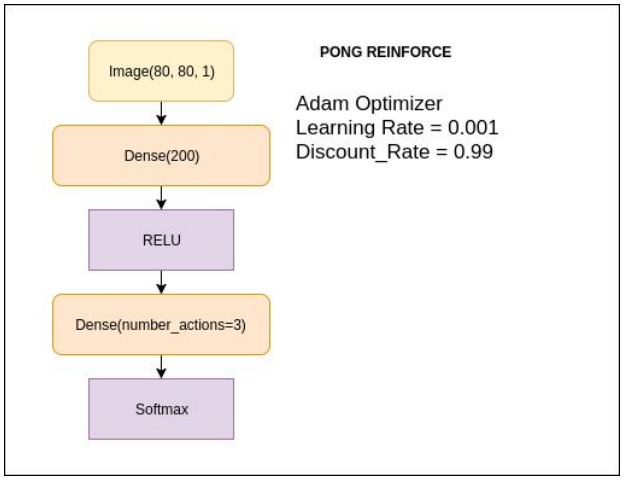
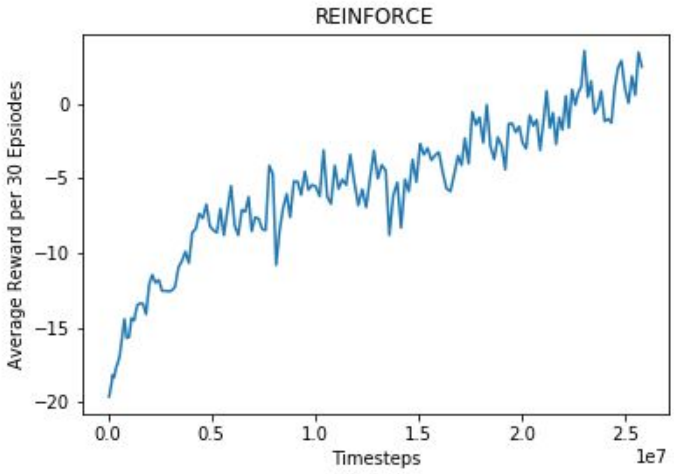
第三作業

Policy Gradient

以下是我的policy gradient模型和參數. 我用的是基本的REINFORCE演算法.

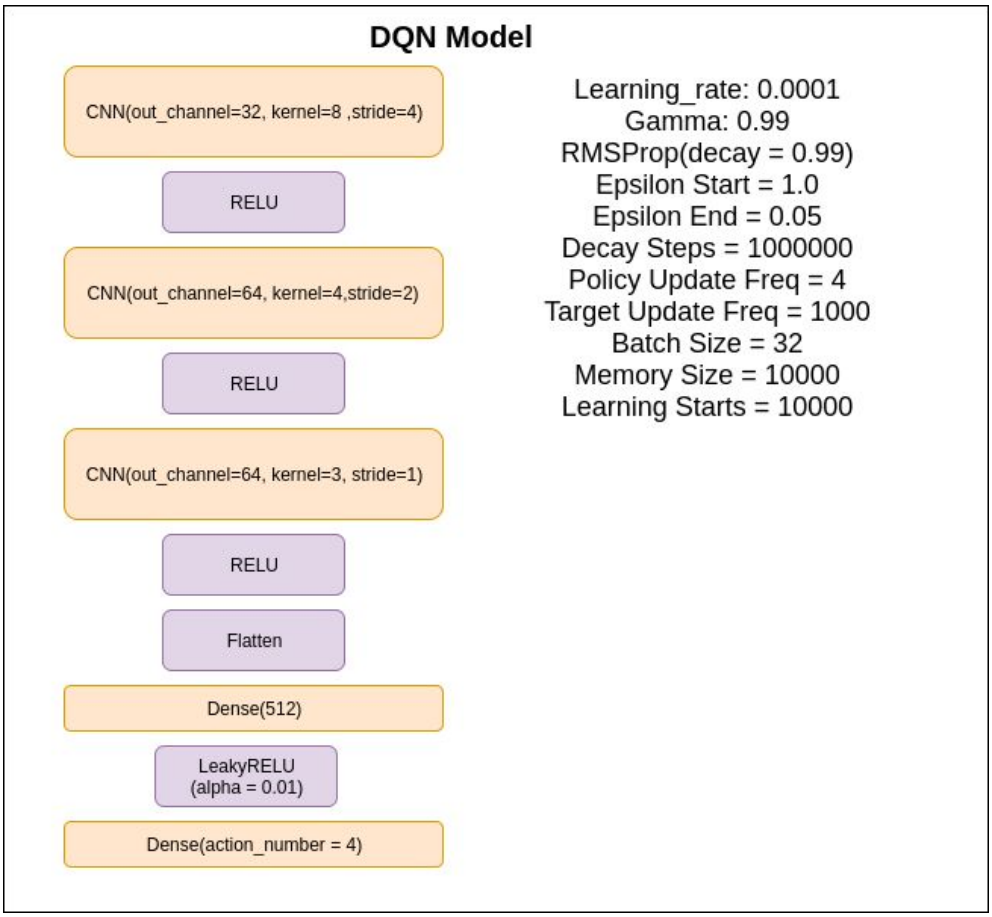


這是學習Pong的學習曲線. 最後測試平均分是2.5

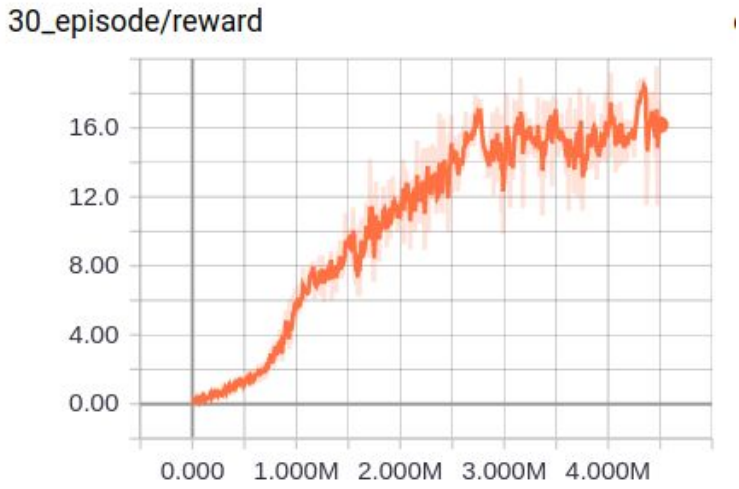


DQN

以下是我的DQN模型和參數.



這是學習Breakout的學習曲線. 最後測試平均分是60-70.



我試了不同的learning rate,不同的network structure，還有一個不同的loss function.
Learning rate: 從0.0001升到0.00025, 主要是知道learning rate 對gradient的影響力很大，尤其有兩個network在freeze和update, 因該就會有影響．結果發現大一點learning rate 反而降低了分數.

Network structure: 改的是最後一個dense的activation function爲relu, 主要是想知道助教爲什麼最後一層要做leaky relu, 所以想跟普通的relu來比較，結果沒有什麼差別

Loss Function: 發現pytorch的 DQN tutorial用的是huber loss, 所以也訓練看看會有什麼結果, 發現效果並沒那麼好．我訓練四百個timestep, 測試平均是30,用MSE算loss反而效果更好.