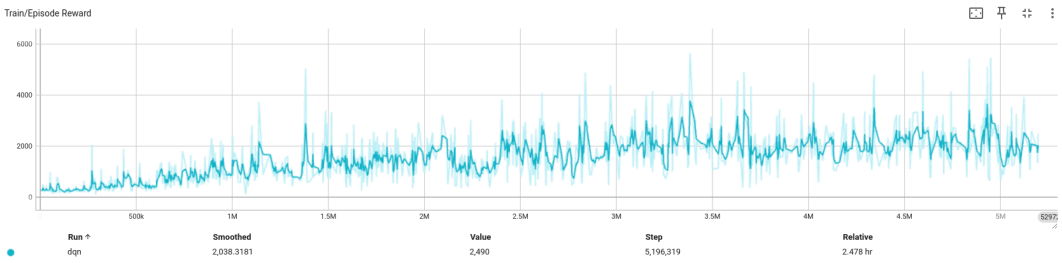
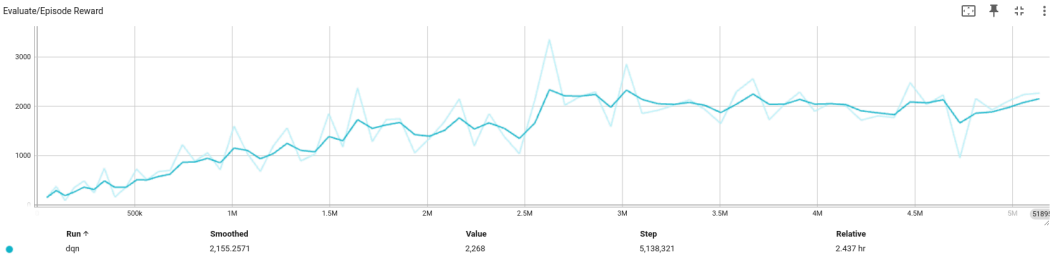
Lab2: DQN, DDQN, and DQN with parallelized rollout

Student Name: 歐庭維
Student ID: 312605001

- **Screenshot of Tensorboard training curve and testing results on DQN**
  - Training curve



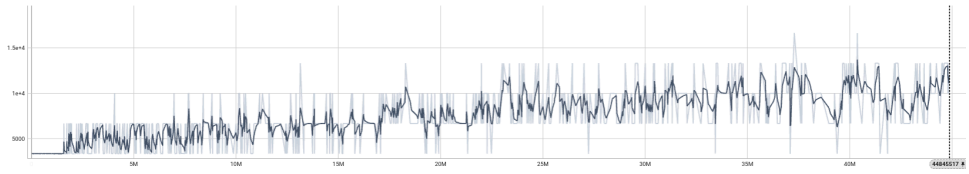
  - Evaluation curve



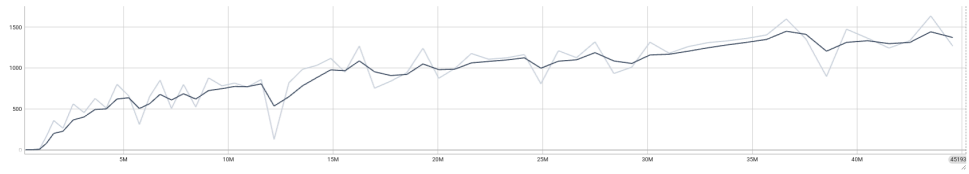
  - Testing results(5 games)

```
episode 1 reward: 3930.0
episode 2 reward: 4440.0
episode 3 reward: 4510.0
episode 4 reward: 4050.0
episode 5 reward: 4040.0
average score: 4194.0
```

- Bonus
  - **Screenshot of Tensorboard training curve and testing results on Enduro-v5 using DQN**
    - Training curve



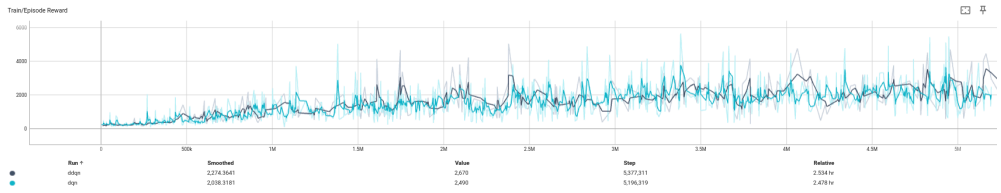
    - Evaluation curve



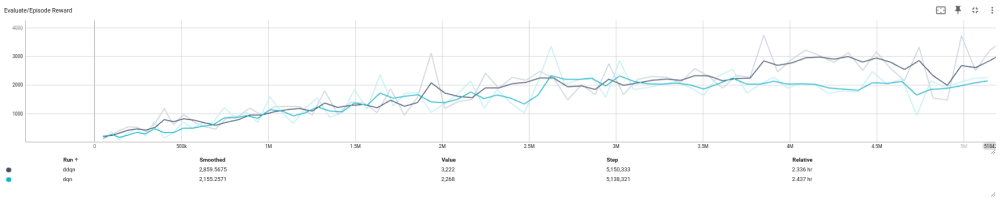
    - Testing results(5 games)

```
episode 1 reward: 1087.0
episode 2 reward: 1373.0
episode 3 reward: 1348.0
episode 4 reward: 1087.0
episode 5 reward: 1092.0
average score: 1197.4
```

- **Screenshot of Tensorboard training curve and testing results on DDQN, and discuss the difference between DQN and DDQN.**
  - Training curve



  - Evaluation curve



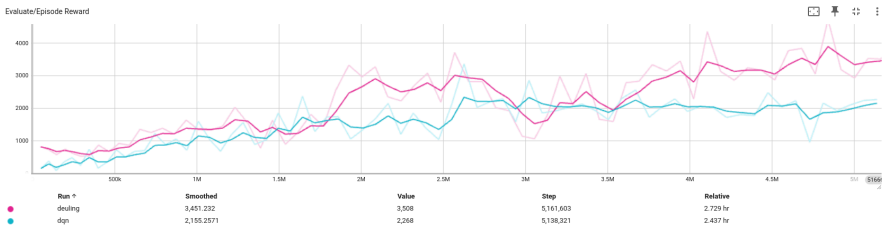
  - Testing results(5 games)

```
episode 1 reward: 3120.0
episode 2 reward: 2850.0
episode 3 reward: 1820.0
episode 4 reward: 2190.0
episode 5 reward: 2690.0
average score: 2534.0
```

  - DQN and DDQN are both used to address the Q-learning problem, with the key difference being how they handle the overestimation of Q-values. DQN employs a deep neural network to approximate the Q-function and uses a target network to stabilize the training process. However, DQN tends to overestimate Q-values. To tackle this issue, DDQN introduces a double network mechanism. In DDQN, a behavior network selects the action, while a target network is used to estimate the Q-value, effectively reducing overestimation.

- **Screenshot of Tensorboard training curve and testing results on Dueling DQN, and discuss the difference between DQN and Dueling DQN.**
  - Training curve

  - Evaluation curve



- Testing results(5 games)

```
episode 1 reward: 4770.0
episode 2 reward: 4770.0
episode 3 reward: 4940.0
episode 4 reward: 4250.0
episode 5 reward: 5080.0
average score: 4762.0
```

- In many states, learning the effect of each action is unnecessary because the actions do not significantly impact the environment. Unlike DQN, Dueling DQN separates the learning process into two streams: the **value stream** and the **advantage stream**. The value stream estimates the value of being in a particular state, reflecting how good or bad the state is, regardless of the action taken. On the other hand, the advantage stream estimates the relative benefit of taking a specific action in that state. By combining these two streams, Dueling DQN allows for more efficient learning by focusing on the state's value and the advantage of different actions within that state, addressing situations where certain actions may have little to no impact on the overall outcome. In conclusion, the Dueling DQN has a smaller variance, which can help improve training performance.