■ Collaboration and Competition

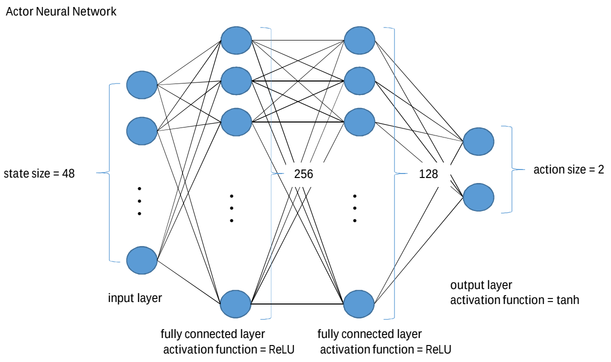
- Learning Argorithm MADDPG

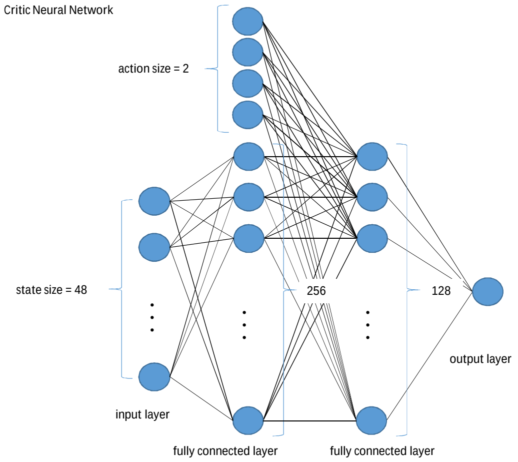
Actor-Critic is a reinforcement learning method that learns by independently estimating the probability of an action and the estimated reward value of a state. DDPG (Deep Deterministic Policy Gradient) is an off-policy actor critic algorithm that combines DPG and DQN. DQN (Deep Q-Network) stabilizes the learning of Q-functions by using experience replay and fixing the target network. DQN works in discrete space, while DDPG is a continuous space algorithm.MADDPG is an application of DDPG for multi-agents, and learns cooperative movements by regarding two tennis players as agents.

- Hyperparameters

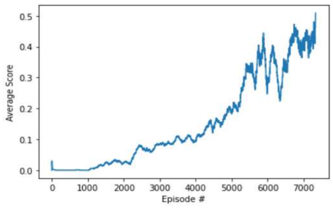
parameter		replay buffer size	1000000
		batch size	128
		discount factor	0.99
		soft update of target parameters	0.001
	actor	learning rate	0.0001
	critic	learning rate	0.0001
		L2 weight decay	0
		maximum number of training episodes	10000
neural network		state size	24 * 2
		action size	2
	actor	number of nodes in first hidden layer	256
		number of nodes in second hidden layer	128
	ctitic	number of nodes in first hidden layer	256
		number of nodes in second hidden layer	128

- **Model Architecture**





- Plot of Rewards



Reward values during training

- Ideas for Future Work

Improvements to MADDPG include adjusting hyperparameters such as increasing the number of nodes in the middle layer and applying Batch Normalization for neural networks.