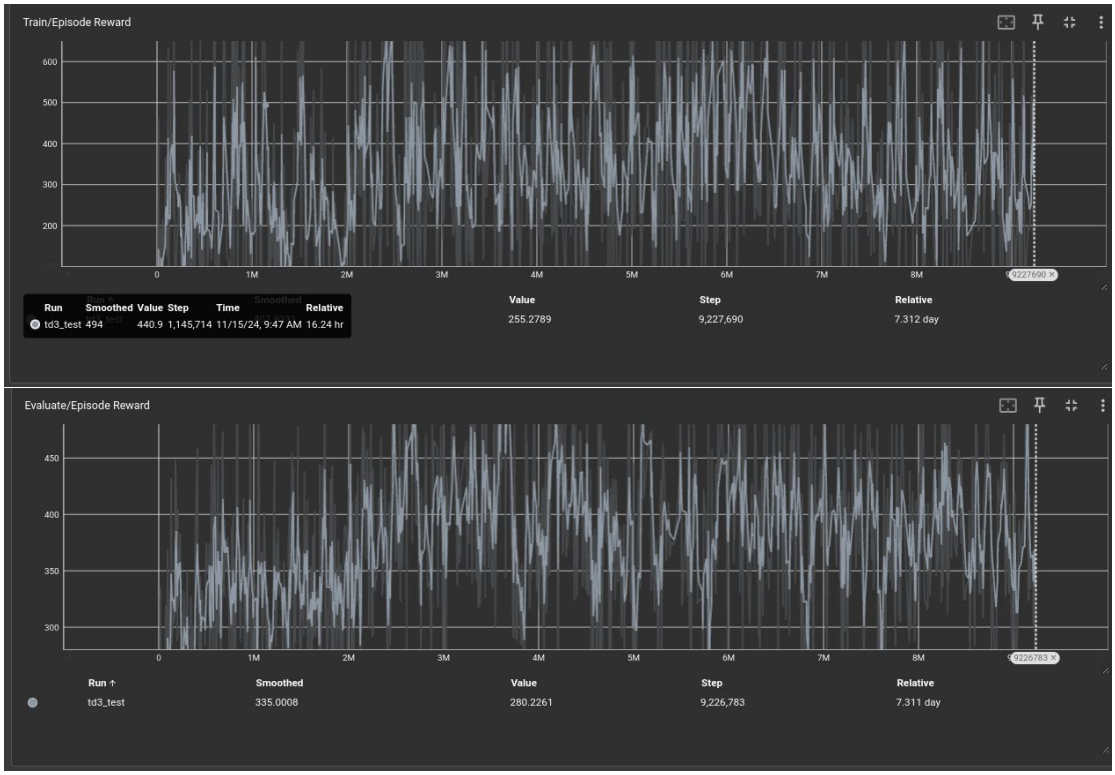
(1) Screenshot of Tensorboard training curve and testing results on TD3.

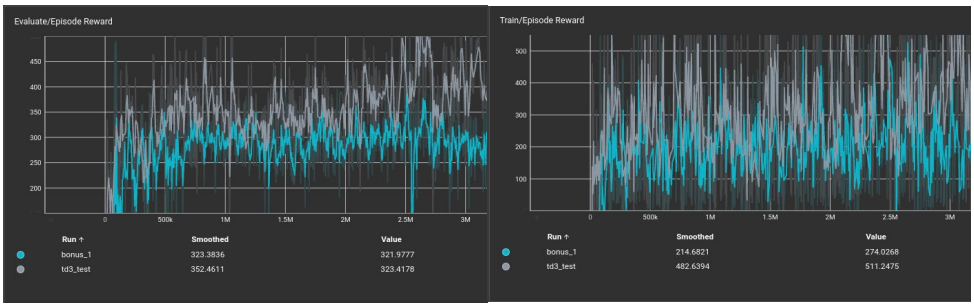
```
Evaluating...
Episode: 1      Length: 216     Total reward: 252.15
Episode: 2      Length: 281     Total reward: 390.17
Episode: 3      Length: 345     Total reward: 471.69
Episode: 4      Length: 596     Total reward: 929.92
Episode: 5      Length: 889     Total reward: 911.00
Episode: 6      Length: 246     Total reward: 332.44
Episode: 7      Length: 999     Total reward: 865.28
Episode: 8      Length: 147     Total reward: 168.79
Episode: 9      Length: 516     Total reward: 948.30
Episode: 10     Length: 255     Total reward: 300.49
average score: 557.0223527565976
```



Bounus:

灰色的曲線是標準 TD3 的實作結果，其他不同顏色的曲線代表相對應 bonus 的實作結果

(1) Screenshot of Tensorboard training curve and compare the performance of using twin Q-networks and single Q-networks in TD3, and explain

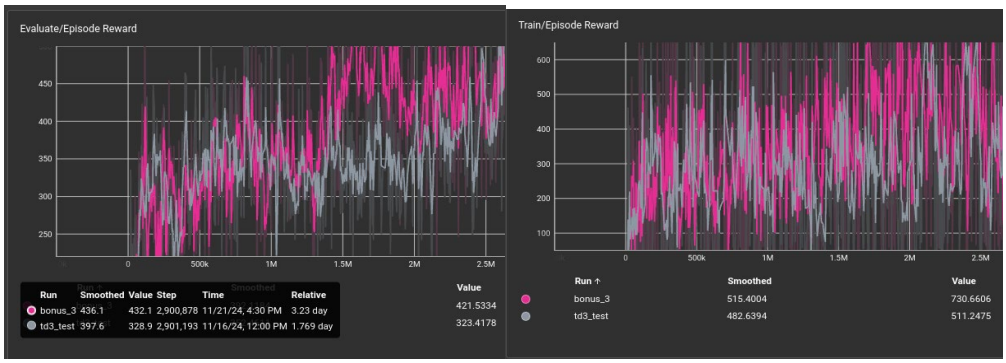


Twin Q network 表現的比只用一個 Q network 的還要好，原因是因為用兩個 Q network 可以讓系統在訓練的過程中比較穩定，也可以減少策略過早收斂到次優解的風險

(2) Screenshot of Tensorboard training curve and compare the impact of enabling and disabling target policy smoothing in TD3, and explain (5%).

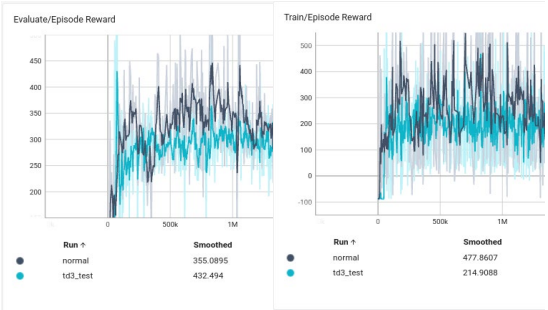
可以看見沒有 noise 的情況 reward 反而比較高，但是也可以觀察到分數的波動也比較明顯，可能的原因為，這個系統中的 noise 可能不適合這個遊戲，因為剪裁的範圍過窄或是更新 actor_net 的頻率不對，所以造成結果反而比較不好。

(3) Screenshot of Tensorboard training curve and compare the impact of delayed update steps and compare the results, and explain (5%).



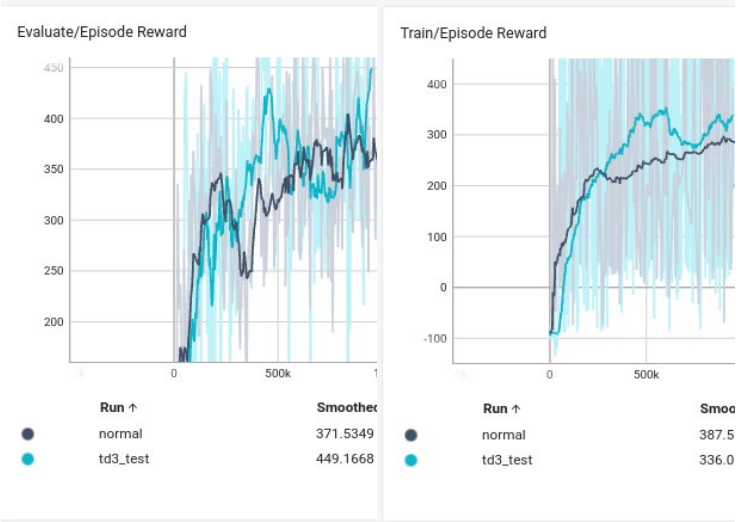
可以看見沒有用延遲更新的曲線會比較不穩定，且最後的結果也沒有比較好，在其他的情況下，可能還會 collapse

(4) Screenshot of Tensorboard training curve and compare the effects of adding different levels of action noise (exploration noise) in TD3, and explain (5%).



調整 noise 的 sigma，固定為 0.001，可以發現這個設置並不適合這個系統，造成 training 的結果比原本的還要差。

(5) Screenshot of Tensorboard training curve and compare your reward function with the original one and explain why your reward function works better



把分數為正的增加權重，分數為負的減少權重，來讓整個系統專注在學習好的 actor，可以看見分數較原本的高一些。