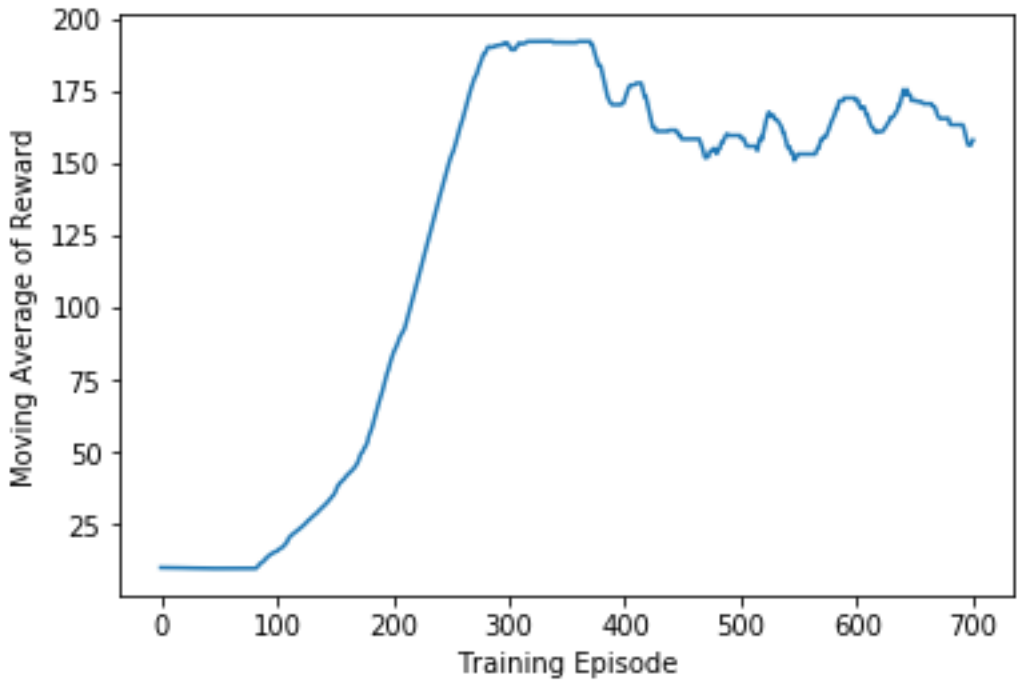
Lifeifan Du

In this project, by now I only finished CartPole. StarGunner is in progress but there are some bugs related to the input of CNN that are still not resolved.

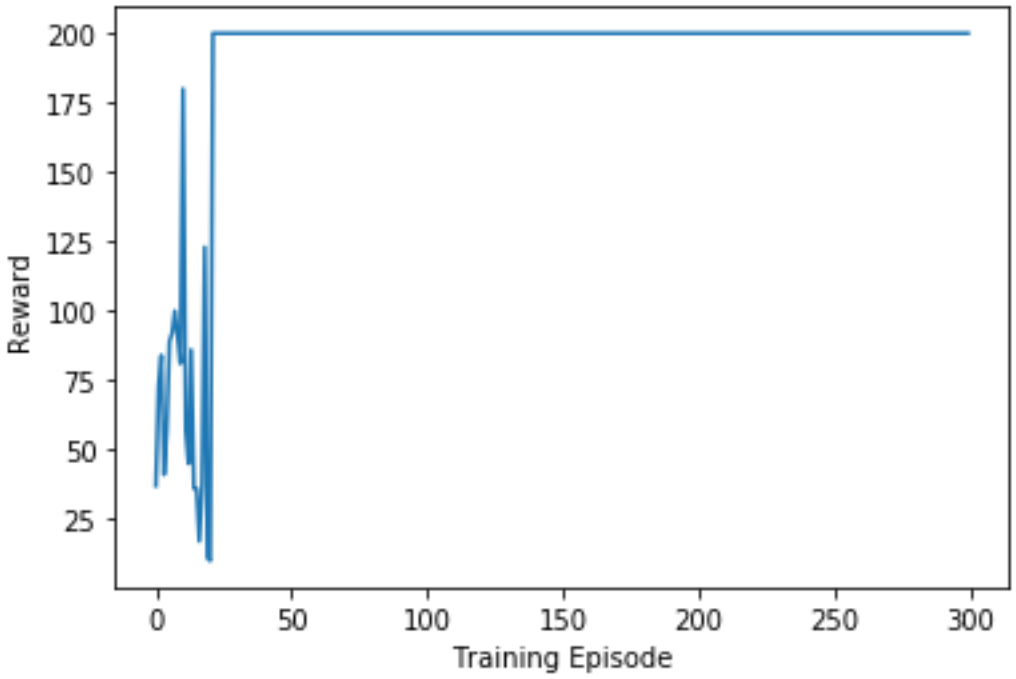
In the CartPole, I tried to change the reward in the batch memory and the training set, but I did not modify the final score given by the env, so the score shown in results is the same as original scores.

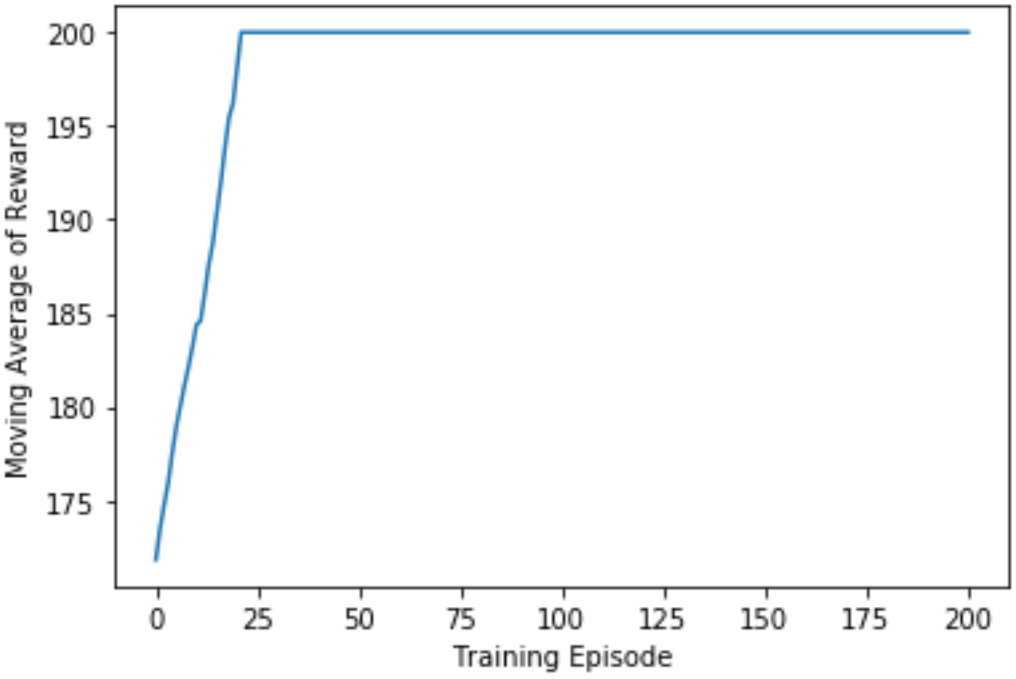
At beginning, I tried to give more penalty to the agent if the agent is "done", which means the agent is "dead". The original reward of "death" is -1, I give it -20. I ran it for 800 episodes, the result is shown below. We can see from the moving average figure that from about 275th episode to about 375th episode, the average is around max score 200, but then maybe diverge and the score is not stable, the average score drops, but the agent is learning. The result is not that good, so I tried another method.





I referred to some materials about the state and reward of CartPole. And one efficient method in the material is to take cart's location and the angle between cart and pole into consider. These factors can be gotten from the state which is in the env. So, I tried this method and used the parameter provided by the material. That is, the agent will get higher reward if it is at the middle location, and the cart and pole are closer to vertical. I just change the reward in the memory and the training set, not the score shown in results. And from the previous method to this method, I just changed this reward given function, and did not change anything in the DQN model. This agent performs so good, and it converges after about 25 episodes. The results are shown below.


Reward
Training Episode
200
175
150
125
100
75
50
25
0
50
100
150
200
250
300


Moving Average of Reward
Training Episode
200
195
190
185
180
175
0
25
50
75
100
125
150
175
200