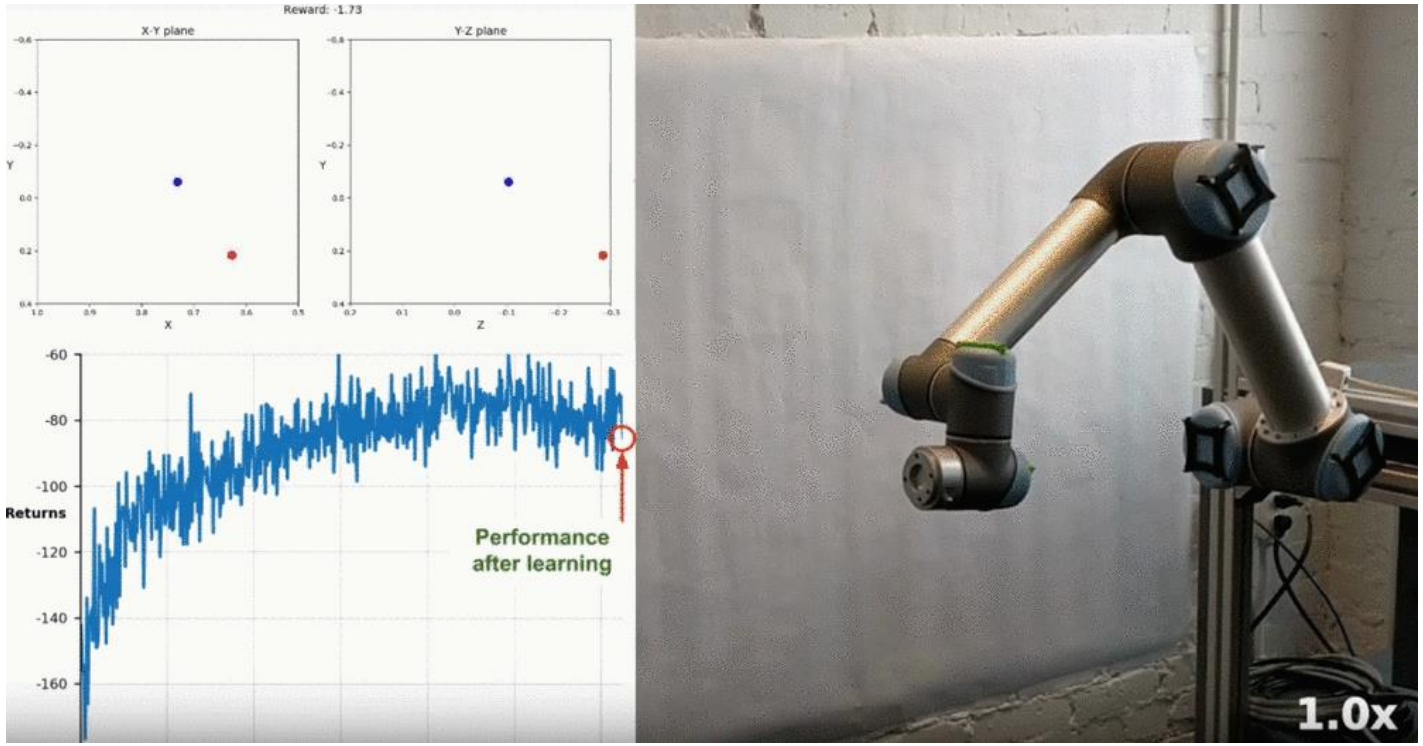
Continuous Control

The Environments

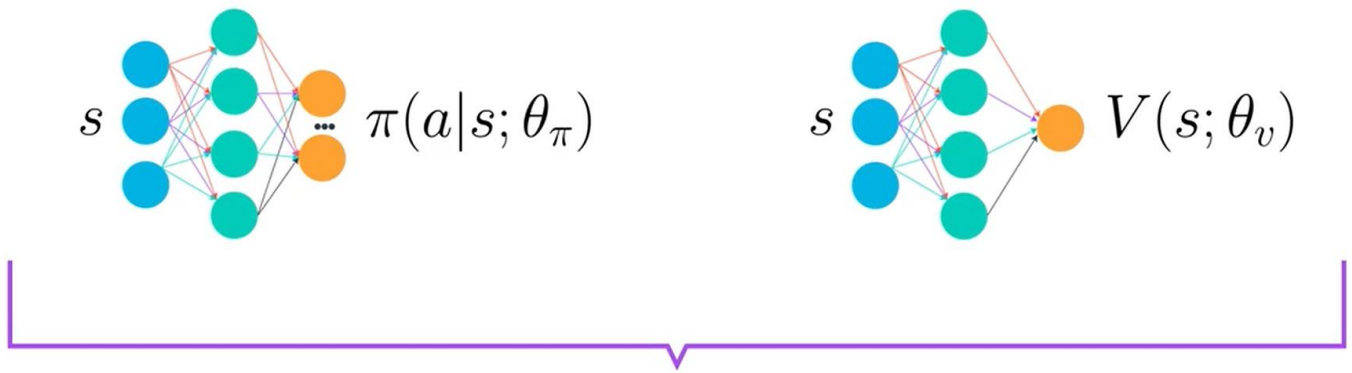
We are working with continuous state-space & continuous action-space, where we having a moving object that we trying to grab (follow)



DDPG

(Deep Deterministic Policy Gradient)

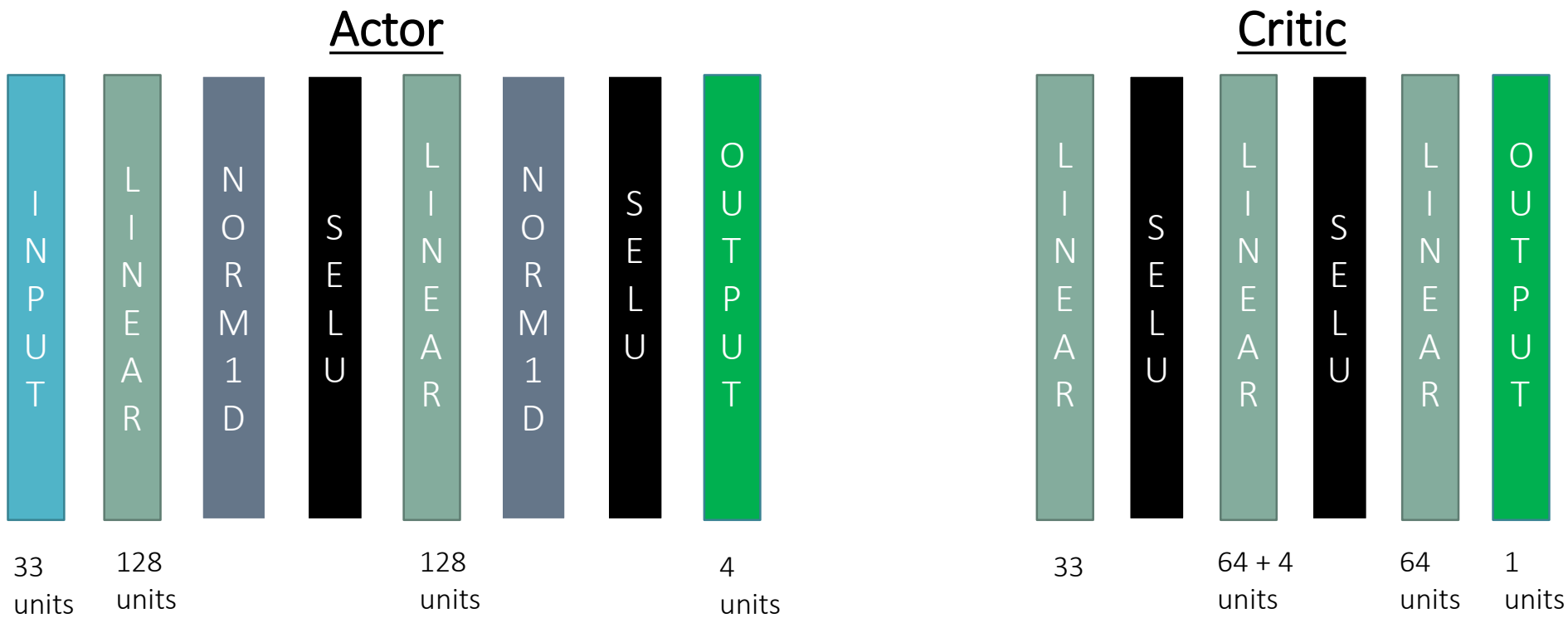
We will have two networks one for the Actor, the other for critic, using the action (the output from the actor-network) to update the value-function at critic-network



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)



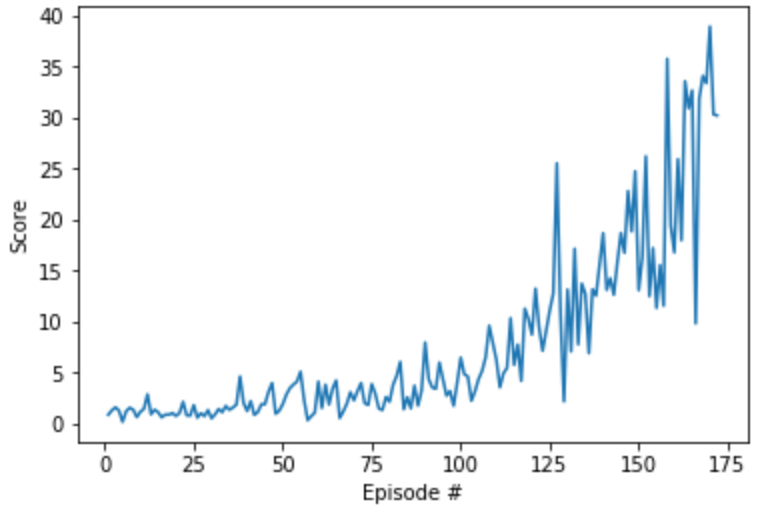
Parameters

- 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>30.0** after 172 episodes

```
Episode 10 max-Score: 1.56  min-Score: 0.14  Average Score: 1.08
Episode 20 max-Score: 2.84  min-Score: 0.58  Average Score: 1.15
Episode 30 max-Score: 2.10  min-Score: 0.48  Average Score: 1.04
Episode 40 max-Score: 4.60  min-Score: 0.87  Average Score: 1.75
Episode 50 max-Score: 3.95  min-Score: 0.82  Average Score: 1.89
Episode 60 max-Score: 5.06  min-Score: 0.30  Average Score: 2.76
Episode 70 max-Score: 4.20  min-Score: 0.51  Average Score: 2.36
Episode 80 max-Score: 3.95  min-Score: 1.30  Average Score: 2.50
Episode 90 max-Score: 7.93  min-Score: 1.37  Average Score: 3.64
Episode 100 max-Score: 6.44  min-Score: 1.72  Average Score: 3.95
Episode 110 max-Score: 9.58  min-Score: 2.21  Average Score: 5.44
Episode 120 max-Score: 11.25  min-Score: 3.53  Average Score: 7.20
Episode 130 max-Score: 25.52  min-Score: 2.14  Average Score: 11.44
Episode 140 max-Score: 18.66  min-Score: 6.88  Average Score: 12.52
Episode 150 max-Score: 24.74  min-Score: 12.61  Average Score: 17.06
Episode 160 max-Score: 35.78  min-Score: 11.32  Average Score: 18.25
Episode 170 max-Score: 38.95  min-Score: 9.79  Average Score: 28.911
Episode 172 max-Score: 38.95  min-Score: 9.79  Average Score: 30.58
solved at 172 Episode Average Score: 30.58
```



Future Work

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

- Implement the multi-agent Knowledge share (where we have 20 Agent)
- Solve the environment using A2C
- I'm facing issue running the script in this Repo (Due to libraries missing), will be easier for future environment solving

Resources/References

Using the codes to solving the pendulum-environment existing in this Repo

Only by changing the parameters like the network architectures and which layers to use and selecting the best architecture for the environment