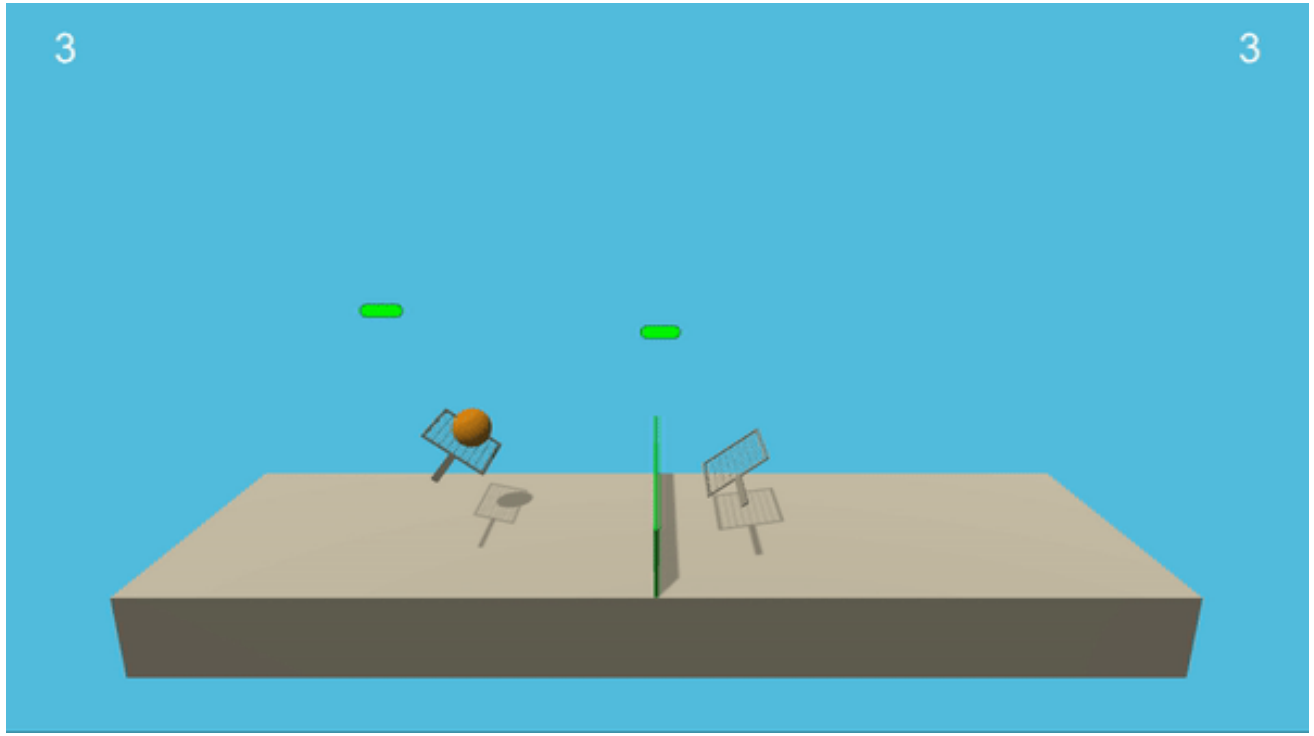
Collaboration & Competition

The Environments

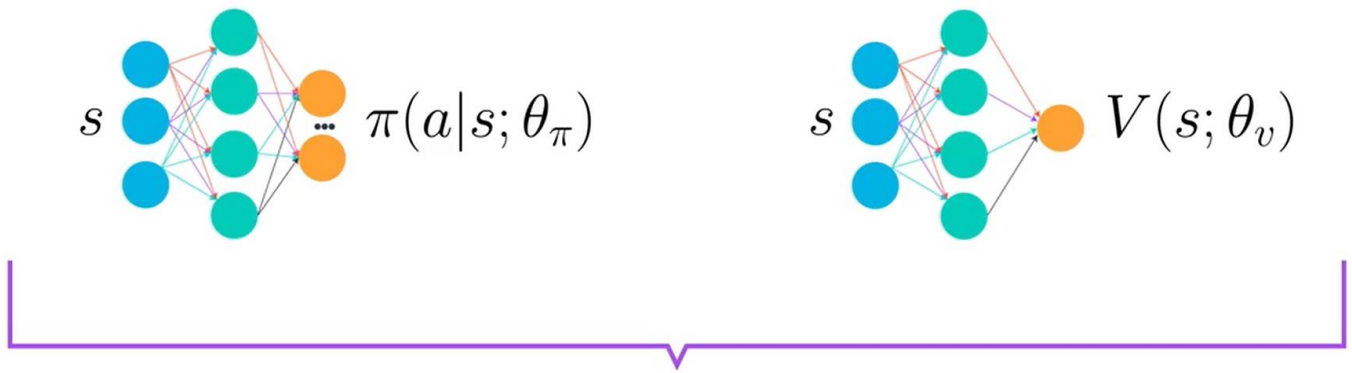
We are working with continuous state-space & continuous action-space, where we have 2 player and a ball, the objective is to bounce the ball as long as possible (collaborative)



DDPG

(Deep Deterministic Policy Gradient)

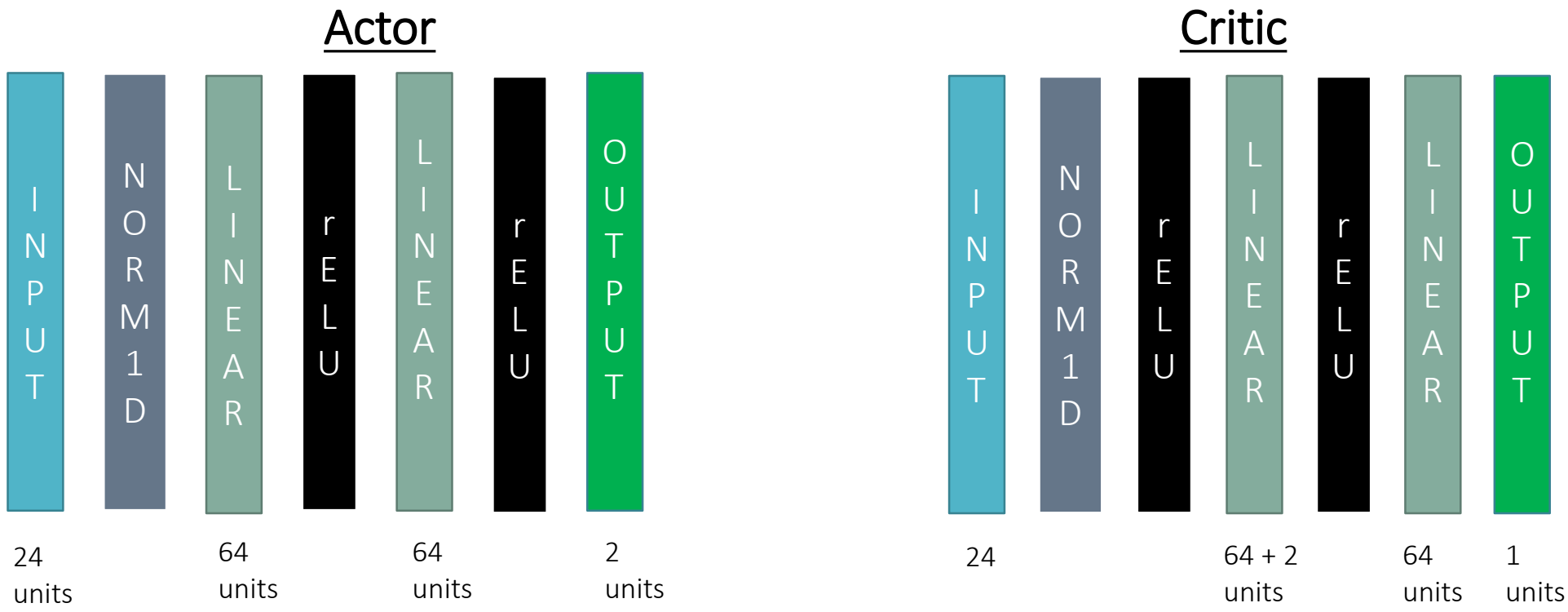
We will have two networks one for the Actor, the other for critic, using the action (the output from the actor-network) to update the value-function at critic-network



And for every network, we will have Regular and Target, and for every step we are going to make target closer to Regular

Convolution Neural Network architecture

We are going to build two networks to train our model (updating the weights) using epochs (iteration) of forward & backpropagation, where is the output (Action) of the first network (Actor) will be the input for the second network (Critic)



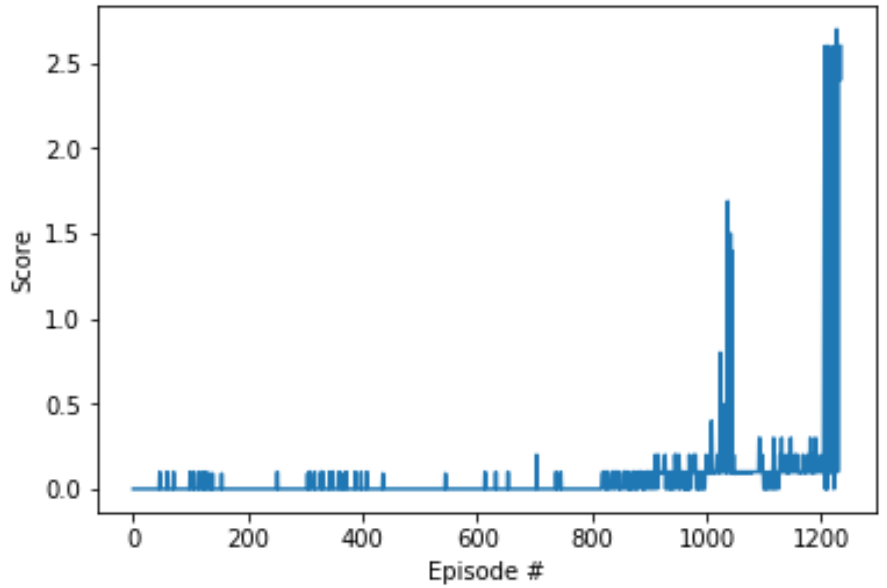
Parameters

- GAMMA (discounted rate) = 0.99 ➔ this specify we are interested in our future reward
- Tau (soft-update) = 5e-4 ➔ help prevent the variance, by how much we are going to update the Target network
- BATCH_SIZE (Replay Buffer) = 64 ➔ more about Replay in the next slide
- BUFFER_SIZE = 10^5 ➔ replay buffer size
- LR_Actor, LR_CRTICT = 5e-4 ➔ learning rate for Actor, and Critic networks
- We are going to update target with Regular for every time-step using Tau percentage of Regular every-time

Result

The Agent reaches **score > 0.5** after 1234 episodes

```
Episode 100     Average Score: 0.00 Score: 0.10
Episode 200     Average Score: 0.01 Score: 0.00
Episode 300     Average Score: 0.00 Score: 0.00
Episode 400     Average Score: 0.01 Score: 0.00
Episode 500     Average Score: 0.00 Score: 0.00
Episode 600     Average Score: 0.00 Score: 0.00
Episode 700     Average Score: 0.00 Score: 0.00
Episode 800     Average Score: 0.00 Score: 0.00
Episode 900     Average Score: 0.04 Score: 0.10
Episode 1000    Average Score: 0.08 Score: 0.20
Episode 1100    Average Score: 0.17 Score: 0.00
Episode 1200    Average Score: 0.11 Score: 0.19
Episode 1233    Average Score: 0.50 Score: 2.40
Episode 1234    Average Score: 0.52
```



Future Work

Trying to get higher score with fewer number of episodes (solving the environment earlier)

- Implement the multi-agent Knowledge share (where we have 2 separate Agents with share Replay Buffer, already tried but bad score, trying tuning parameters)
- Solve the environment using MARL

Resources/References

Using the codes that used for solving the pendulum-environment existing in this Repo also the codes for Continuous-Control

Only by changing the parameters like the network architectures and which layers

to use and selecting the best architecture for the environment