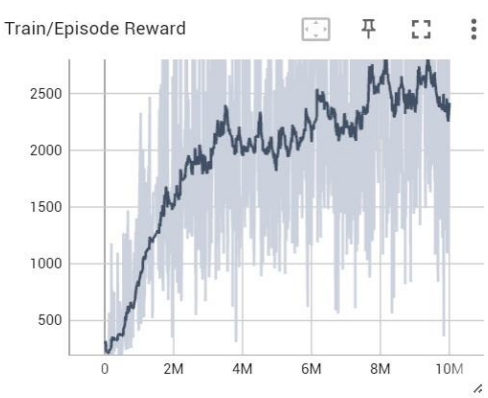
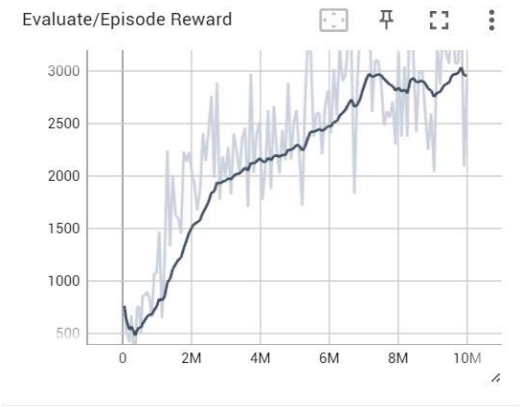
RL lab2

Name: 陳以瑄 Student ID:109705001

- **DQN. Tensorboard training curve**

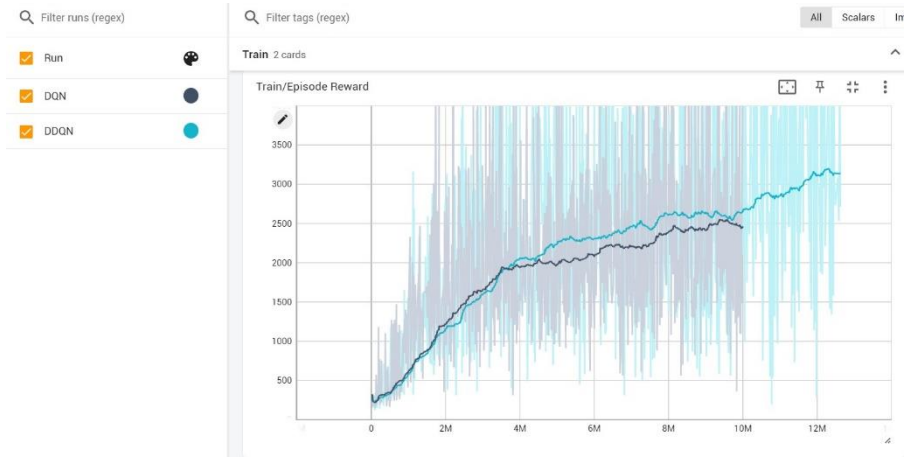


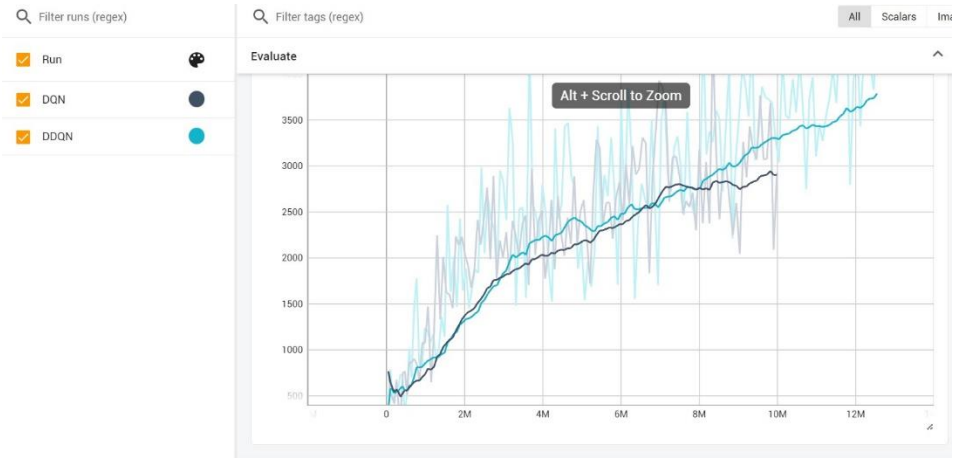


- **DQN. Testing result**

```
episode 1 reward: 2530.0
episode 2 reward: 3060.0
episode 3 reward: 4380.0
episode 4 reward: 2790.0
episode 5 reward: 3290.0
average score: 3210.0
```

- **DDQN. Tensorboard training curve**





- **DDQN. Testing result**

```
episode 1 reward: 4250.0
episode 2 reward: 5090.0
episode 3 reward: 4480.0
episode 4 reward: 4870.0
episode 5 reward: 4620.0
average score: 4662.0
```

- **Discussion of the difference between DQN and DDQN**

  As the plot shows, the performance of DDQN is slightly higher than DQN. The difference between these two algorithms is that in DQN we calculate the sum of the reward and the maximum Q value based on the target network to evaluate the Target Q value $Y_t^Q = r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a'|\theta^-)$, but this will result in overestimation of Q-Values. While in DDQN, we first choose the action that has the maximum Q value based on the behavior network, then we calculate the sum of reward and the Q value of this action based on the target network $Y_t^{DoubleQ} = r_{t+1} + \gamma Q\left(S_{t+1}, \operatorname{argmax}_a Q(S_{t+1}, a|\theta)|\theta^-\right)$.