Firth Project UDACITY

1. THE GOAL

This Project have the goal of build a deep reiforcement learning (DRL). The main idea its the agente navigates on the enviroment and collect the yellow bananas while avoiding purple bananas.

The Rewards of the enviroment its +1 if the agente collects a yellow banana and -1 if colect a purple banana. This environment is solve when during 100 consecutive episodes the escore must get a score of +13.

2. DESCRIPTION OF THE ENVIRONMENT

The task at hand consists of navigating a flat 3D environment (pictured above) with the goal of collecting (stepping over) yellow bananas while avoiding blue bananas.

We train a reinforcement learning agent to solve this task. The agent observers the environment though a 37-dimensional real vector (state space), consisting of the agents' current velocity, along with ray-based perception of objects around agent's forward direction.

The time is divided into turns. An episode ends after 300 turns. At each turn, the agent chooses one of the following actions:

0 - move forward.

1 - move backward.

2 - turn left.

3 - turn right.

3. RL ALGORITHM

The reinforcement learning (RL) framework is characterized by an agente to interact with with the ambience, and learning in the environment contínuos or discrete.

At each time step or episode, the agent receives the environment's state. The environment presents a situation, estate, to the agente. The agent have to choise an appropriate action. After this step, the agent receives a reward indicates whether the agent has responded appropriately to the state and a new state with this informations in a loop.

All agents have the objective, maximize the reward, or the provided sum of rewards sun for all time or steps.

3.1. DEEP Q-NETWORKS

The DQN algoritm started develope to play Atari Games, only use pixels like a human. This tecnics input the images from the environment, producing thestates, and the return the beast action for each action estate of game. This algoritm is a simple neural network of the images are first processed by a couple of convolution layers, allows the system to explored the dates and this convulution layers work with four layears to extrat some temporal properties across those frames. after these layers the dates go to fully conected linear output layer to excute the best action.

3.2. DUELING Q-NETWORK

This algoritm is compose by a sequence of convolution layers, and after is conected by a couple of fully connected layers to produce the Q value. The core idea of dueling networks is use two stream . In the image 2, the V(s) represents the state value and the another layer, representes the advantege values. The Q Value its the sum of this two layers.

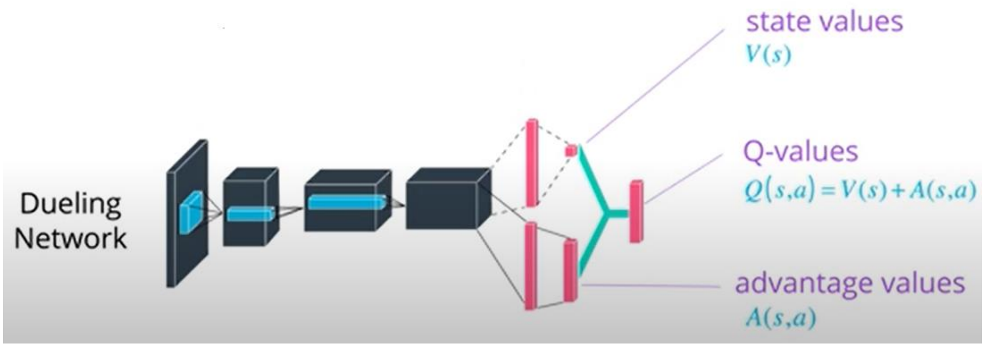
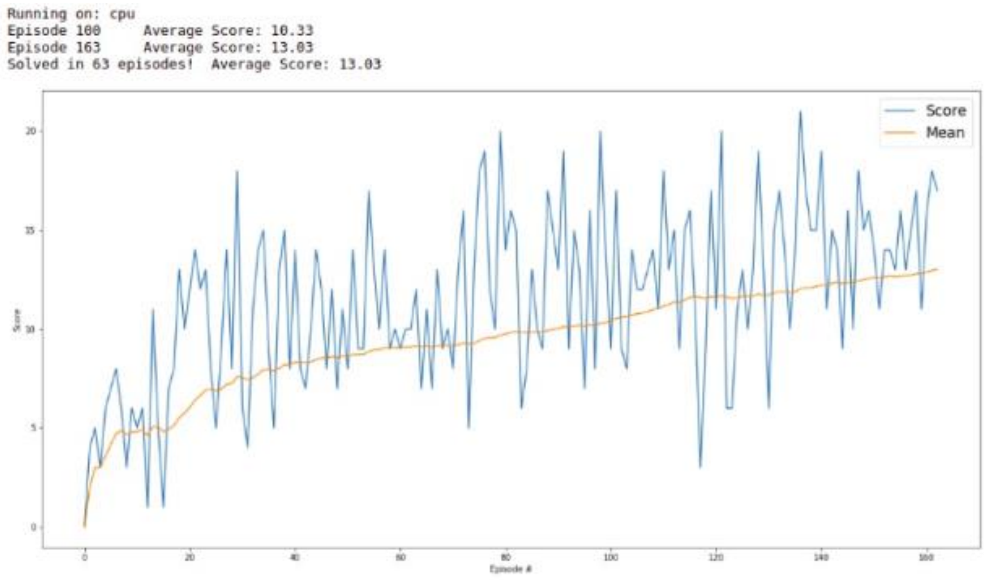


Figure 1. Dueling Network

4. EXPERIMENTS

In Below have a graph with score by step in blue and the averange in Orange.



```
Running on: cpu
Episode 100      Average Score: 10.33
Episode 163      Average Score: 13.03
Solved in 63 episodes!  Average Score: 13.03
```

Figure 2. Training result.