DQN_loss

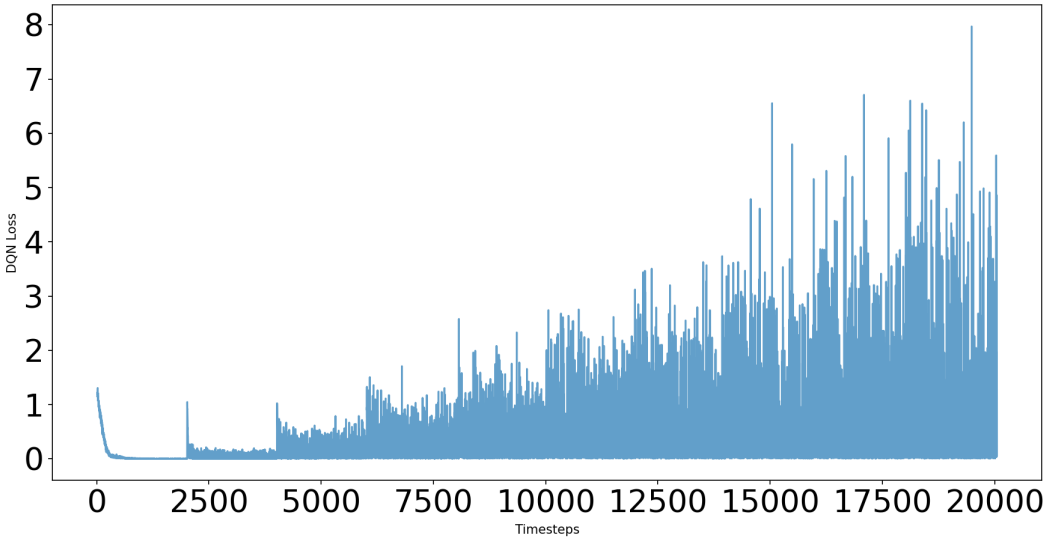

Figure 1: dqn_loss

From Figure 1 we can see that the loss of dqn keeps a linear growth. Every 2000 timesteps, there will be a spike. The trend of the loss of dqn is different from that in supervised learning. The typical supervised learning neural network is designed to minimise the loss between the model and the training data, while the dqn is designed to minimise the loss between the current neural network and the target network. If the variable target_network_update_frequencyis set to 2000 as shown in the figure, the loss will decrease in this period. When the target network is updated, the gradient in current network will be initialized as zero and the loss between current network and target network will be larger than before. Also because of the hard update, the loss will get big immediately and the spike appears.