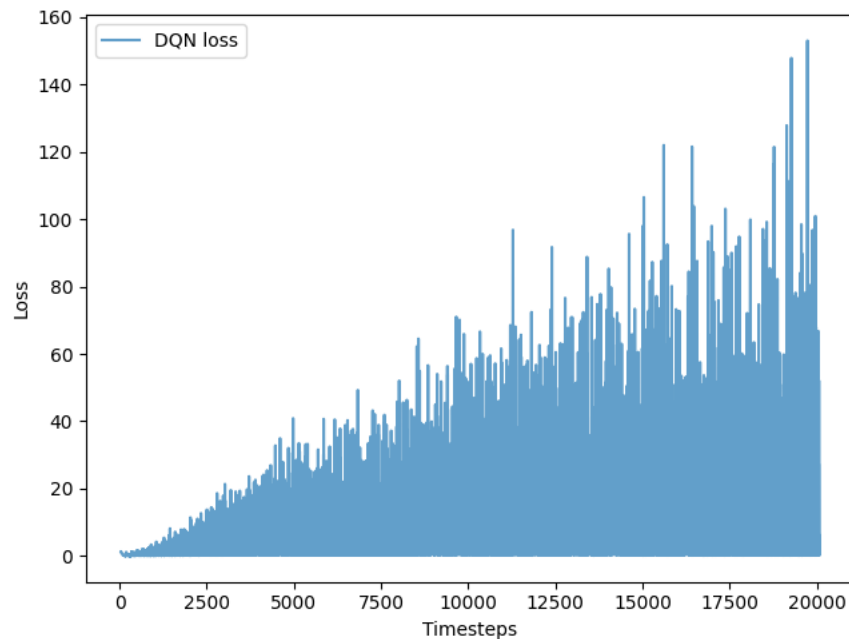


Understanding the Loss



Different from supervised learning, in the training process of DQN, loss does not measure the performance, it is just used to update parameters.

While supervised learning training, the target value of calculating loss is fixed, so loss will decrease smoothly. But the target value in DQN will change during the training process. When *update counter* is equal to *target update frequency*, the parameters of the target network (`critics_target`) update based on the value network (`critics_net`), causing Q-learning target to change. It will cause the loss (MSE between Q-network and Q-learning targets) to change suddenly, and this is the reason for the difference, also for the spikes.