Reinforcement Learning in Different Games: *Cartpole & Mountain Car Game & Super Mario Bros*

BACKGROUND: Previous studies implemented Deep Q Network (DQN), Q learning, or some policy-based methods in Super Mario Bros . But we found out that there are several new methods proposed recently for reinforcement learning, such as Deep Double Q Network (DDQN), Proximal Policy Optimisation (PPO), which we assume will achieve much better result for Mario Game since the Mario is such a sophisticated game which DQN and Q learning is not capable of handling]. So we decide to take PPO instead of just DQN, and then make a comparison of those two methods.

METHODS

Cartpole & Mountain Car Game

Network structure: 3 linear layers with Relu as its activation function and 1 layer output.

For Cartpole Game: we tried DQN and DDQN. For Mountain Car is Deep Imitation Learning). We also implemented with several techniques mentioned. We use this trained agent to play the games lots of times to calculate the successful rates.

Mario Bros

We use PPO for Mario Playing. We take two things as input, the current frame and the previous actions taken by Mario. Then we make some process to those data according to the figure on the right side to get the final reward and how to do the next step.

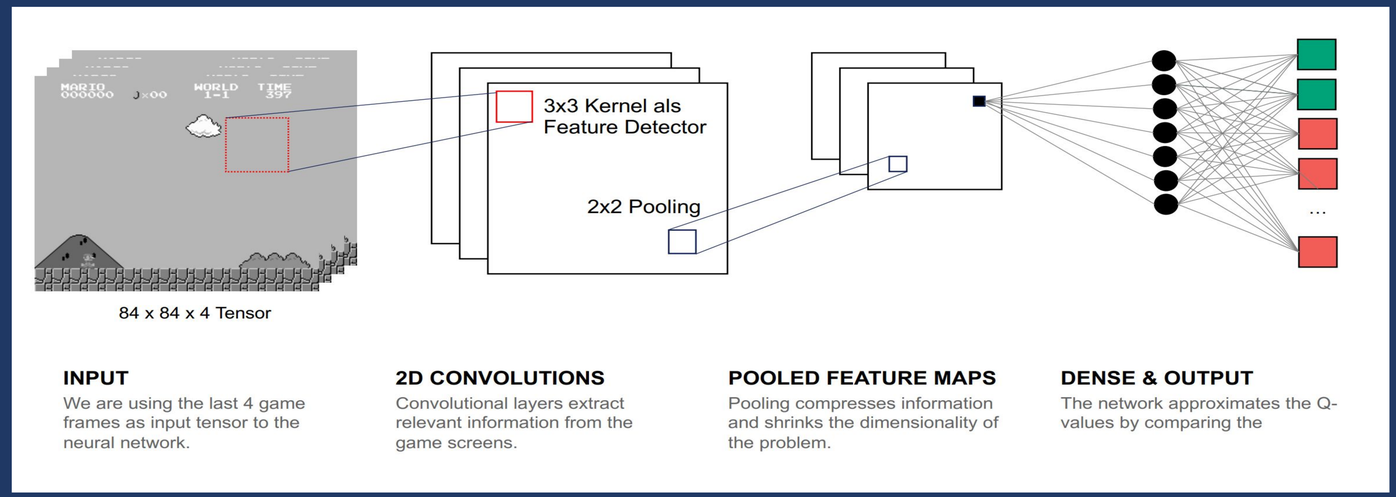


Figure 1: Simple Reinforcement Learning with Convolutional Neural Network

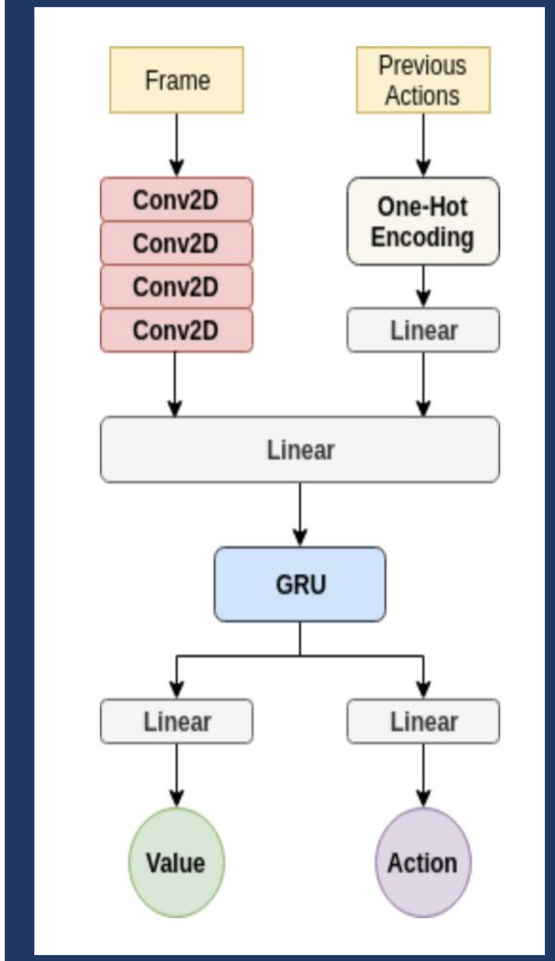
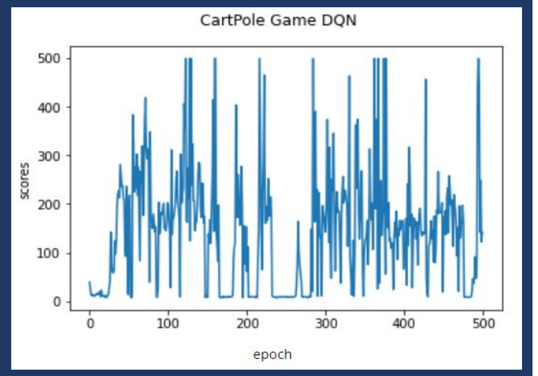


Figure 2: Architecture for PPO

Figure 3: Scores for DQN on Cartpole Game



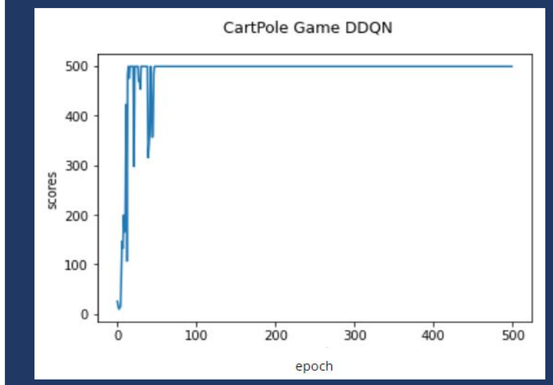
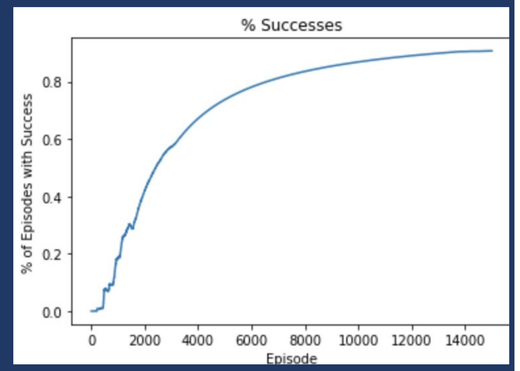


Figure 4: Scores for DDQN on Cartpole Game

Figure 5: Successful Ratio Chart of Mountain Car Game



RESULTS

- Cartpole Game & Mountain Car Game
  Successfully reach 499 seconds of stable time
  Successfully reach the top of mountain
- Mario
  We have successfully beaten every metric compared to DQN, more specific results and comparison metrics are shown below

Cartpole Game		
Model	Score/Stable Time (Max 499)	Success Rate
Q-learning	Max 208	0.37
Deep Q learning	Max 499	0.63
Double Deep Q learning	Max 499	0.95

Mountain Car Game		
Model	Score (Max -2)	Success Rate
Q-learning	-151	0.5
Deep Imitation Learning	-85	0.86
Better Network	-57	0.94

Mario Bros Game			
Model	Passing Rate	Passing time (unit time)	Score
Deep Q Learning	41%	92	700
Proximal Policy	78%	65	1300

Conghao Xiong(cxiong5)
Nan Huo (nhuo1)
Shuxian Zhao (szhao35)
Luchao Qi (lqi9)

JOHNS HOPKINS UNIVERSITY