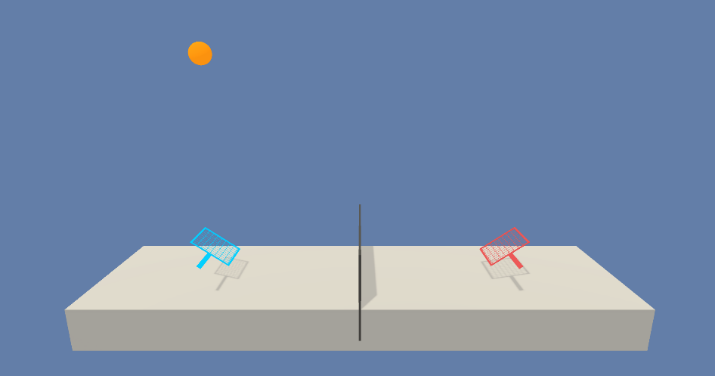
# Project 3 – Collaboration and competition

## Introduction:

The purpose of this project is to train two agents which control rackets to bounce a ball over a net. If an agent hits the ball over the net, it receives a reward of +0.1. If an agent lets a ball hit the ground or hits the ball out of bounds, it receives a reward of -0.01. Thus, the goal of each agent is to keep the ball in play.

For this project, we will work with the Unity Tennis environment.

The observation space consists of 8 variables corresponding to the position and velocity of the ball and racket. Each agent receives its own, local observation. Two continuous actions are available, corresponding to movement toward (or away from) the net, and jumping.

The task is episodic, and in order to solve the environment, your agents must get an average score of +0.5 (over 100 consecutive episodes, after taking the maximum over both agents). Specifically:

* After each episode, we add up the rewards that each agent received (without discounting), to get a score for each agent. This yields 2 (potentially different) scores. We then take the maximum of these 2 scores.
* This yields a single **score** for each episode.

The environment is considered solved, when the average (over 100 episodes) of those **scores** is at least +0.5.

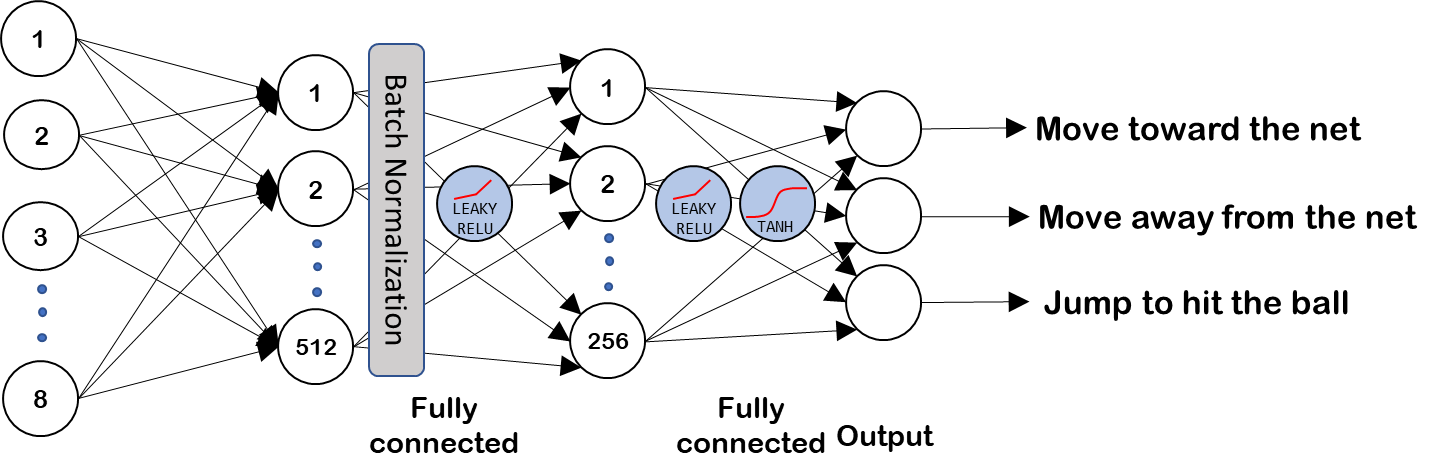
## Learning Algorithm:

We may consider that both agents are learning to play a game of trying to get the ball staying in the air, so it may be considered as a unique game played by two agents, instead of a sum-zero (competitive) game where both agents want the other to fail.

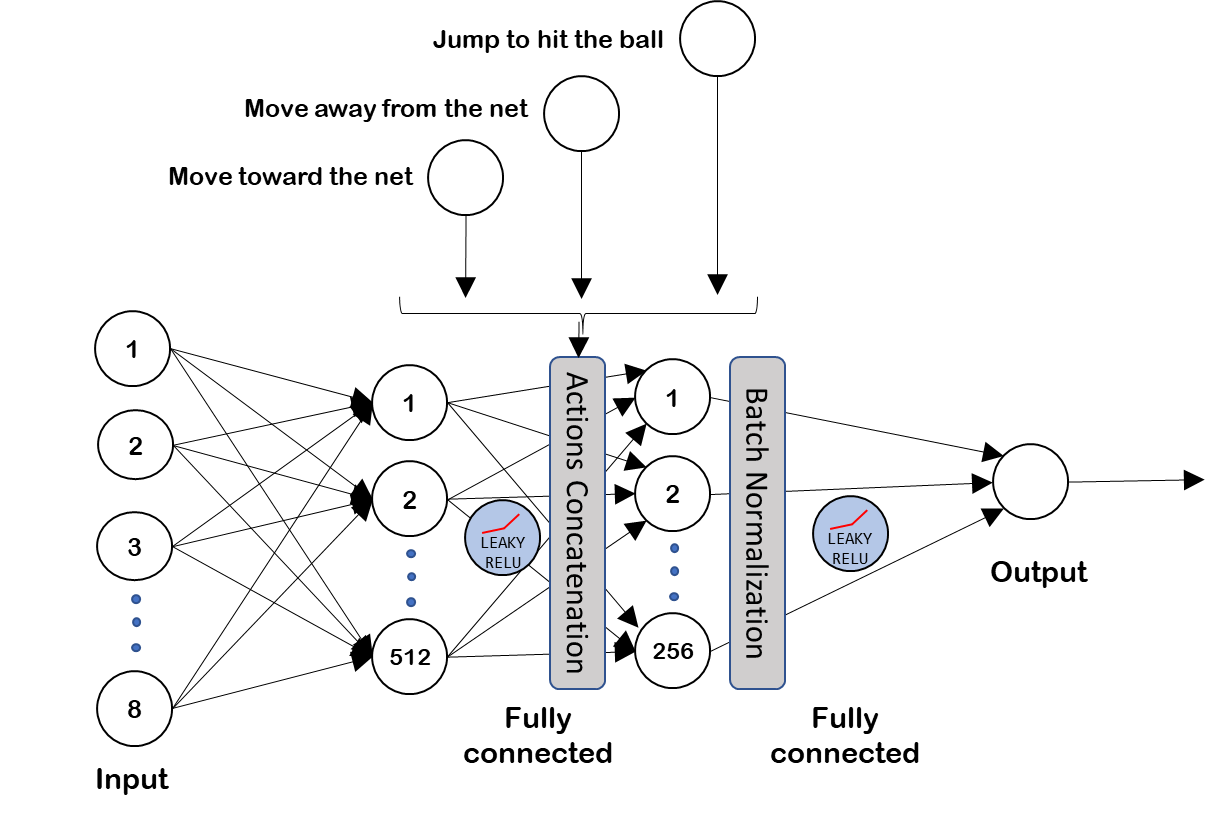
Based on the above premise, we will start testing with the DDPG algorithm used in the previous project (Continuous Control). Both agents will use the same neural network architecture and the **same replay buffer**.

## Actor Neural Network:

The actor NN used is the following one, the ending function used is a Tanh.



## Critic Neural Network:



Thanks to the implementation of **experience replay**, the algorithm stores states which are rare and actions

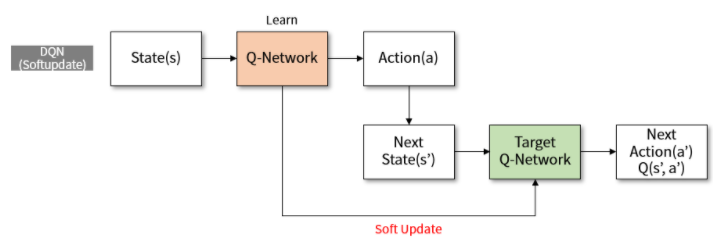
that are costly to be able to recall them, each experience (Which is formed by State, Action, Reward and Next

State) is stored in a buffer as the agent is interacting with the environment, afterwards the agent randomly samples a small batch of these experiences in order to learn from them. Thanks to this:

* it learns from individual state-actions multiple times, so it recalls rare occurrences and make a better use of the experience obtained.
* Because of the random sampling, it helps breaking the correlation and prevent actions from oscillating or diverging in wrong ways.

Another critic functionality of the DDPG algorithm is a **soft update** to the target network.

The implementation of this soft update means that the algorithm has two neural networks. The called “Regular” neural network and the “Target” neural network.



In our algorithm, this soft update is done every step, so the networks are blended with a percentage of merging considering the TAU variable, so considering the TAU variable is set to 0,05%, every step a 0,05% of the regular network will be merged with the target network.

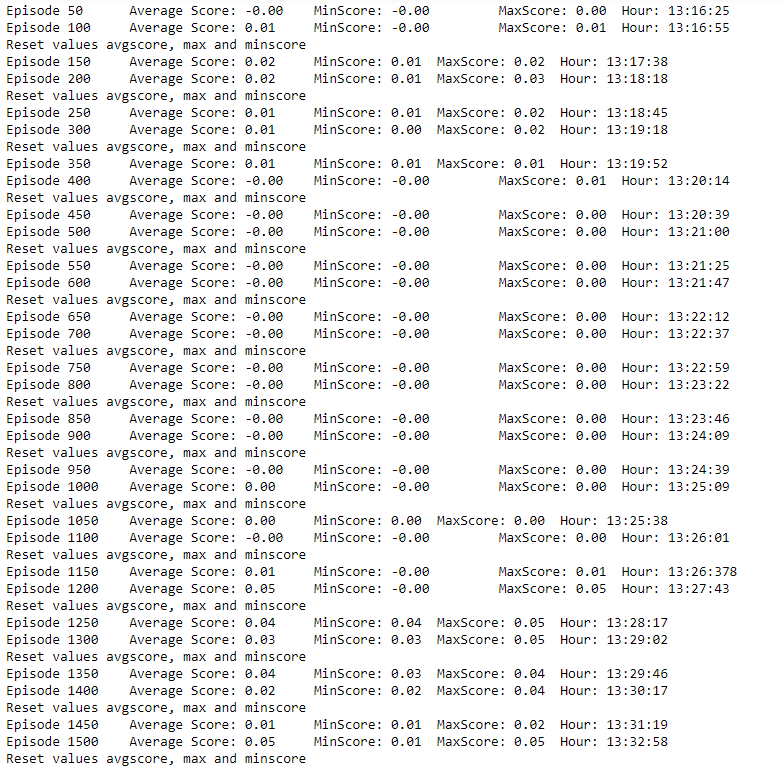
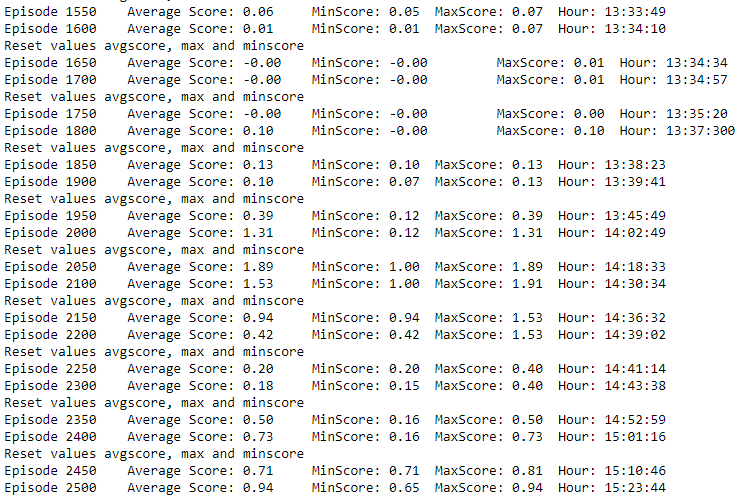
Thanks to soft update, the DDPG algorithm gets faster convergence. Soft update logic could be implemented in other off-policy algorithms that use target networks, such as DQN.

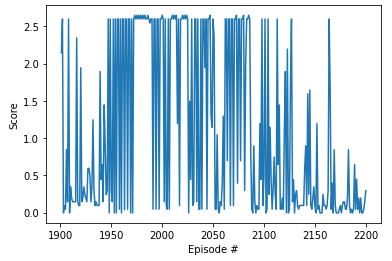
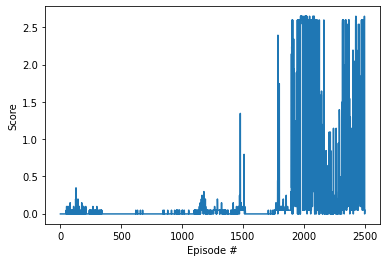
To encourage exploration in the training process, we added **noise** using the [Ornstein-Uhlenbeck](http://entsphere.com/pub/pdf/1930%20Uhlenbeck,%20on%20the%20theory%20of%20the%20Brownian%20motion.pdf) noise process, the noise is reduced through time, because when the agent is more experienced, the need for exploration decreases.

We used [**batch normalization**](https://arxiv.org/pdf/1502.03167.pdf) in both networks for stabilizing the learning process and reduce the training time for deep neural networks, batch normalization is especially useful to limit covariate shift in environments where the input distribution changes with each step, like this.

On the critic side, we have **used leaky relu** activation functions for dealing better with the possible negative values coming from the learning (*Dying ReLU problem***)** and improve training performance.

## Algorithm performance:





## Future improvements:

Some future improvement that could be done in this project are:

1. Use **MaDDPG** algorithm, the learning process should be faster.
2. Implement **Prioritized Experience replay**, the learning process should be more efficient
3. Try making a hard-update on the target network **periodically** to make the learning process faster without losing the soft updates improvements to the learning process.
4. Decrease exploration through the learning process
5. Implementing some features described in the [benchmarking DeepRL for Continuous control paper](https://arxiv.org/abs/1604.06778):
   * Rescale the reward by a factor of 0.1 through the learning process, this should improve stability in the learning process.
   * Change learning rates to 5x10-3, usually this value improves performance.
6. Analyze if it’s possible to improve the neural network architecture by using Archai:

<https://www.microsoft.com/en-us/research/blog/archai-can-design-your-neural-network-with-state-of-the-art-neural-architecture-search-nas/>

2nd Iteration:

Episode 50 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:04:40

Episode 100 Average Score: 0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:06:39

Reset values avgscore, max and minscore

Episode 150 Average Score: 0.00 MinScore: 0.00 MaxScore: 0.00 Hour: 13:08:04

Episode 200 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:09:26

Reset values avgscore, max and minscore

Episode 250 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:10:50

Episode 300 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:12:12

Reset values avgscore, max and minscore

Episode 350 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:13:38

Episode 400 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:15:12

Reset values avgscore, max and minscore

Episode 450 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:16:31

Episode 500 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:17:54

Reset values avgscore, max and minscore

Episode 550 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:19:14

Episode 600 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:20:36

Reset values avgscore, max and minscore

Episode 650 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:22:00

Episode 700 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:23:20

Reset values avgscore, max and minscore

Episode 750 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:24:38

Episode 800 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:25:56

Reset values avgscore, max and minscore

Episode 850 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:27:16

Episode 900 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:28:38

Reset values avgscore, max and minscore

Episode 950 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:30:02

Episode 1000 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:31:26

Reset values avgscore, max and minscore

Episode 1050 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:32:53

Episode 1100 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:34:14

Reset values avgscore, max and minscore

Episode 1150 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:35:32

Episode 1200 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:36:51

Reset values avgscore, max and minscore

Episode 1250 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:38:12

Episode 1300 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:39:28

Reset values avgscore, max and minscore

Episode 1350 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:40:48

Episode 1400 Average Score: -0.00 MinScore: -0.00 MaxScore: 0.00 Hour: 13:42:13

Reset values avgscore, max and minscore

Episode 1450 Average Score: 0.01 MinScore: -0.00 MaxScore: 0.01 Hour: 13:43:583

Episode 1500 Average Score: 0.01 MinScore: -0.00 MaxScore: 0.01 Hour: 13:45:21

Reset values avgscore, max and minscore

Episode 1550 Average Score: 0.00 MinScore: 0.00 MaxScore: 0.01 Hour: 13:46:57

Episode 1600 Average Score: 0.01 MinScore: 0.00 MaxScore: 0.01 Hour: 13:48:53

Reset values avgscore, max and minscore

Episode 1650 Average Score: 0.02 MinScore: 0.01 MaxScore: 0.02 Hour: 13:51:01

Episode 1700 Average Score: 0.03 MinScore: 0.01 MaxScore: 0.03 Hour: 13:53:21

Reset values avgscore, max and minscore

Episode 1750 Average Score: 0.08 MinScore: 0.03 MaxScore: 0.08 Hour: 13:58:23

Episode 1800 Average Score: 0.12 MinScore: 0.03 MaxScore: 0.12 Hour: 14:03:00

Reset values avgscore, max and minscore

Episode 1850 Average Score: 0.11 MinScore: 0.11 MaxScore: 0.13 Hour: 14:08:49

Episode 1900 Average Score: 0.38 MinScore: 0.11 MaxScore: 0.38 Hour: 14:28:59

Reset values avgscore, max and minscore

Episode 1950 Average Score: 0.77 MinScore: 0.41 MaxScore: 0.77 Hour: 15:00:13

Episode 2000 Average Score: 0.55 MinScore: 0.41 MaxScore: 0.77 Hour: 15:08:40

Reset values avgscore, max and minscore

Episode 2050 Average Score: 0.41 MinScore: 0.38 MaxScore: 0.53 Hour: 15:29:51

Episode 2100 Average Score: 0.51 MinScore: 0.38 MaxScore: 0.53 Hour: 15:44:12

Reset values avgscore, max and minscore

Episode 2150 Average Score: 0.46 MinScore: 0.46 MaxScore: 0.55 Hour: 16:06:33

Episode 2200 Average Score: 0.36 MinScore: 0.36 MaxScore: 0.55 Hour: 16:16:55

Reset values avgscore, max and minscore

Episode 2250 Average Score: 0.17 MinScore: 0.17 MaxScore: 0.34 Hour: 16:24:14

Episode 2300 Average Score: 0.17 MinScore: 0.16 MaxScore: 0.34 Hour: 16:33:09

Reset values avgscore, max and minscore

Episode 2350 Average Score: 0.20 MinScore: 0.17 MaxScore: 0.21 Hour: 16:41:45

Episode 2400 Average Score: 0.16 MinScore: 0.15 MaxScore: 0.21 Hour: 16:47:09

Reset values avgscore, max and minscore

Episode 2450 Average Score: 0.14 MinScore: 0.11 MaxScore: 0.15 Hour: 16:56:09

Episode 2500 Average Score: 0.21 MinScore: 0.11 MaxScore: 0.21 Hour: 17:06:04

Reset values avgscore, max and minscore

Episode 2550 Average Score: 0.21 MinScore: 0.21 MaxScore: 0.25 Hour: 17:14:09

Episode 2600 Average Score: 0.16 MinScore: 0.16 MaxScore: 0.25 Hour: 17:20:34

Reset values avgscore, max and minscore

Episode 2650 Average Score: 0.17 MinScore: 0.14 MaxScore: 0.19 Hour: 17:29:40

Episode 2700 Average Score: 0.28 MinScore: 0.14 MaxScore: 0.28 Hour: 17:41:07

Reset values avgscore, max and minscore

Episode 2750 Average Score: 0.36 MinScore: 0.28 MaxScore: 0.37 Hour: 17:57:21

Episode 2800 Average Score: 0.43 MinScore: 0.28 MaxScore: 0.45 Hour: 18:15:18

Reset values avgscore, max and minscore

Episode 2850 Average Score: 0.33 MinScore: 0.32 MaxScore: 0.42 Hour: 18:24:14

Episode 2900 Average Score: 0.14 MinScore: 0.14 MaxScore: 0.42 Hour: 18:30:38

Reset values avgscore, max and minscore

Episode 2950 Average Score: 0.15 MinScore: 0.12 MaxScore: 0.16 Hour: 18:41:16

Episode 3000 Average Score: 0.13 MinScore: 0.12 MaxScore: 0.16 Hour: 18:45:49

Reset values avgscore, max and minscore

Episode 3050 Average Score: 0.10 MinScore: 0.09 MaxScore: 0.13 Hour: 18:53:46

Episode 3100 Average Score: 0.20 MinScore: 0.09 MaxScore: 0.21 Hour: 19:05:53

Reset values avgscore, max and minscore

Episode 3150 Average Score: 0.46 MinScore: 0.20 MaxScore: 0.47 Hour: 19:30:42

Episode 3200 Average Score: 0.54 MinScore: 0.20 MaxScore: 0.55 Hour: 19:48:30

Reset values avgscore, max and minscore

Episode 3250 Average Score: 0.31 MinScore: 0.31 MaxScore: 0.54 Hour: 19:57:08

Episode 3300 Average Score: 0.19 MinScore: 0.18 MaxScore: 0.54 Hour: 20:05:31

Reset values avgscore, max and minscore

Episode 3350 Average Score: 0.27 MinScore: 0.19 MaxScore: 0.29 Hour: 20:20:42

Episode 3400 Average Score: 0.24 MinScore: 0.19 MaxScore: 0.29 Hour: 20:27:08

Reset values avgscore, max and minscore

Episode 3450 Average Score: 0.12 MinScore: 0.12 MaxScore: 0.24 Hour: 20:33:08

Episode 3500 Average Score: 0.13 MinScore: 0.11 MaxScore: 0.24 Hour: 20:40:35

Reset values avgscore, max and minscore

Episode 3550 Average Score: 0.12 MinScore: 0.12 MaxScore: 0.13 Hour: 20:45:58

Episode 3600 Average Score: 0.13 MinScore: 0.12 MaxScore: 0.14 Hour: 20:54:34

Reset values avgscore, max and minscore

Episode 3650 Average Score: 0.16 MinScore: 0.13 MaxScore: 0.16 Hour: 21:01:55

Episode 3700 Average Score: 0.12 MinScore: 0.12 MaxScore: 0.16 Hour: 21:07:46

Reset values avgscore, max and minscore

Episode 3750 Average Score: 0.10 MinScore: 0.09 MaxScore: 0.12 Hour: 21:13:30

Episode 3800 Average Score: 0.16 MinScore: 0.09 MaxScore: 0.16 Hour: 21:23:44

Reset values avgscore, max and minscore

Episode 3850 Average Score: 0.19 MinScore: 0.15 MaxScore: 0.20 Hour: 21:33:11

Episode 3900 Average Score: 0.10 MinScore: 0.10 MaxScore: 0.20 Hour: 21:37:00

Reset values avgscore, max and minscore

Episode 3950 Average Score: 0.11 MinScore: 0.10 MaxScore: 0.14 Hour: 21:46:47

Episode 4000 Average Score: 0.45 MinScore: 0.10 MaxScore: 0.45 Hour: 22:15:52

Reset values avgscore, max and minscore

Episode 4050 Average Score: 0.39 MinScore: 0.39 MaxScore: 0.46 Hour: 22:20:03

Episode 4100 Average Score: 0.05 MinScore: 0.05 MaxScore: 0.46 Hour: 22:23:19

Reset values avgscore, max and minscore

Episode 4150 Average Score: 0.05 MinScore: 0.05 MaxScore: 0.05 Hour: 22:27:18

Episode 4200 Average Score: 0.07 MinScore: 0.05 MaxScore: 0.07 Hour: 22:32:03

Reset values avgscore, max and minscore

Episode 4250 Average Score: 0.11 MinScore: 0.07 MaxScore: 0.11 Hour: 22:39:27

Episode 4300 Average Score: 0.14 MinScore: 0.07 MaxScore: 0.14 Hour: 22:45:22

Reset values avgscore, max and minscore

Episode 4350 Average Score: 0.08 MinScore: 0.08 MaxScore: 0.13 Hour: 22:49:01

Episode 4400 Average Score: 0.03 MinScore: 0.03 MaxScore: 0.13 Hour: 22:51:38

Reset values avgscore, max and minscore

Episode 4450 Average Score: 0.02 MinScore: 0.02 MaxScore: 0.03 Hour: 22:53:58

Episode 4500 Average Score: 0.02 MinScore: 0.02 MaxScore: 0.03 Hour: 22:56:46

Reset values avgscore, max and minscore

Episode 4550 Average Score: 0.03 MinScore: 0.02 MaxScore: 0.03 Hour: 22:59:51

Episode 4566 Average Score: 0.03 MinScore: 0.02 MaxScore: 0.04 Hour: 23:01:00