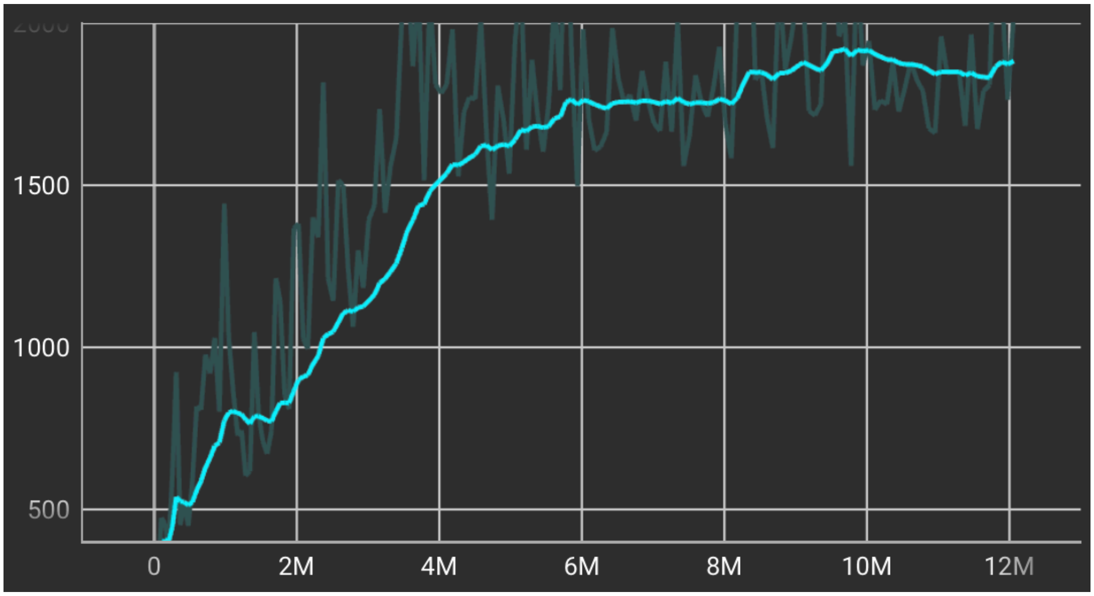
RL Lab2 Report

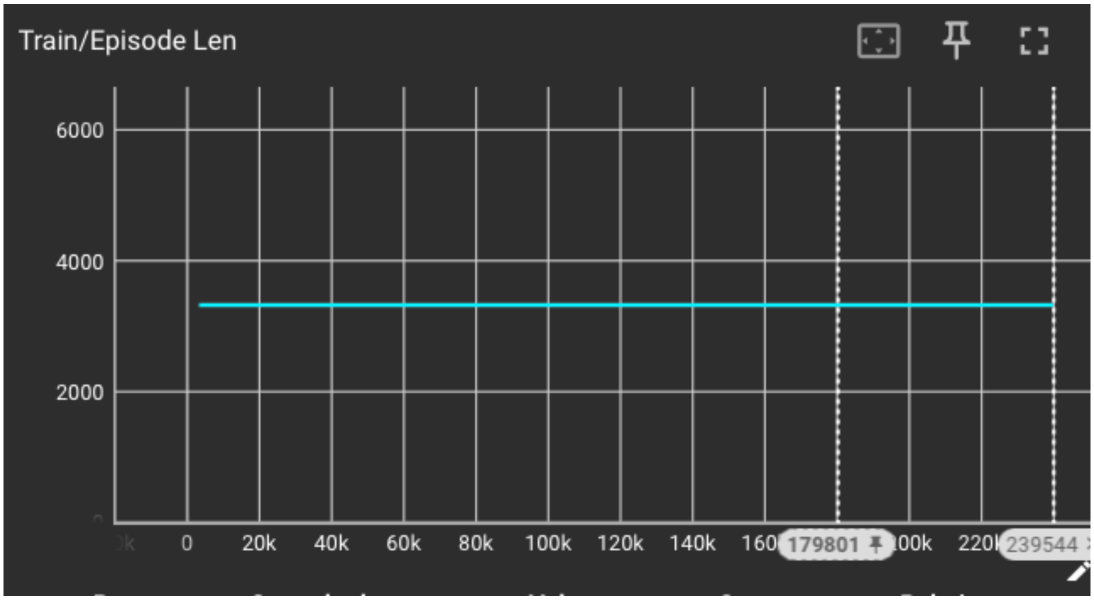
Name : 林伯偉　　Student ID : 109612019

- **Experiment Results (30%)**



- **Bonus**

1. **Screenshot of Tensorboard training curve and testing results on Enduro-v5 (10%).**



2. **Screenshot of Tensorboard training curve and testing results on DDQN, and discuss the difference between DQN and DDQN (3%)**

Target Q-Network:

DQN: DQN uses a single Q-network for both selecting actions and estimating their values. The Q-network is trained to minimize the temporal difference (TD) error between the predicted Q-values and the target Q-values.

DDQN: DDQN introduces the concept of a target Q-network to address overestimation bias in Q-value estimates. It maintains two separate Q-networks – the online Q-network (used for action selection) and the target Q-network (used to estimate Q-values for target computation).

Update:

DQN: DQN uses a simple Q-learning update rule, where the Q-network is updated by minimizing the TD error between the predicted Q-values and the target Q-values.

DDQN: DDQN employs a more sophisticated double Q-learning update rule. It uses the online Q-network to select the best action and the target Q-network to estimate the Q-value of that action, resulting in a more accurate estimate of the Q-value.

3. **Screenshot of Tensorboard training curve and testing results on Dueling DQN, and discuss the difference between DQN and Dueling DQN (3%).**

Network architecture:

DQN: DQN uses a simple Q-learning update rule, where the Q-network is updated by minimizing the TD error between the predicted Q-values and the target Q-values.

DDQN: DDQN employs a more sophisticated double Q-learning update rule. It uses the online Q-network to select the best action and the target Q-network to estimate the Q-value of that action, resulting in a more accurate estimate of the Q-value.

Value function:

DQN: DQN estimates the Q-values directly, where each output of the network corresponds to the Q-value of a specific action.
Dueling DQN: Dueling DQN decomposes the Q-value into the sum of a state value and the advantages of each action. The state value represents the value of being in a particular state regardless of the action taken, and the advantages represent how each action deviates from the state value.

4. **Screenshot of Tensorboard training curve and testing results on DQN with parallelized rollout, and discuss the difference between DQN and DQN with parallelized rollout (4%).**

The primary difference between DQN and DQN with parallelized rollout is that DQN with parallelized rollout uses parallelization to collect experiences simultaneously from multiple instances of the environment. Instead of waiting for one instance to complete an episode before starting the next, several instances run in parallel, generating more experiences in a given time frame. The parallelized approach aims to make more efficient use of computational resources, accelerate data collection, and potentially improve exploration during the learning process.