

1. Implement Advanced RL algorithm

a) Advanced RL algorithm : DQN

b) The difference between DQN and Policy Gradient : DQN 屬於 Value-based 演算法，透過學習並找出最佳的 Value Function，設法找到最佳的 Policy；而 Policy Gradient 屬於 Policy-based 演算法，透過直接對長期的 Total Reward 進行最佳化，設法找到最佳的 Policy。

c) Implementation : 建構一個有三層 FC-Layers 的 DQN Network，以及 DQN Agent，其中包含 Replay Memory 用來儲存資訊。Agent 在執行 Action 時，會透過生成隨機數並與 epsilon 比較的 Epsilon-greedy action selection 方法，解決 Exploration and Exploitation dilemma 問題，並選擇出一個 Action，進一步執行該 Action 以獲得 Reward 與下一個 State，而後將這些資訊儲存至 Replay Memory 中，再隨機從 Replay Memory 取出 minibatch 進行訓練，最後進行梯度遞減與參數更新流程。

2. (C) It utilize the MCTS algorithm during training