# Homework 3 DRL

Mohamed Chedhli Bourguiba
DQN

## Question 1

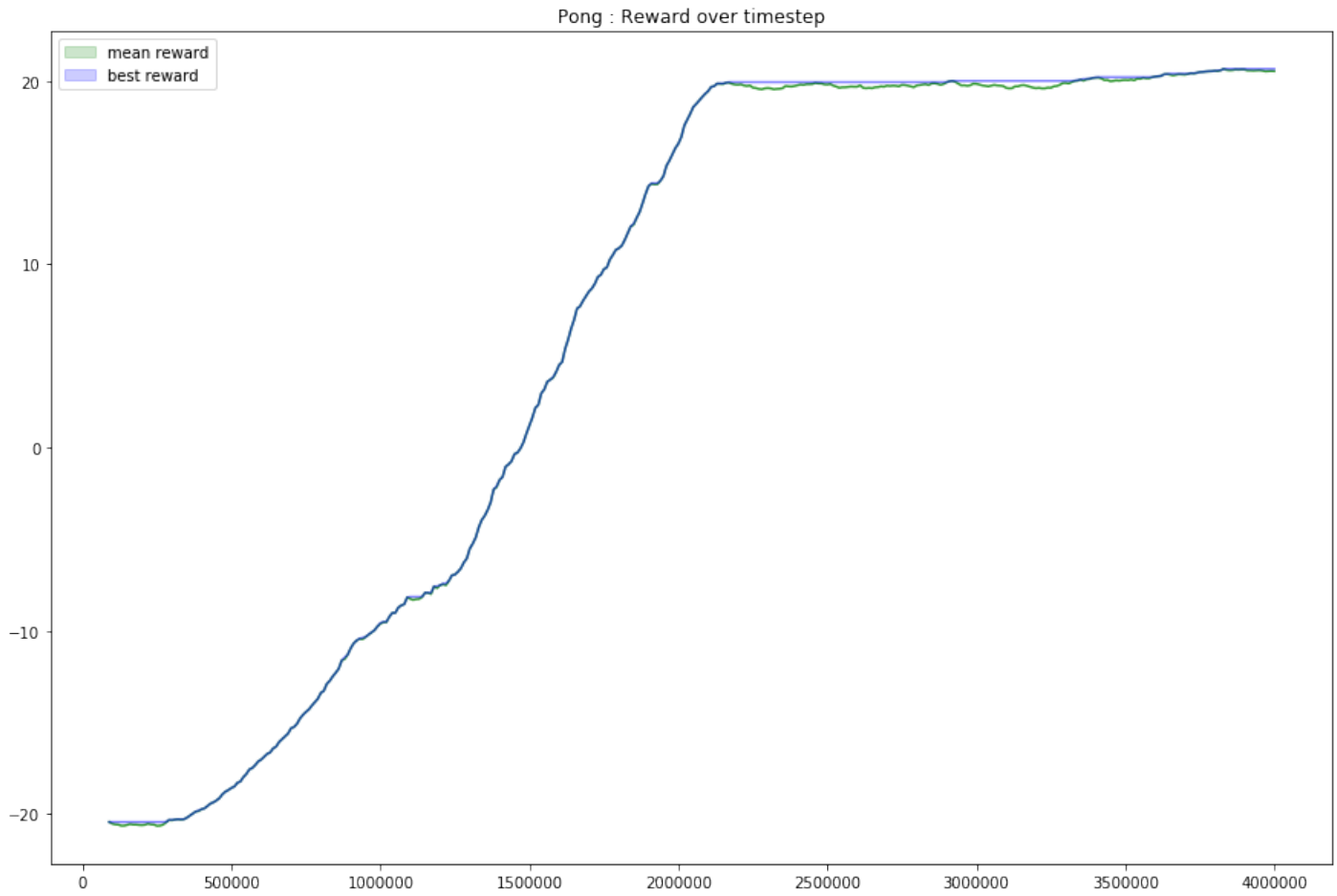


Figure 1: Pong reward. The mean reward is computed over 100 episodes. Default hyper-parameters were used.

# Question 2

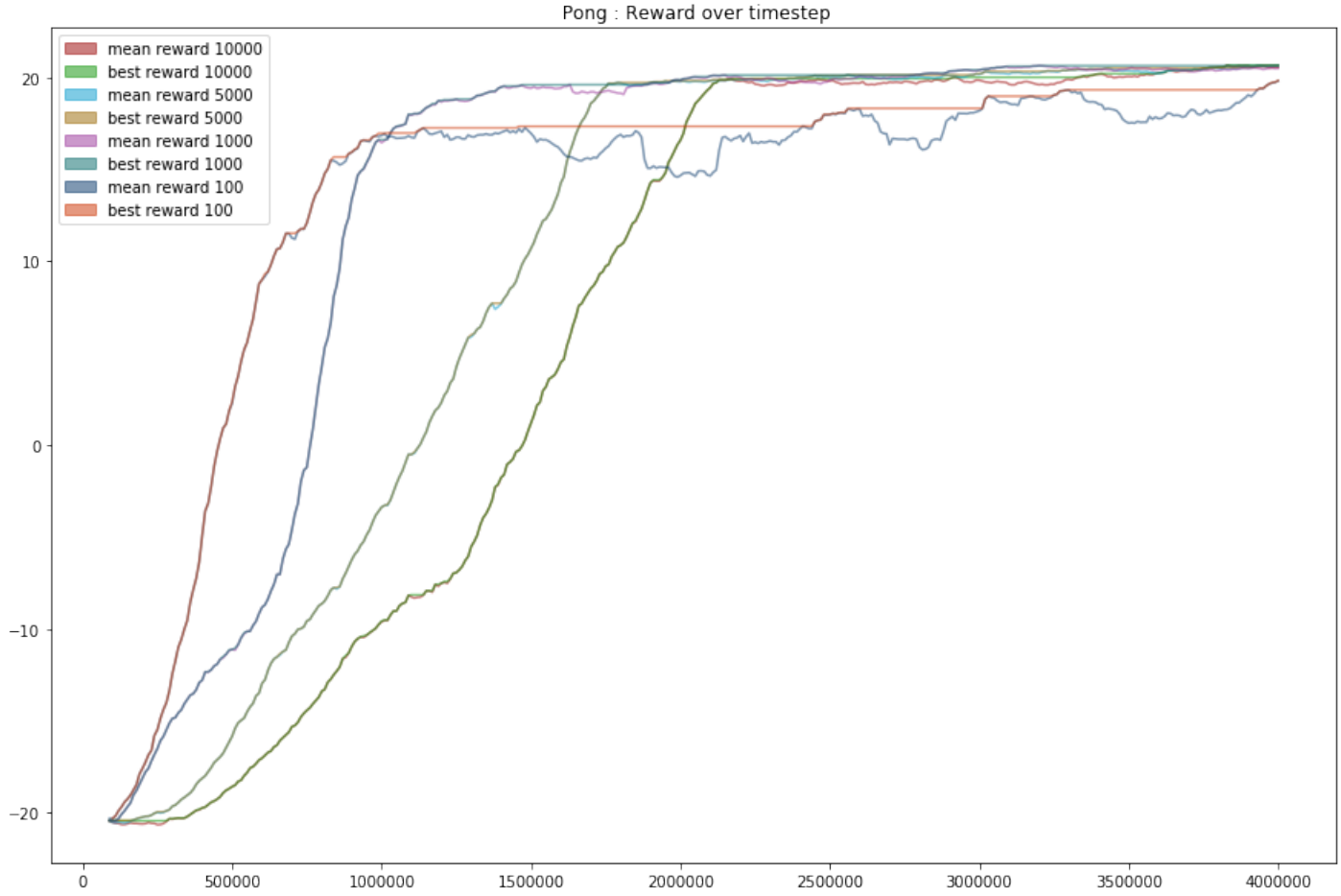


Figure 2: The changed hyper-parameter is the update frequency of the target network. With the decrease of the update frequency of the target network, one would expect there is less lag and the current network can easier fit the target values. Nevertheless there is a trade off between speed/convergence and stability as it can be noticed when the frequency drops to 100, the target is moving too fast and the algorithm does not seem to be as stable as for the other tested frequencies even though the best reward eventually gets close to the maximum reward without reaching it.Besides it requires more training time

# Other games

## Breakout

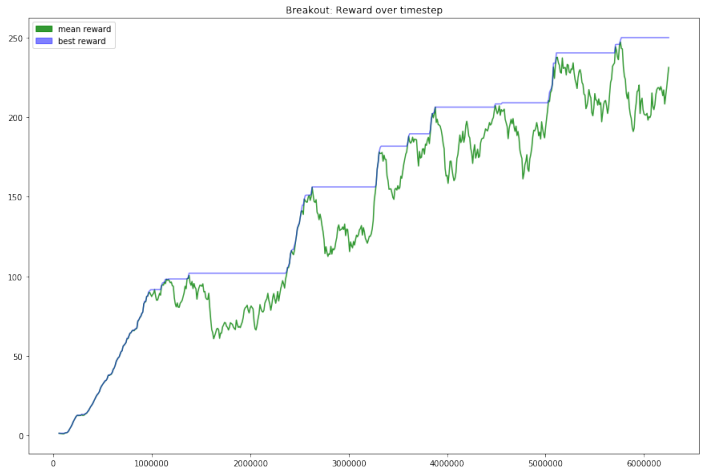


Figure 3: Breakout reward. The mean reward is computed over 100 episodes. Default hyper-parameters were used except target network update frequency which was 1000. The best reward is slightly above average compared to the results published in Openai website. Yet the algorithm is not stable. With a greater target network update frequency and more time steps one would except better results

## Enduro

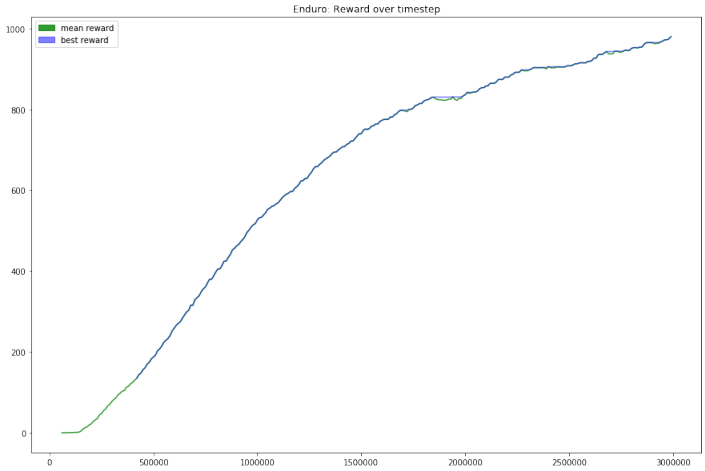


Figure 4: Enduro reward. The mean reward is computed over 100 episodes. Default hyper-parameters were used except target network update frequency which was 1000. The best reward is better than the results published in Openai website. Yet it is still far from the maximum reward which is 5000. For this game, the algorithm behavior is more stable. As for Breakout, the algorithm would probably achieve better results with more time steps and greater update frequency.