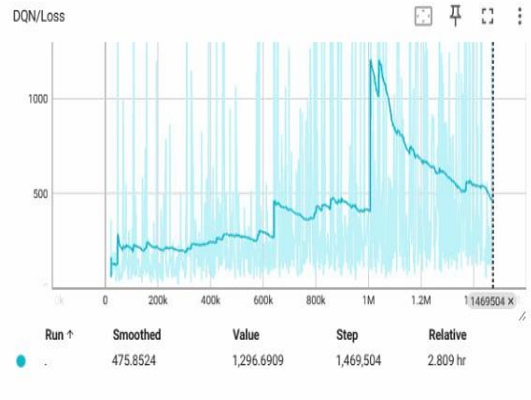
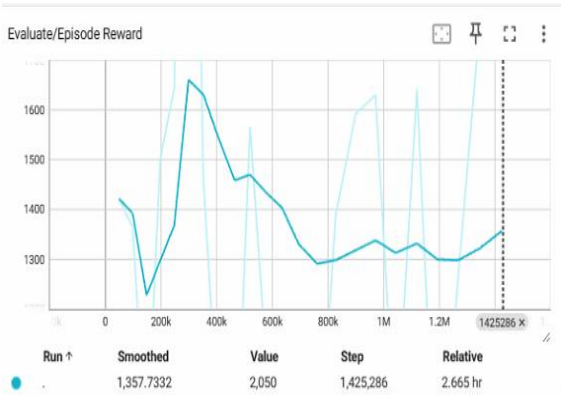
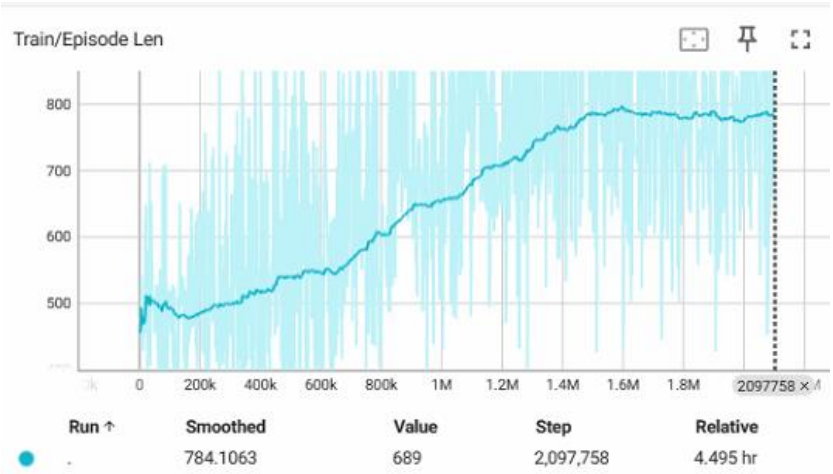
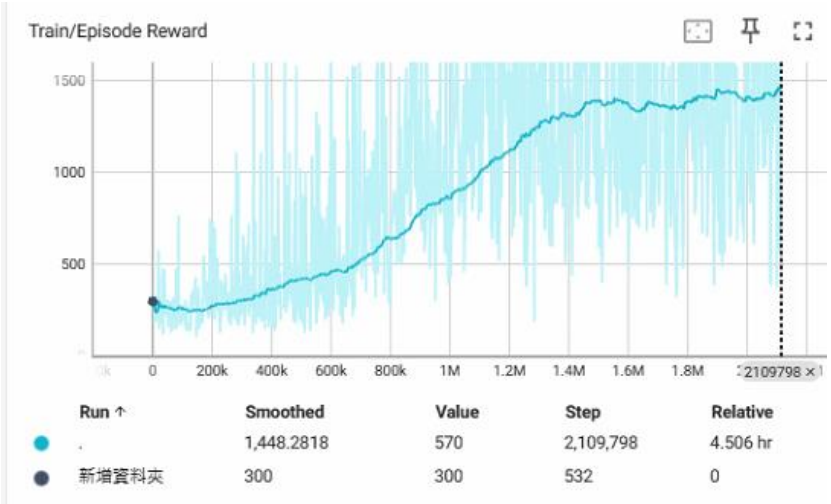
RESULT



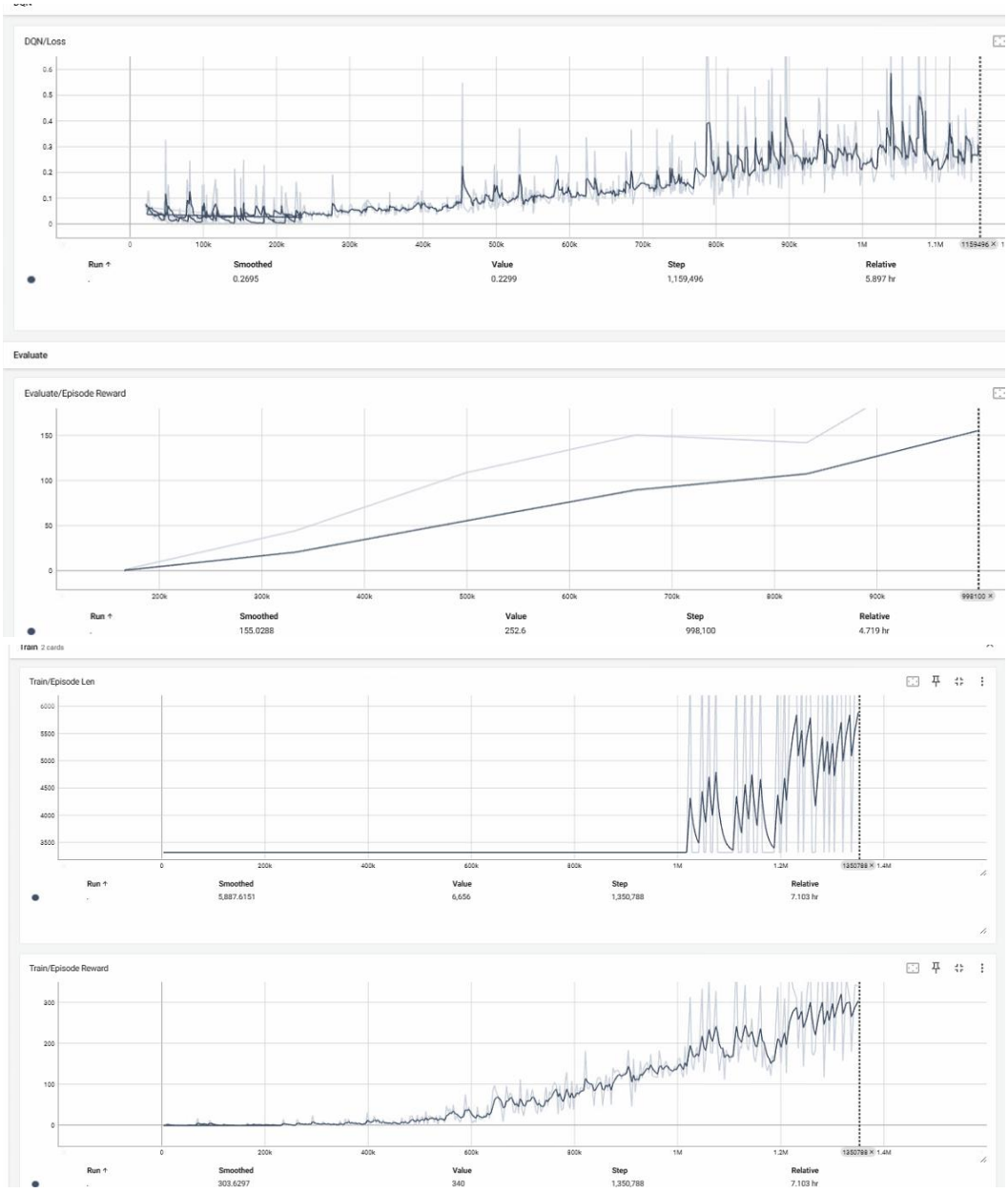


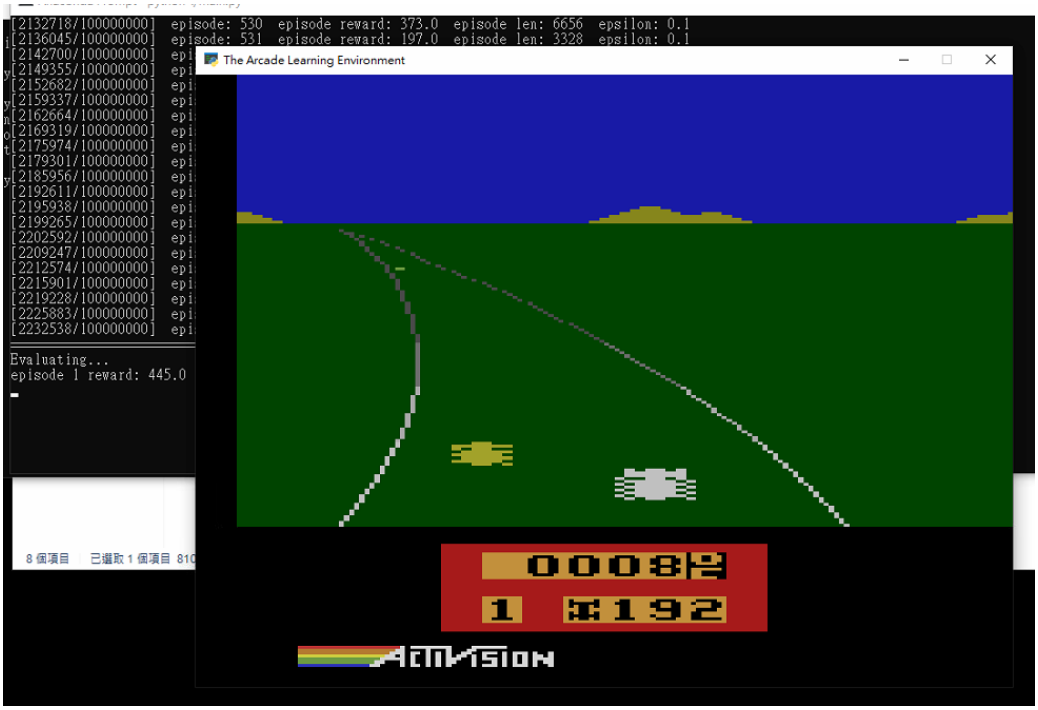




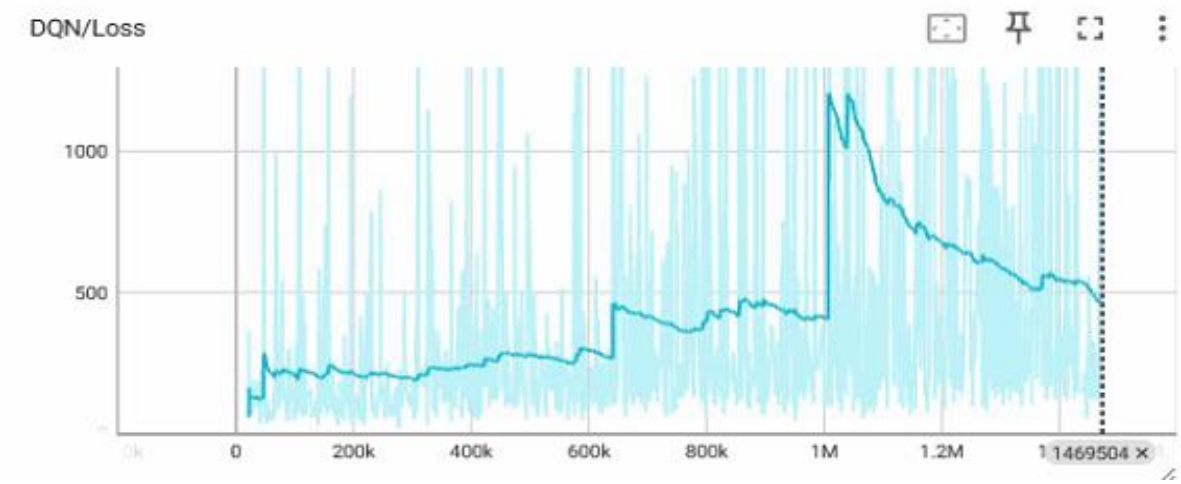
BONUS

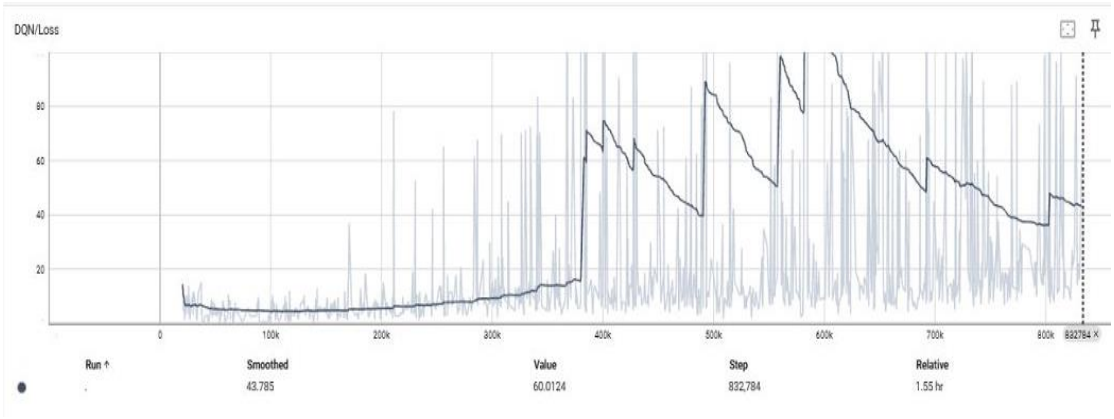
1. Screenshot of Tensorboard training curve and testing results on Enduro-v5 (10%)



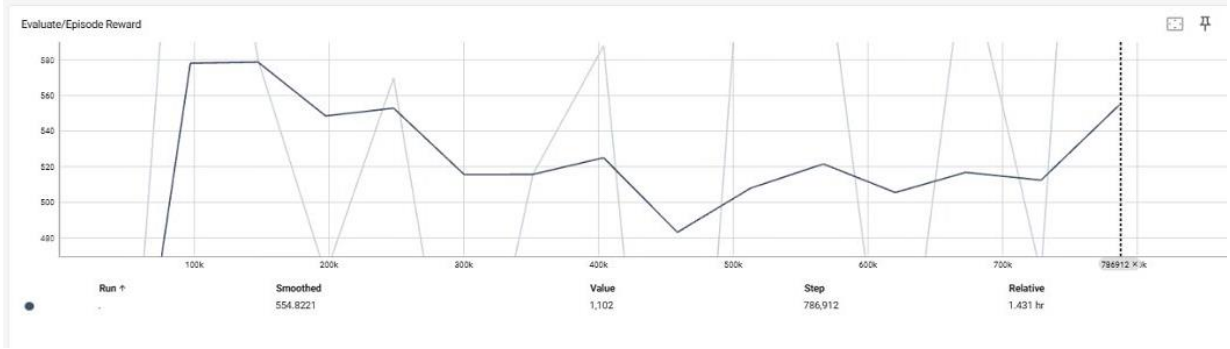


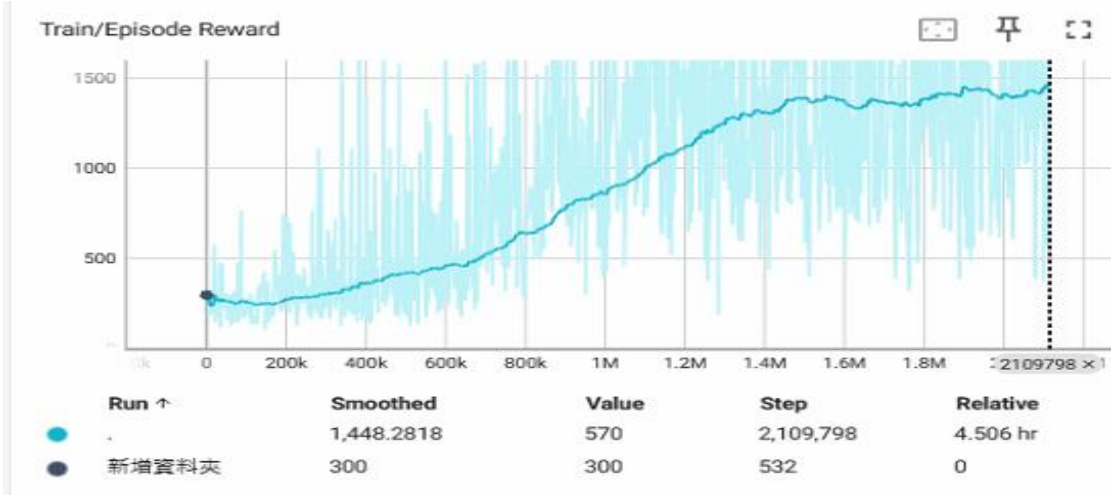
2. Screenshot of Tensorboard training curve and testing results on DDQN, and discuss the difference between DQN and DDQN (3%)
   - 藍色現為 DQN，黑色線則為 DDQN
   - 最為明顯的地方是 DDQN 的 loss 明顯低於 DQN，這主要是因為 DDQN 的設計就是為了減小 Q-NETWORK 過度高估的問題。
   - 收斂速度以及最後 performance 則看起來都差不多
   - DDQN 主要改變是再算 target value 時所用到 next value 不是直接由 target network 計算，而是先由 behavior network 算出對應 action 再用此 action 從 target network 尋找對應的值

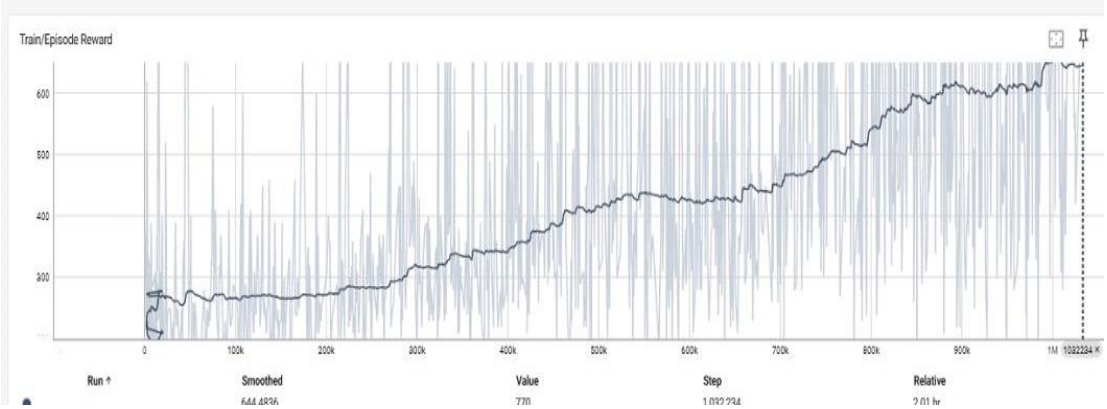



DQN/Loss
80
60
40
20
0
100k
200k
300k
400k
500k
600k
700k
800k
832784
Run
Smoothed
Value
Step
Relative
43.785
60.0124
832,784
1.55 hr

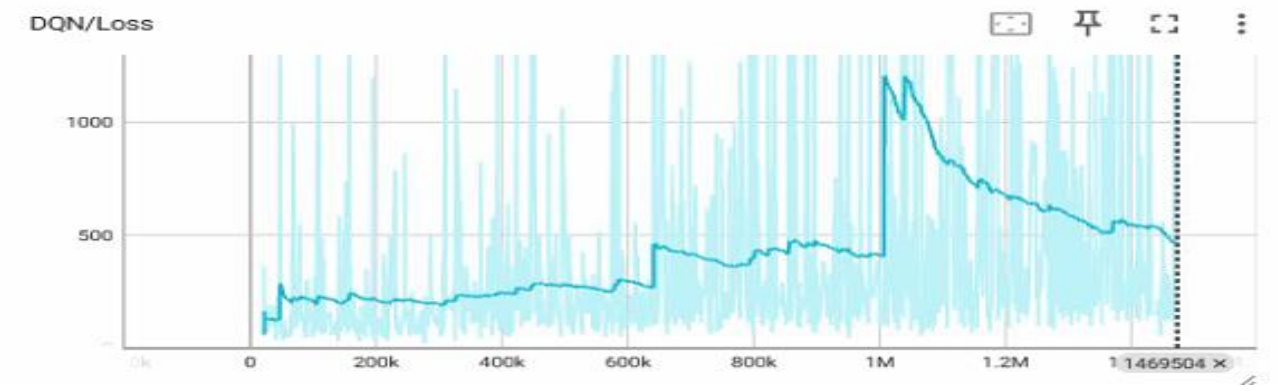

Evaluate/Episode Reward
1600
1500
1400
1300
0
200k
400k
600k
800k
1M
1.2M
1425286
Run
Smoothed
Value
Step
Relative
1,357.7332
2,050
1,425,286
2.665 hr

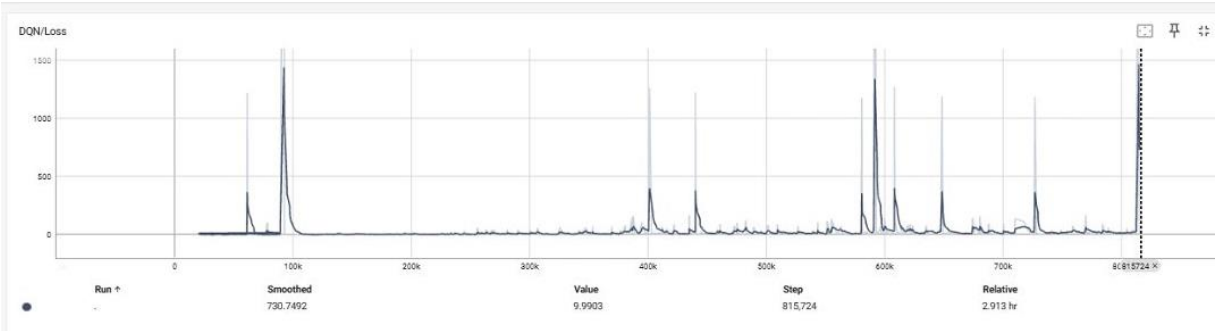

Evaluate/Episode Reward
580
560
540
520
500
480
100k
200k
300k
400k
500k
600k
700k
786912
Run
Smoothed
Value
Step
Relative
554.8221
1,102
786,912
1.431 hr

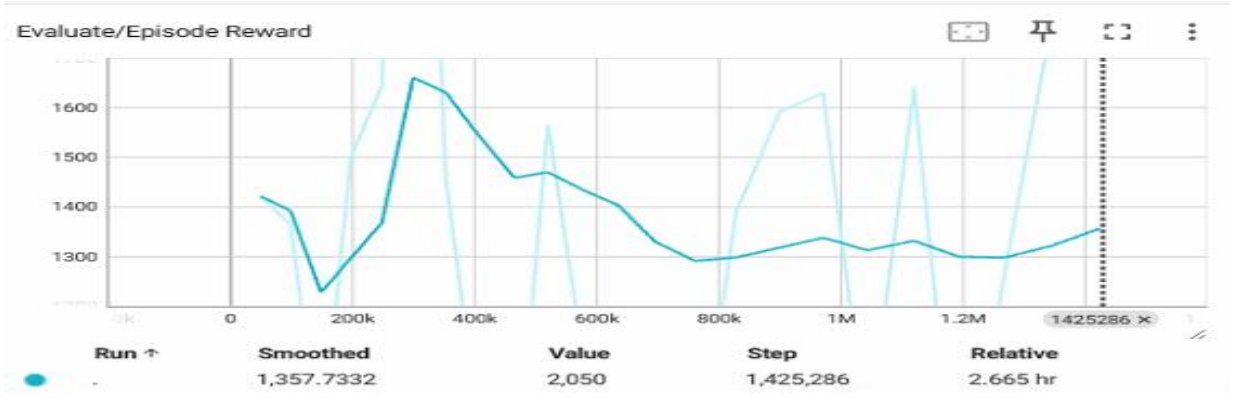

Train/Episode Reward
1500
1000
500
0
200k
400k
600k
800k
1M
1.2M
1.4M
1.6M
1.8M
2109798
Run
Smoothed
Value
Step
Relative
1,448.2818
570
2,109,798
4.506 hr
新增資料夾
300
300
532
0

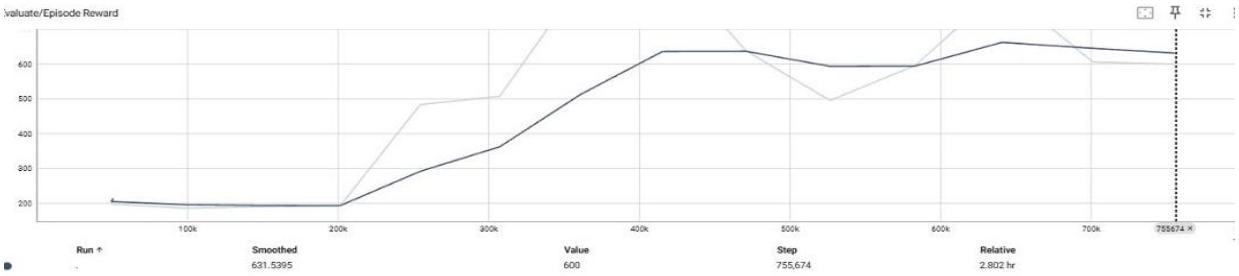


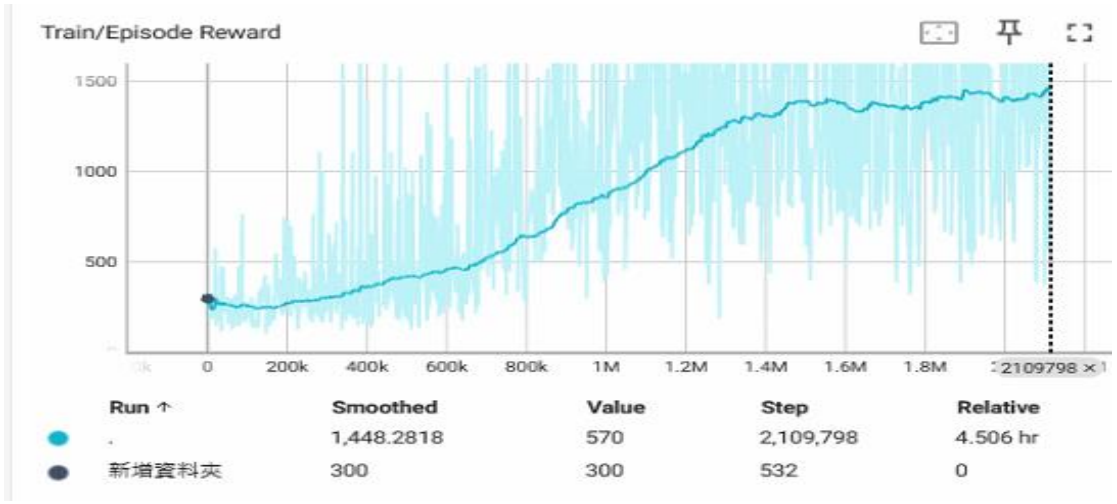
3. Screenshot of Tensorboard training curve and testing results on Dueling DQN, and discuss the difference between DQN and Dueling DQN (3%)
   - 藍色現為 DQN，黑色線則為 Dueling-DQN
   - 最為明顯的地方是 Dueling-DQN 的 loss 低於 DQN，這主要是因為 DDQN 的設計就是為了減小 Q-NETWORK 過度高估的問題。其採用 advantage value 來計算 q-value，英此更為穩定
   - 收斂速度以及最後 performance 則看起來都差不多
   - Dueling DQN 與 DQN 差別是在 Dueling DQN 的 network 是先算出 state value 以及 advantage 再用兩者來算出 q value，如此可以確保 q value 更為穩定

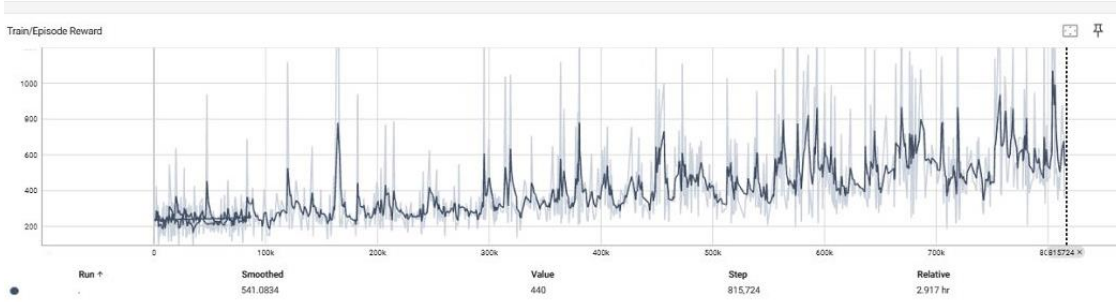












4. Screenshot of Tensorboard training curve and testing results on DQN with parallelized rollout, and discuss the difference between DQN and DQN with parallelized rollout (4%).