Udacity Deep Reinforcement Learning Nanodegree

Project 2: Continuous control

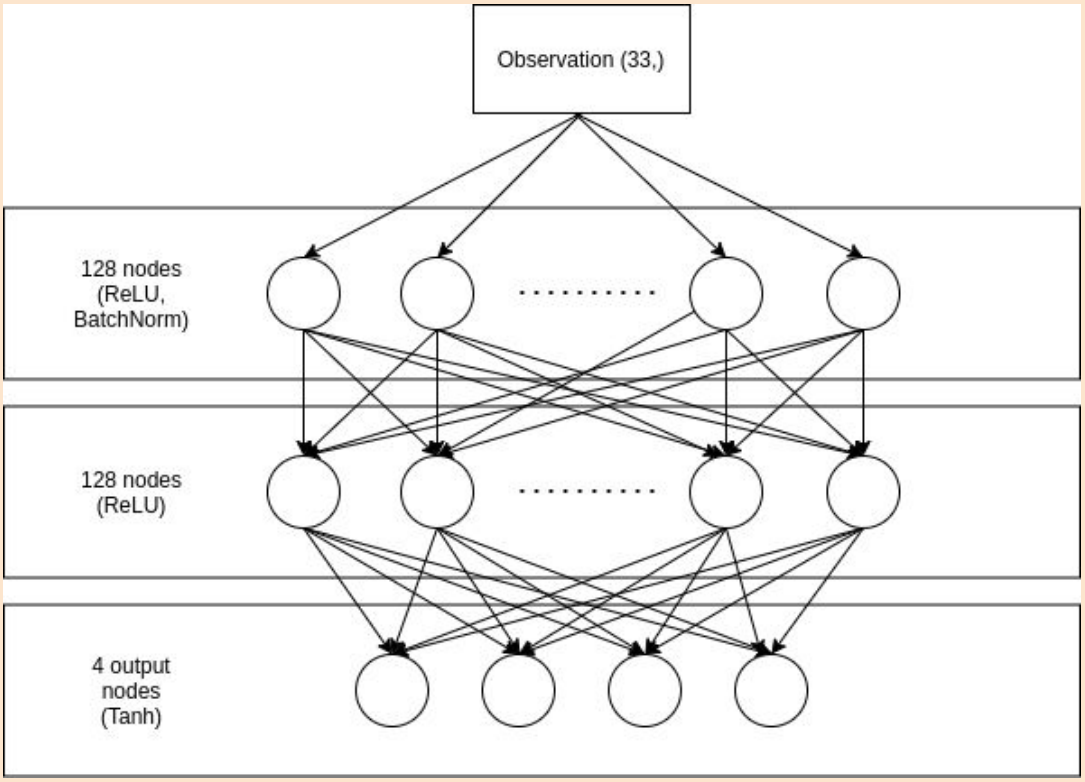
Khanh Nguyen Vu

I. Approach

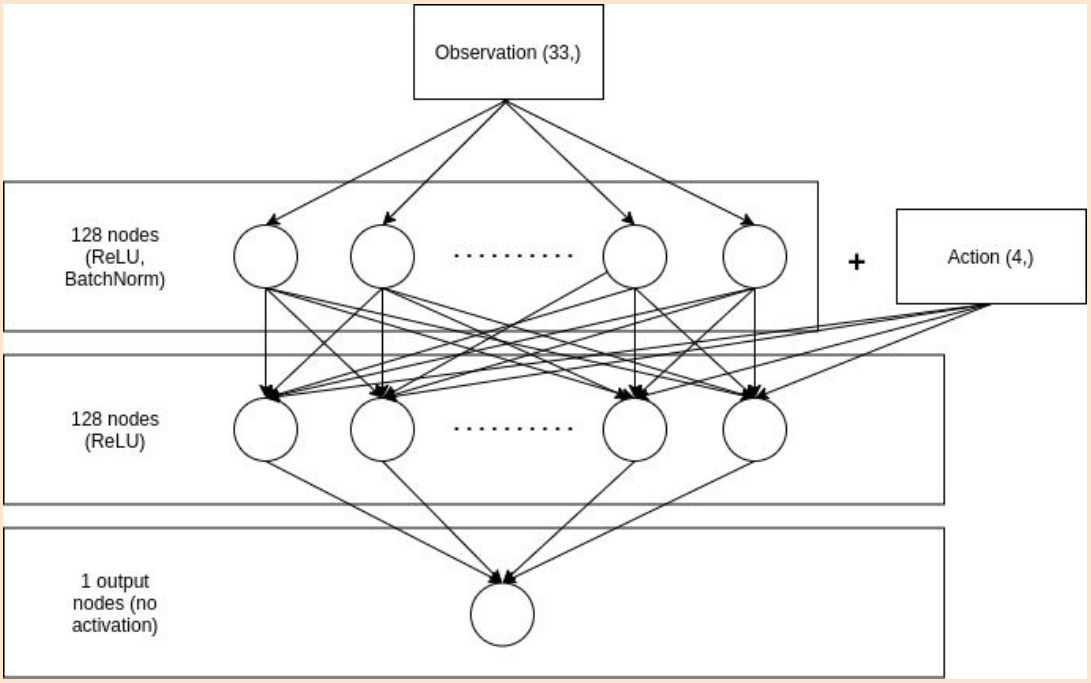
The **Deep Deterministic Policy Gradient (DDPG)** algorithm was adopted for this project. I reused some of the components from the first project (Navigation) since the DDPG is pretty much similar to DQN algorithm (it has a ReplayBuffer and the same networks updating scheme).

Network architecture

Actor network



Critic network



For the exploration factor, Ornstein–Uhlenbeck process was added to the actions vector at every time-step.

Hyperparameters

Networks:

- Actor optimizer: Adam, learning rate = 0.0003.
- Critic optimizer: Adam, learning rate = 0.0003.
- Soft update: TAU = 0.001.

Memory buffer:

- Buffer size: 1000000 (one million).
- Batch size: 64.
- Uniformly sampling.

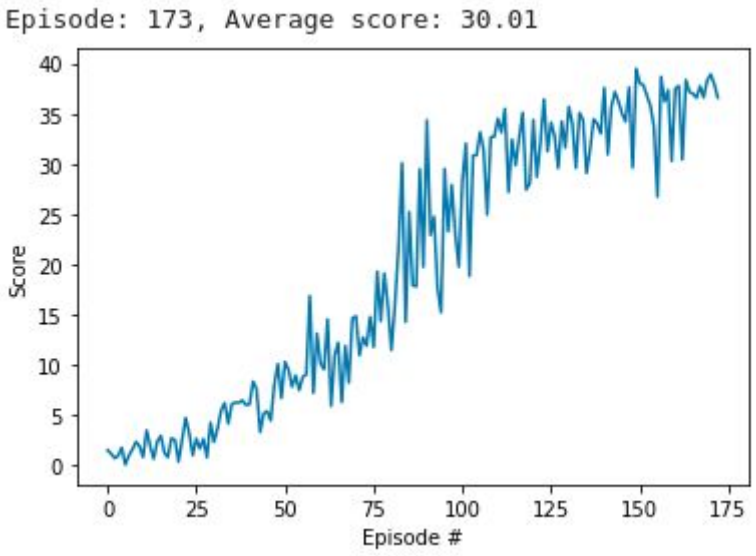
OUNoise:

- Mu: 0
- Theta: 0.1
- Sigma: 0.2
- Sigma_min: 0.1
- Sigma_decay: 0.99 (reduce exploration rate as the agent learns)

Training:

- Max episode: 1000.
- Max steps per episode: 1000.

II. Result



III. Ideas for improvements

- Use prioritized experience replays buffer to improve learning speed.
- Try PPO instead.