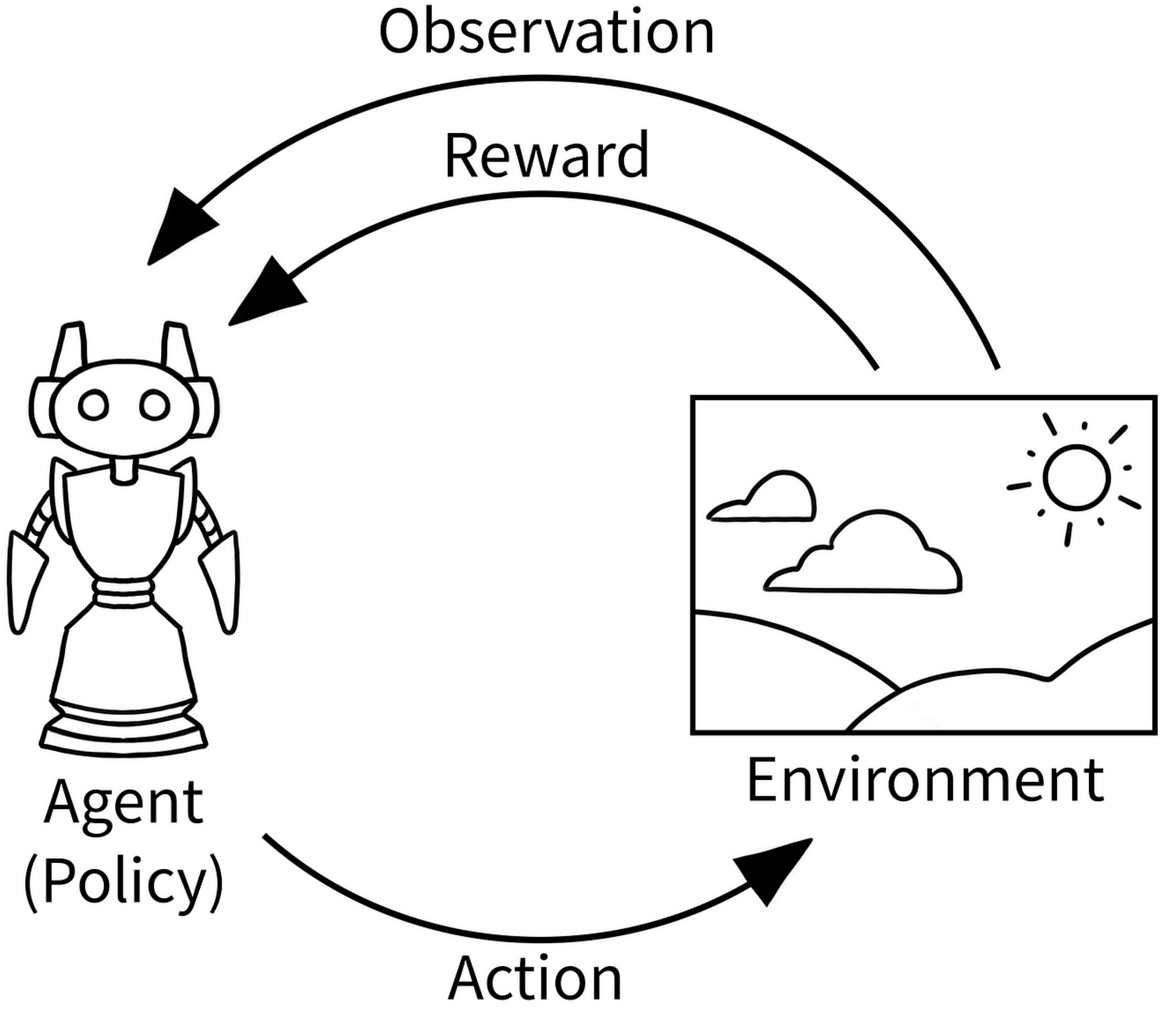
Multi-agent Deep Reinforcement Learning for Pursuit Game

AE 598 - Reinforcement Learning
April 25, 2023

Maulik Bhatt

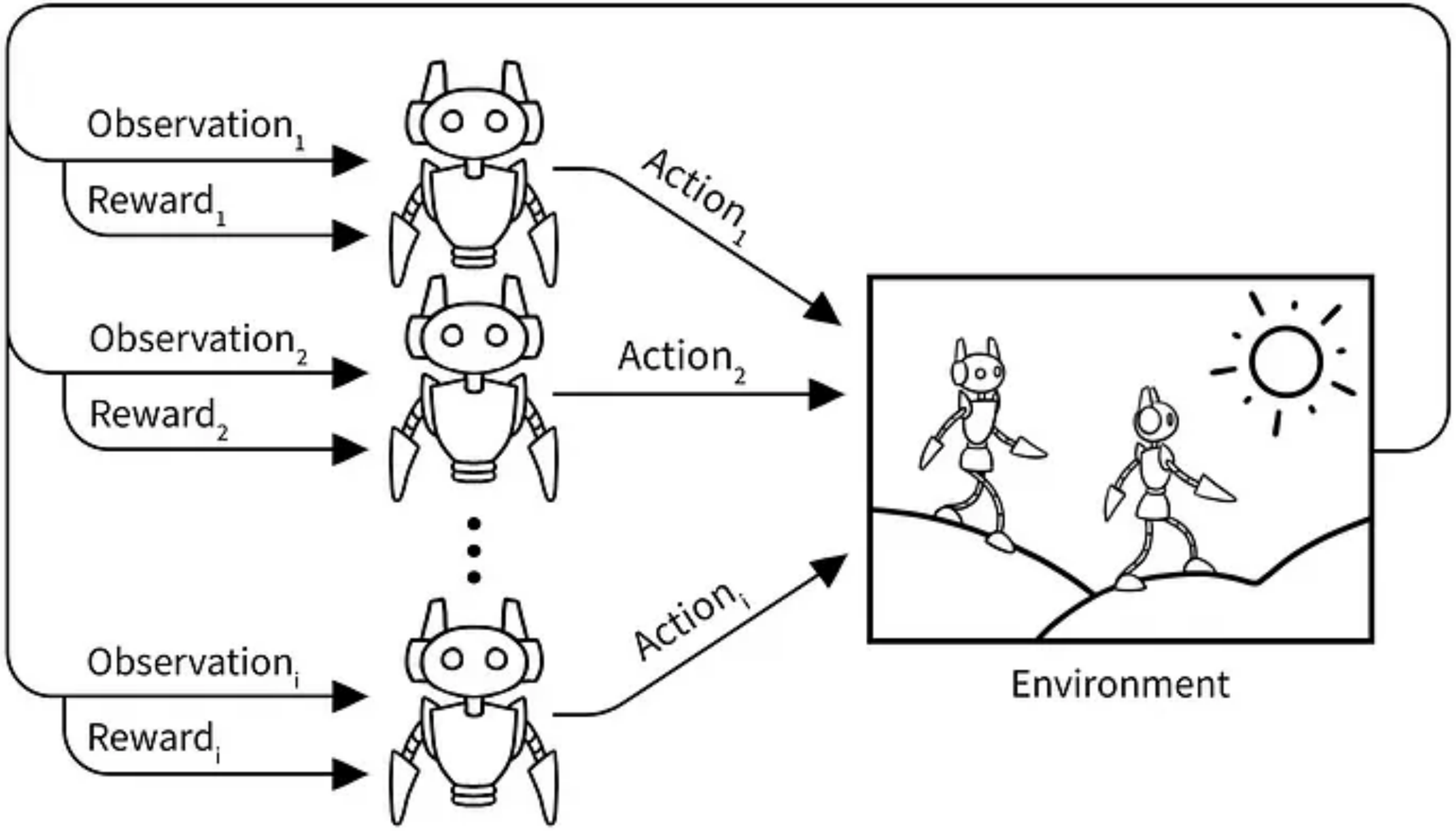
Department of Aerospace Engineering, UIUC

Source: https://miro.medium.com/v2/resize:fit:720/format:webp/1*Ews7HaMiSn2l8r70eeIszQ.png

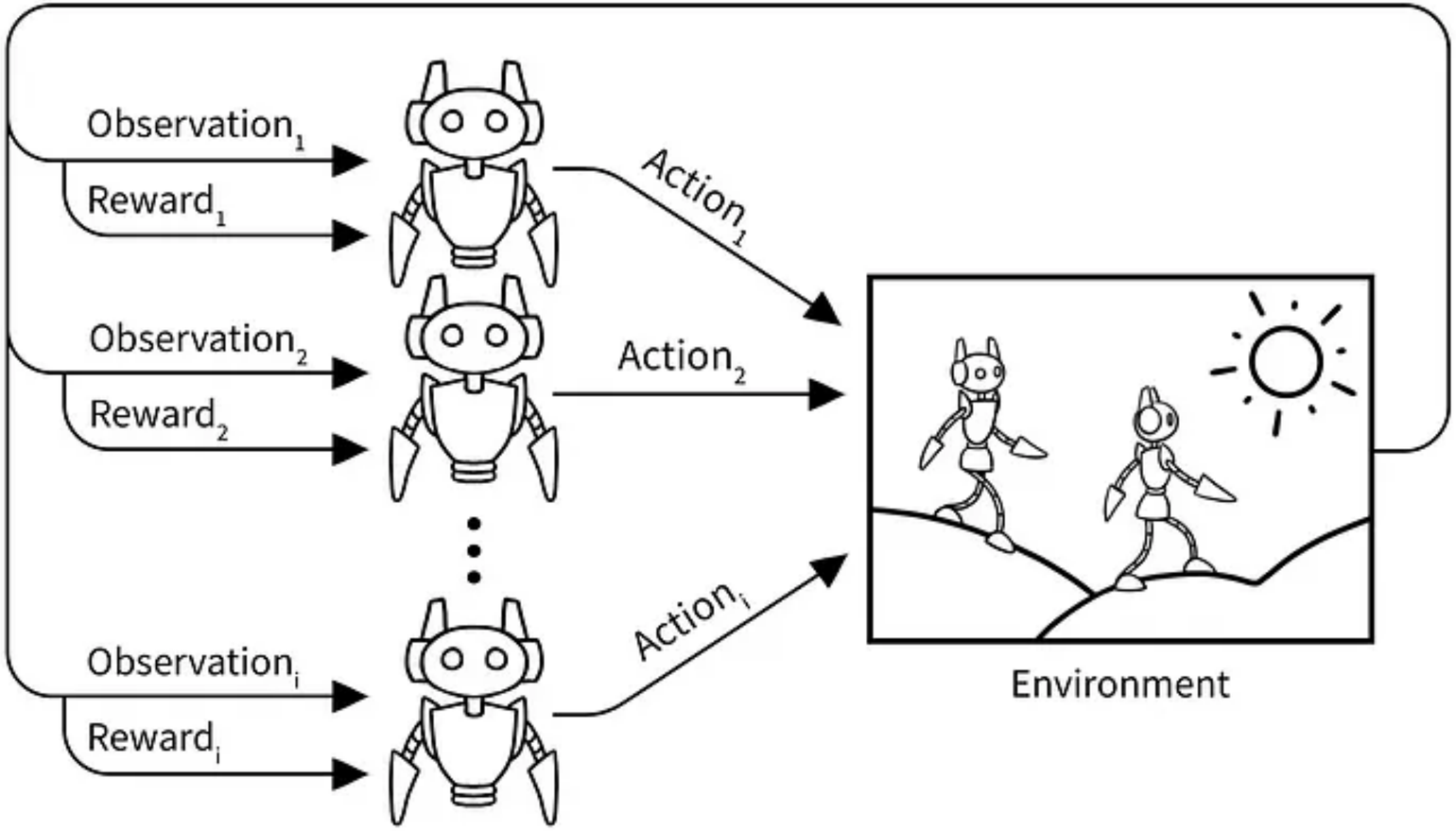


Source: https://miro.medium.com/v2/resize:fit:720/format:webp/1*Ews7HaMiSn2l8r70eeIszQ.png

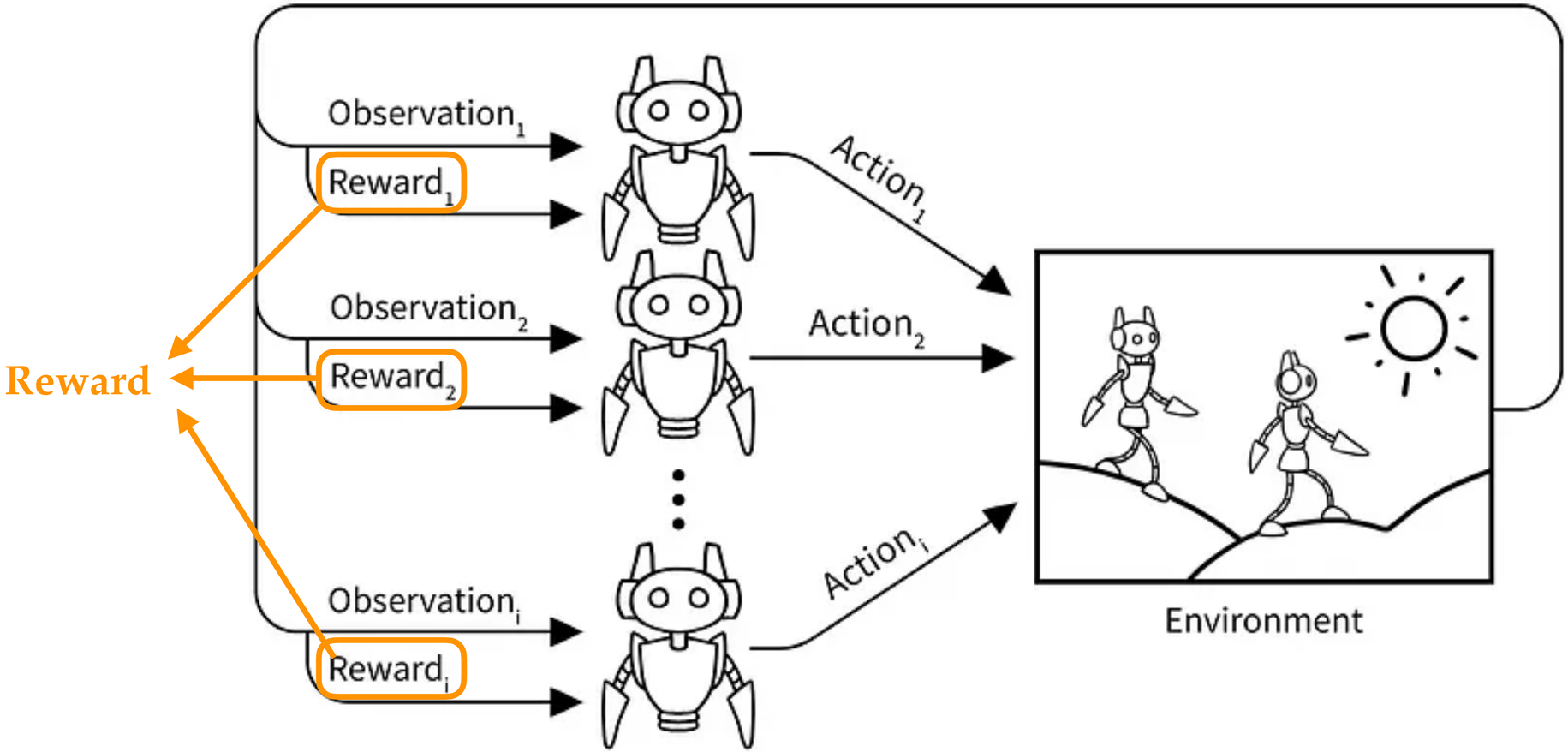
Source: https://miro.medium.com/v2/resize:fit:720/format:webp/1*1o1oeH3vpzsfJukLbFsekw.png



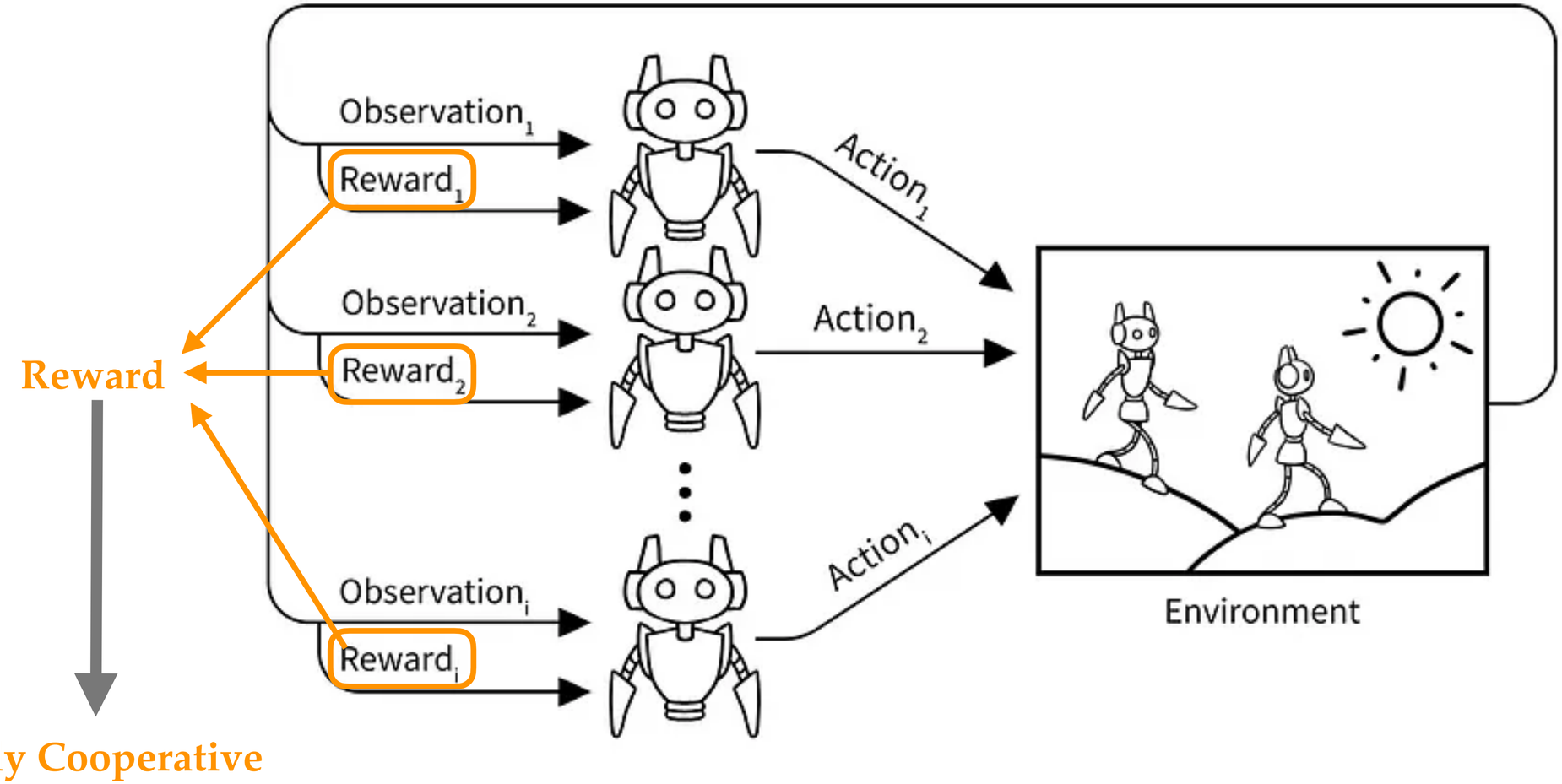
Source: https://miro.medium.com/v2/resize:fit:720/format:webp/1*1o1oeH3vpzsfJukLbFsekw.png



Source: https://miro.medium.com/v2/resize:fit:720/format:webp/1*1o1oeH3vpzsfJukLbFsekw.png



Source: https://miro.medium.com/v2/resize:fit:720/format:webp/1*1o1oeH3vpzsfJukLbFsekw.png



Source: https://miro.medium.com/v2/resize:fit:720/format:webp/1*1o1oeH3vpzsfJukLbFsekw.png

A fully cooperative multi-agent sequential decision-making task can be described as

decentralised partially observable Markov decision process

A fully cooperative multi-agent sequential decision-making task can be described as

decentralised partially observable Markov decision process

$$G = \langle S, U, P, r, Z, O, n, \gamma \rangle$$

A fully cooperative multi-agent sequential decision-making task can be described as

decentralised partially observable Markov decision process

$$G = \langle S, U, P, r, Z, O, n, \gamma \rangle$$

Aim is to minimize the *discounted return*

A fully cooperative multi-agent sequential decision-making task can be described as

decentralised partially observable Markov decision process

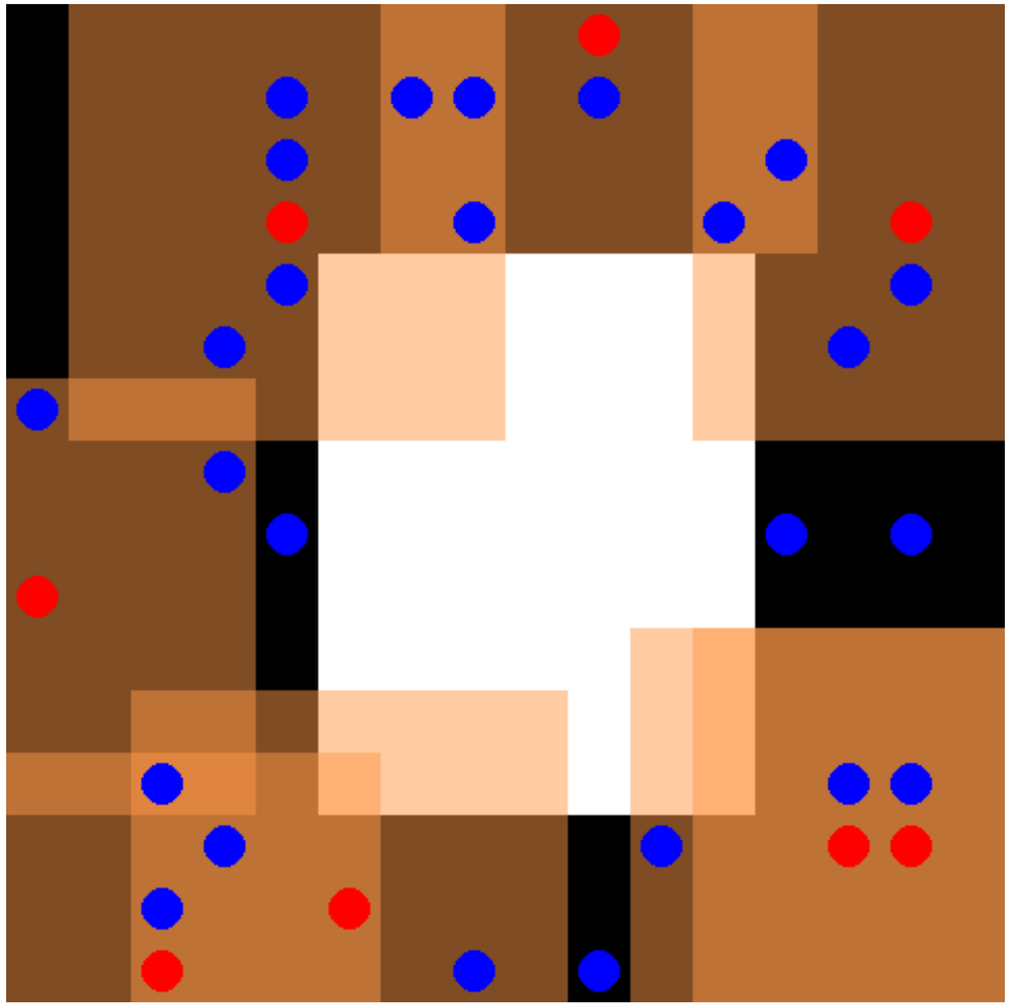
$$G = \langle S, U, P, r, Z, O, n, \gamma \rangle$$

Aim is to minimize the *discounted return*

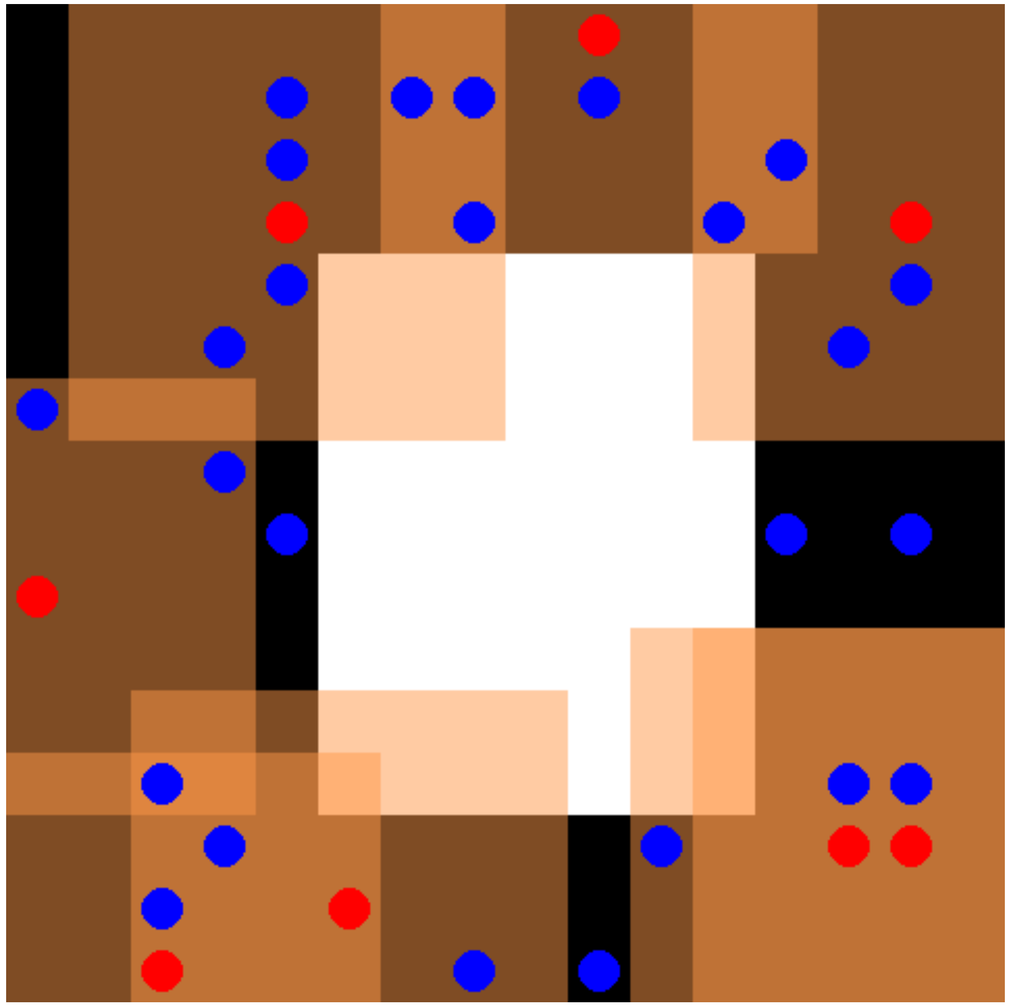
$$R_t = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$$

Pursuit Environment [1]

Pursuit Environment [1]



Pursuit Environment [1]



Number of pursuers: 8

Number of evaders: 25

Grid size: 16*16

Observation size: 7*7*3

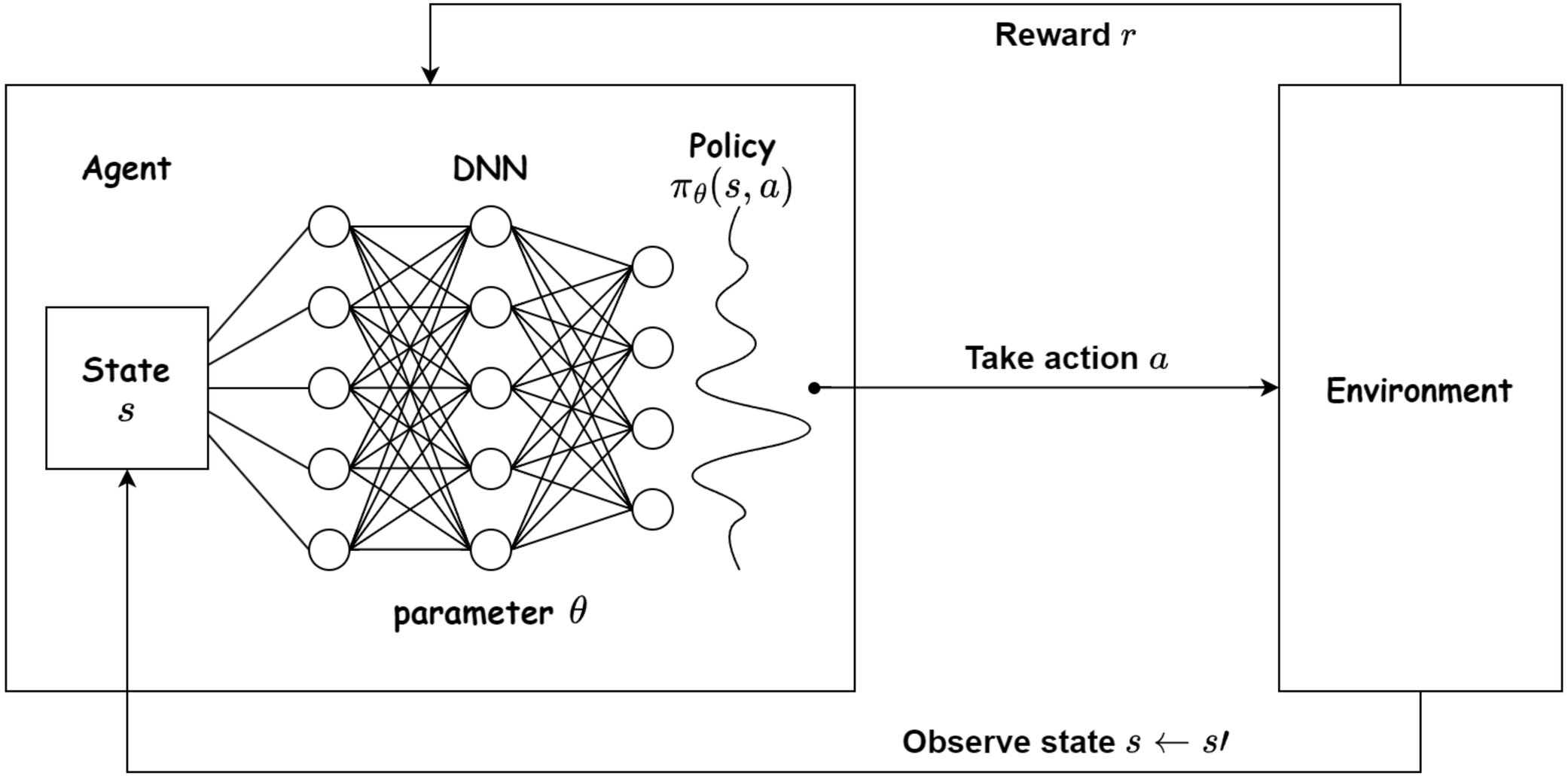
Actions: up, down, left, right, stay

Max Cycles: 500

Deep Q-Learning

Source: https://wikidocs.net/images/page/169312/20_Q_Network.png

Deep Q-Learning



Source: https://wikidocs.net/images/page/169312/20_Q_Network.png

Independent Q-learning (IQL) decomposes a multi-agent problem into a collection of simultaneous single-agent problems that share the same environment.

Independent Q-learning (IQL) decomposes a multi-agent problem into a collection of simultaneous single-agent problems that share the same environment.

Each learning agent is an *independent learner* without considering its influence into the environment of other agents

Independent DQN-1

Independent DQN-1

- Define the networks and agents

Independent DQN-1

- Define the networks and agents
- Initialize separate networks for each agent

Independent DQN-1

- Define the networks and agents
- Initialize separate networks for each agent
- Define the shared Replay buffer

Independent DQN-1

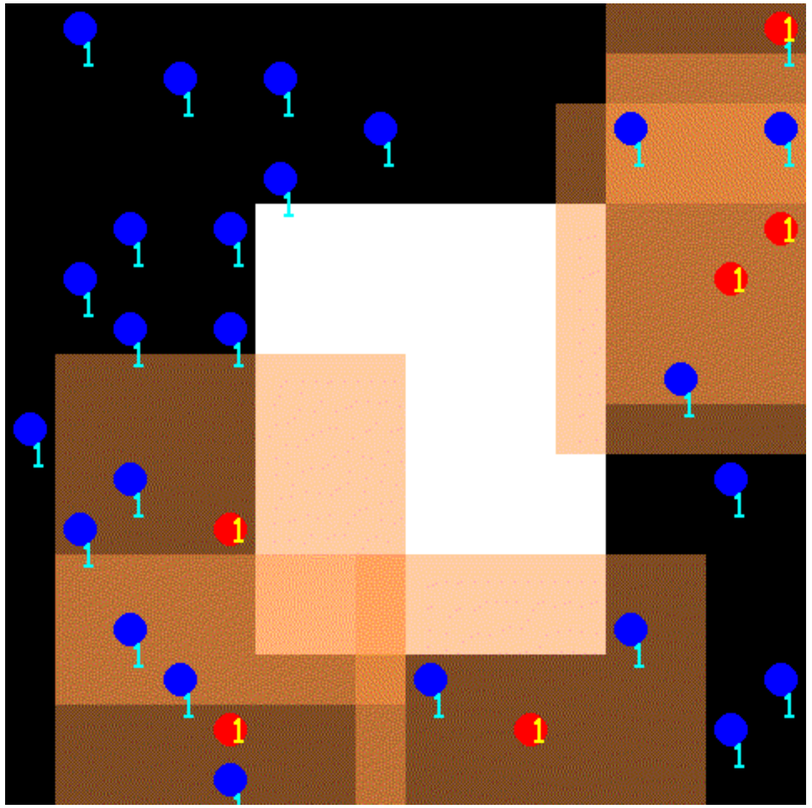
- Define the networks and agents
- Initialize separate networks for each agent
- Define the shared Replay buffer
- Implement the training loop (each agent learns independently)

Independent DQN-1

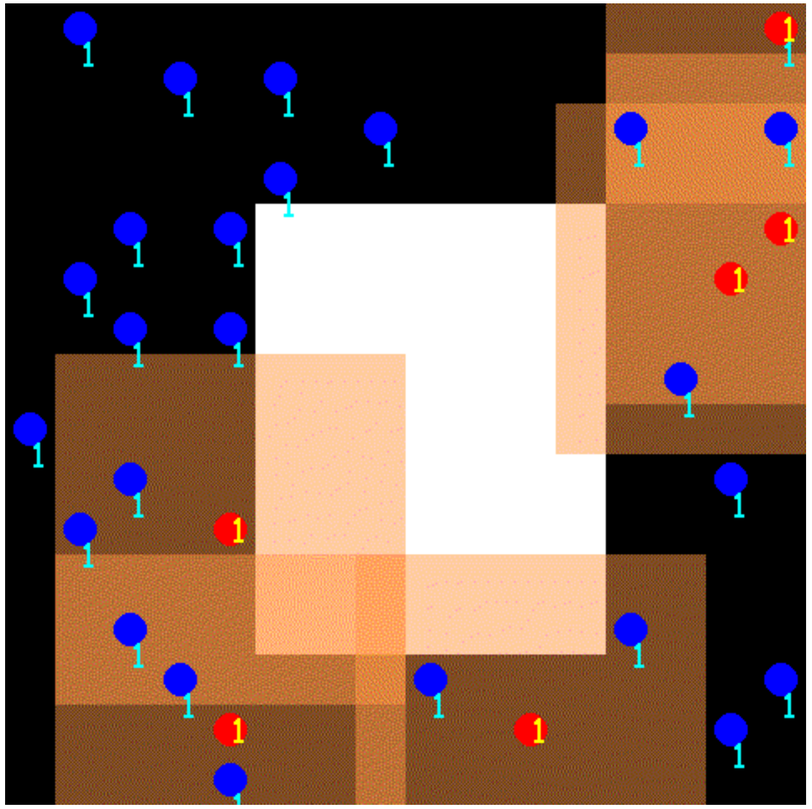
- Define the networks and agents
- Initialize separate networks for each agent
- Define the shared Replay buffer
- Implement the training loop (each agent learns independently)
- Implement the exploration strategy

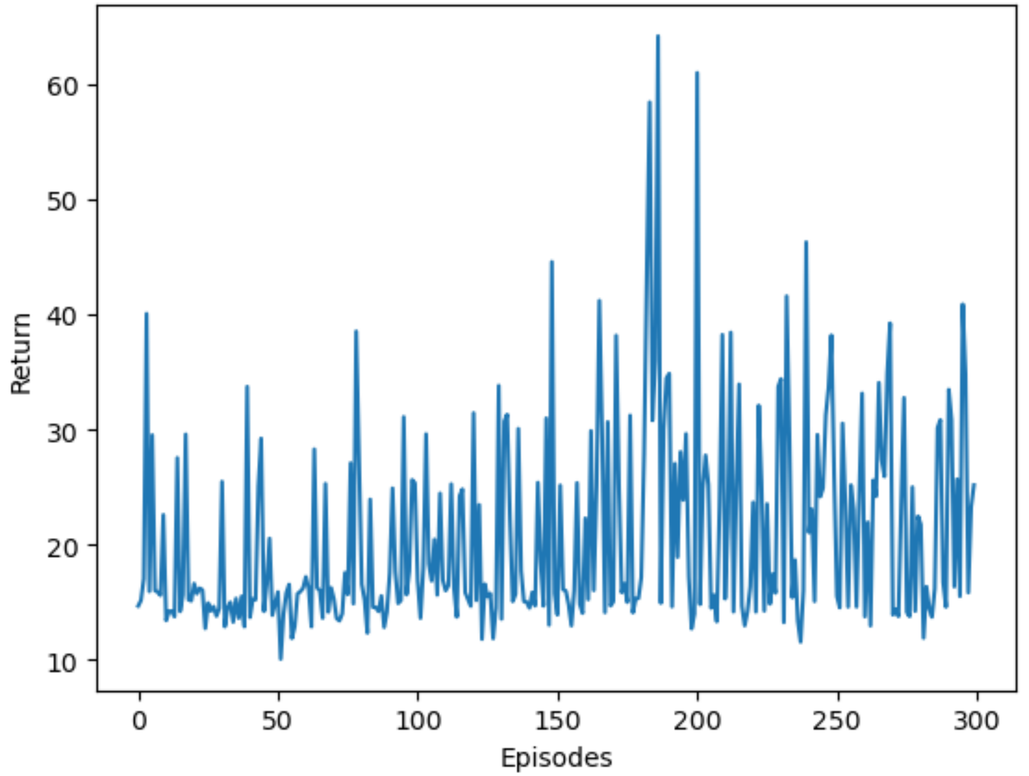
Results

Results

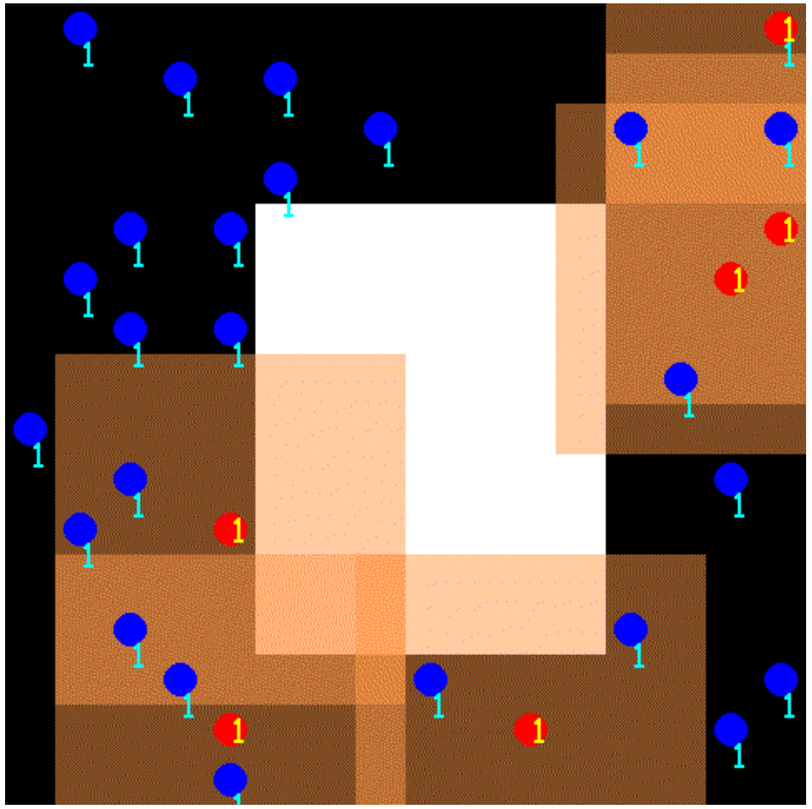


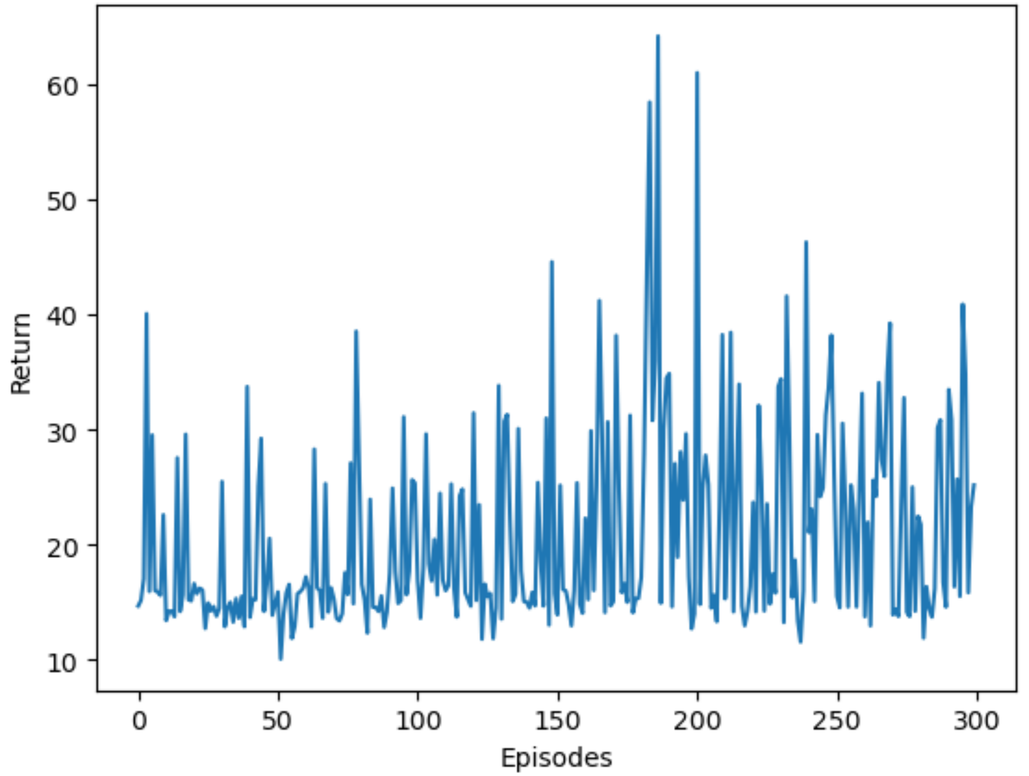
Results





Results





Batch-size = 32, Discount factor = 1, Eps range: 1-0.1
Episodes: 300, Optimizer: RMSProp(lr = 0.00025, alpha = 0.95), Replay Memory Size: 1000000
Target network updated after 5000 iterations

Independent DQN-2

Independent DQN-2

- Define the networks and agents

Independent DQN-2

- Define the networks and agents
- Initialize a single network for all agents(as agents are identical, parameter sharing)

Independent DQN-2

- Define the networks and agents
- Initialize a single network for all agents(as agents are identical, parameter sharing)
- Define the shared Replay buffer

Independent DQN-2

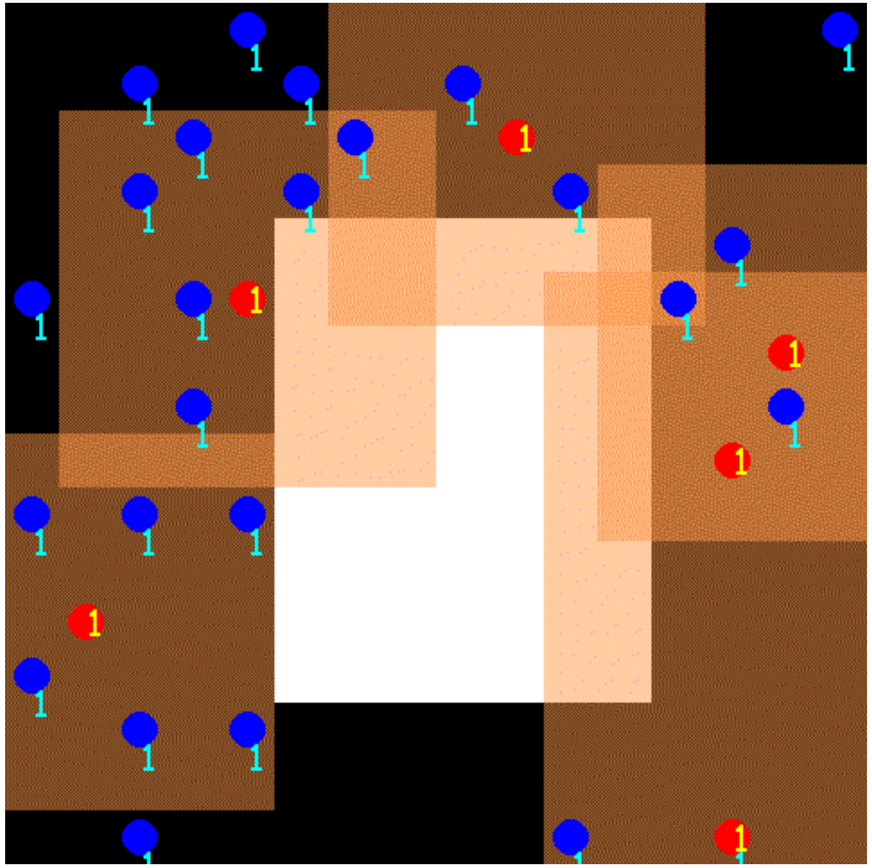
- Define the networks and agents
- Initialize a single network for all agents(as agents are identical, parameter sharing)
- Define the shared Replay buffer
- Implement the training loop (each agent learns independently)

Independent DQN-2

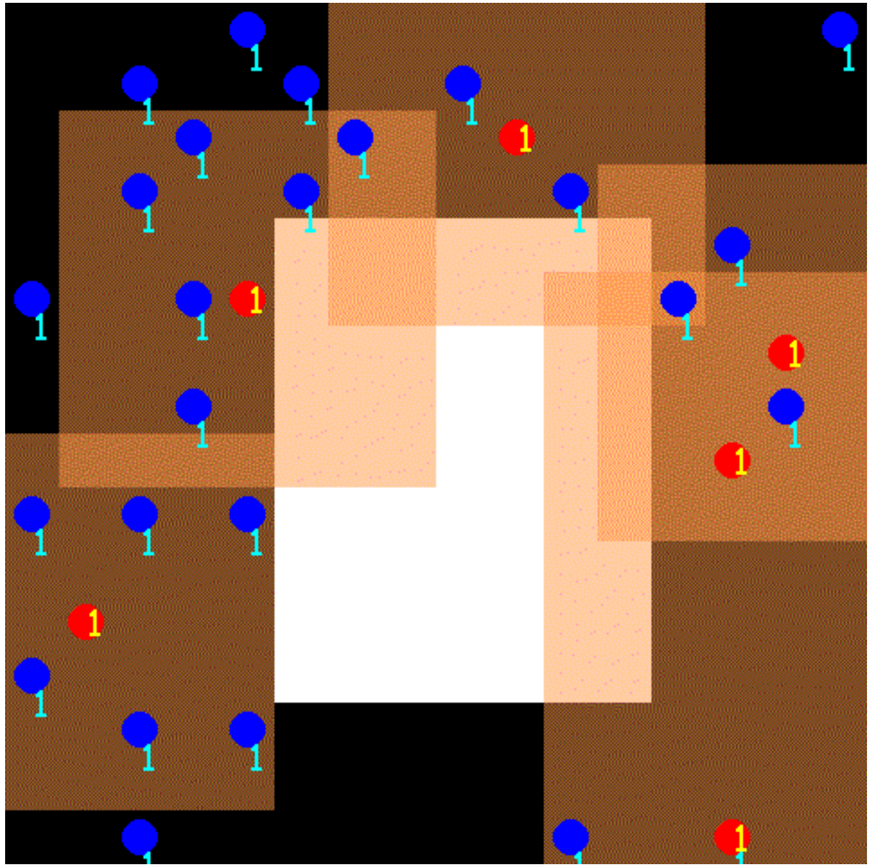
- Define the networks and agents
- Initialize a single network for all agents(as agents are identical, parameter sharing)
- Define the shared Replay buffer
- Implement the training loop (each agent learns independently)
- Implement the exploration strategy

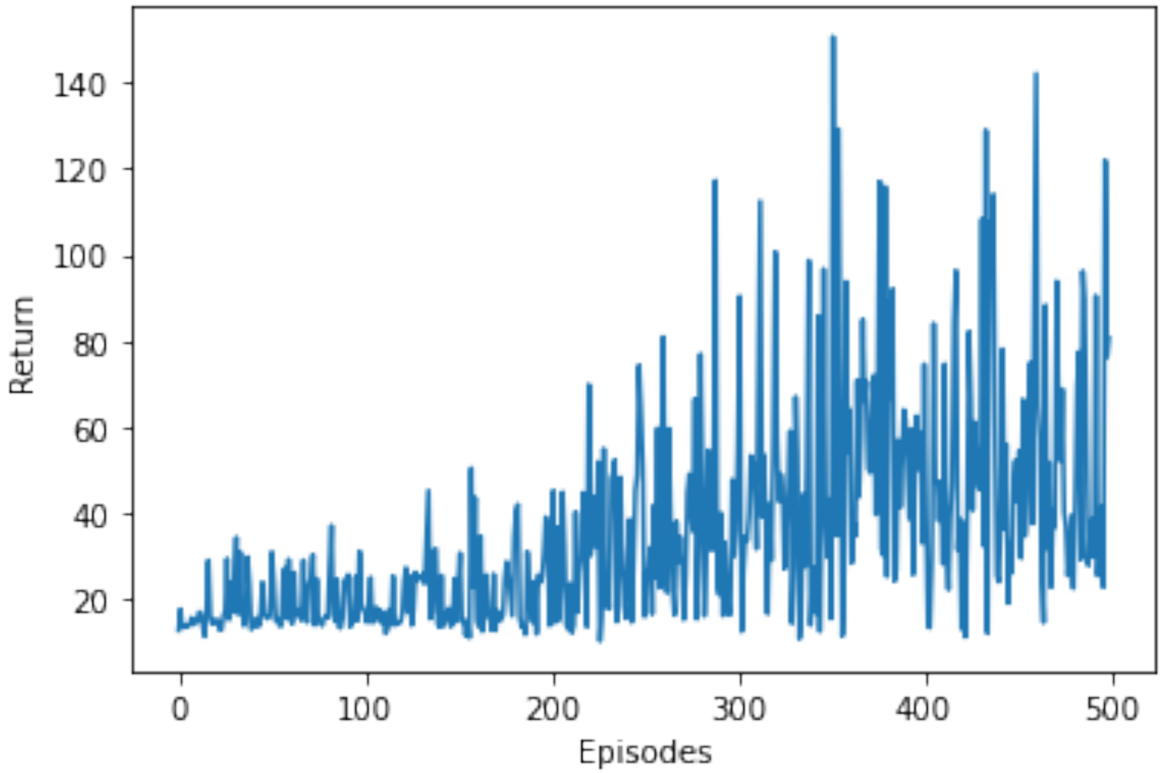
Results

Results

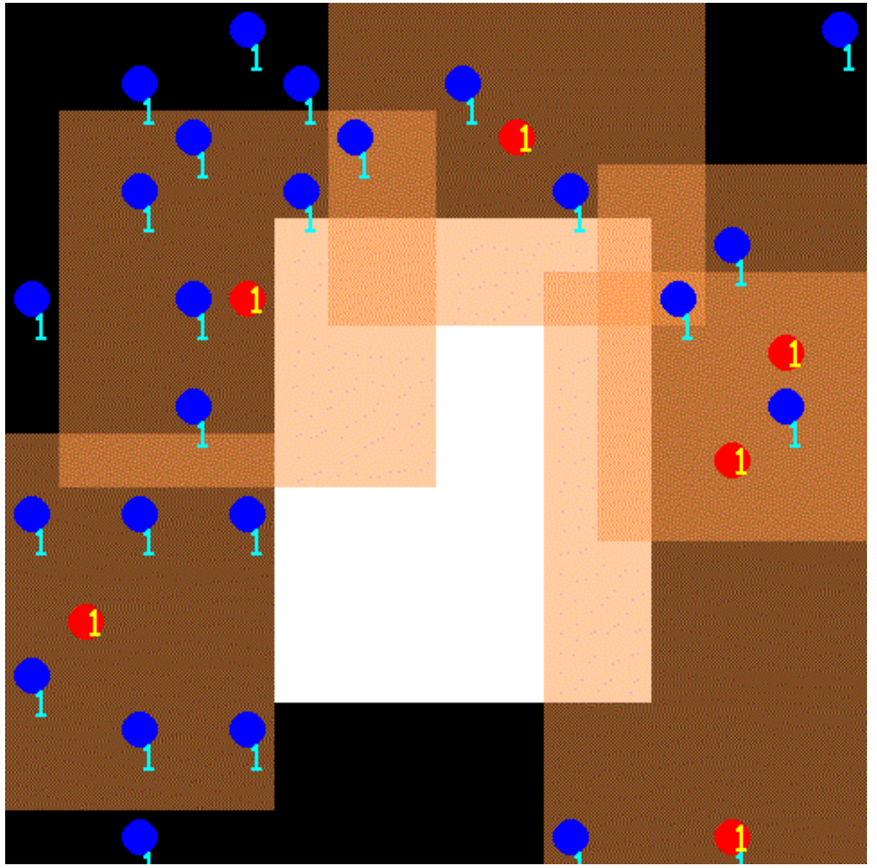


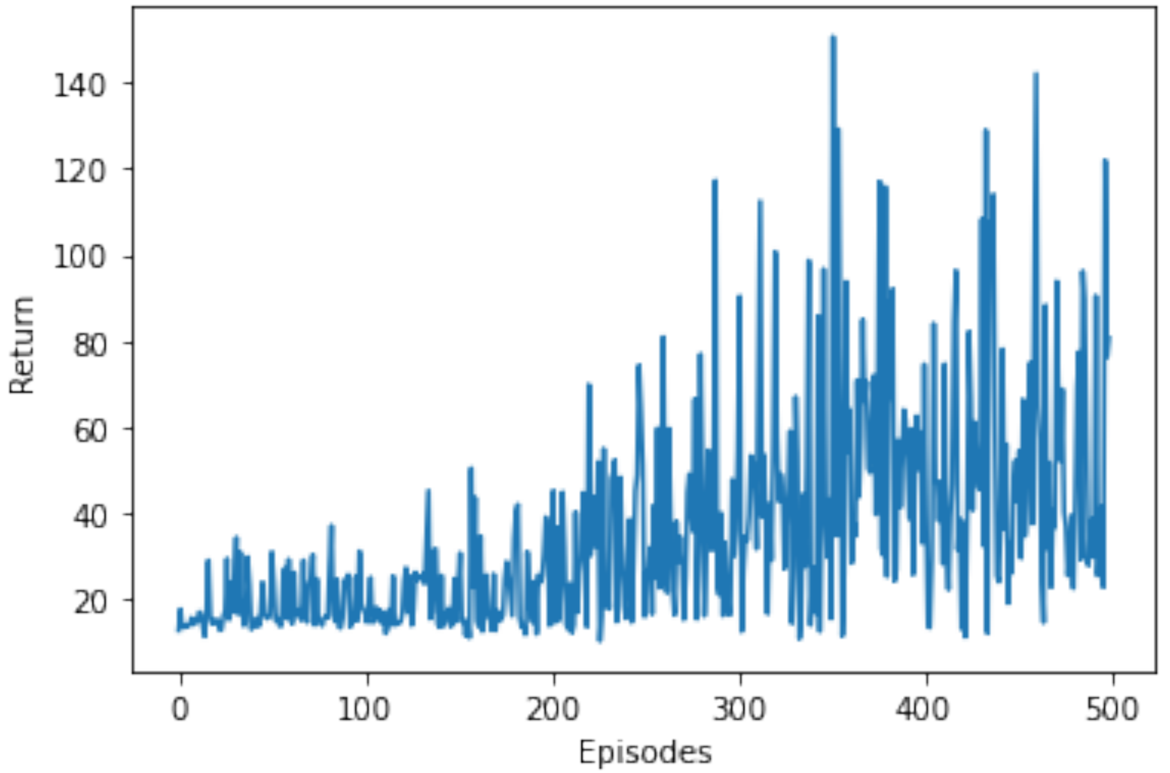
Results





Results

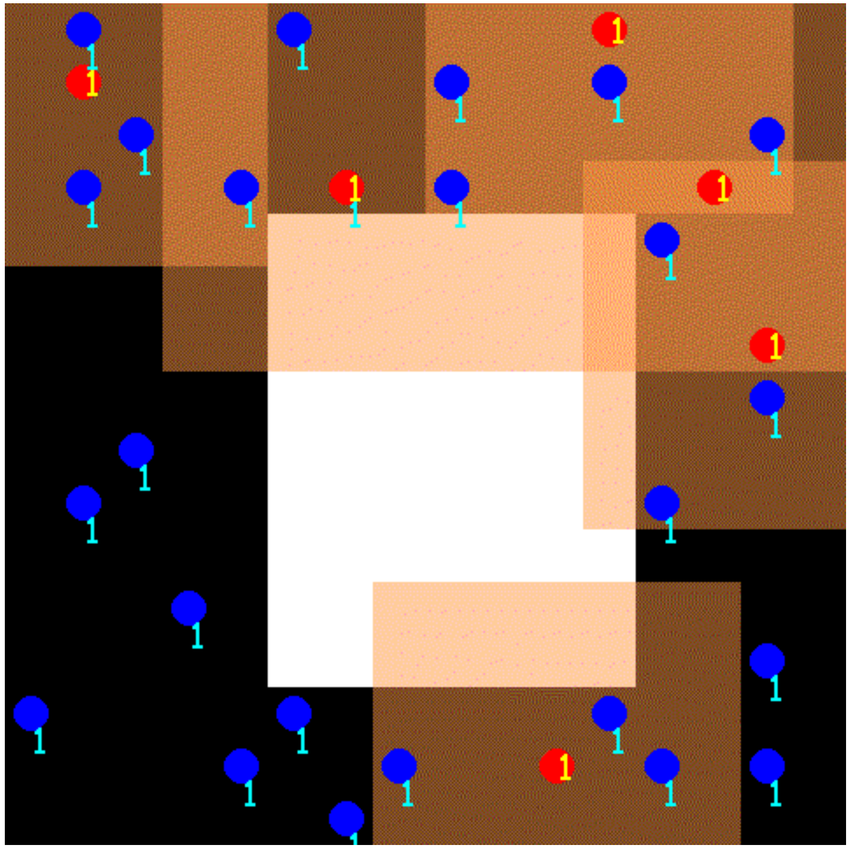




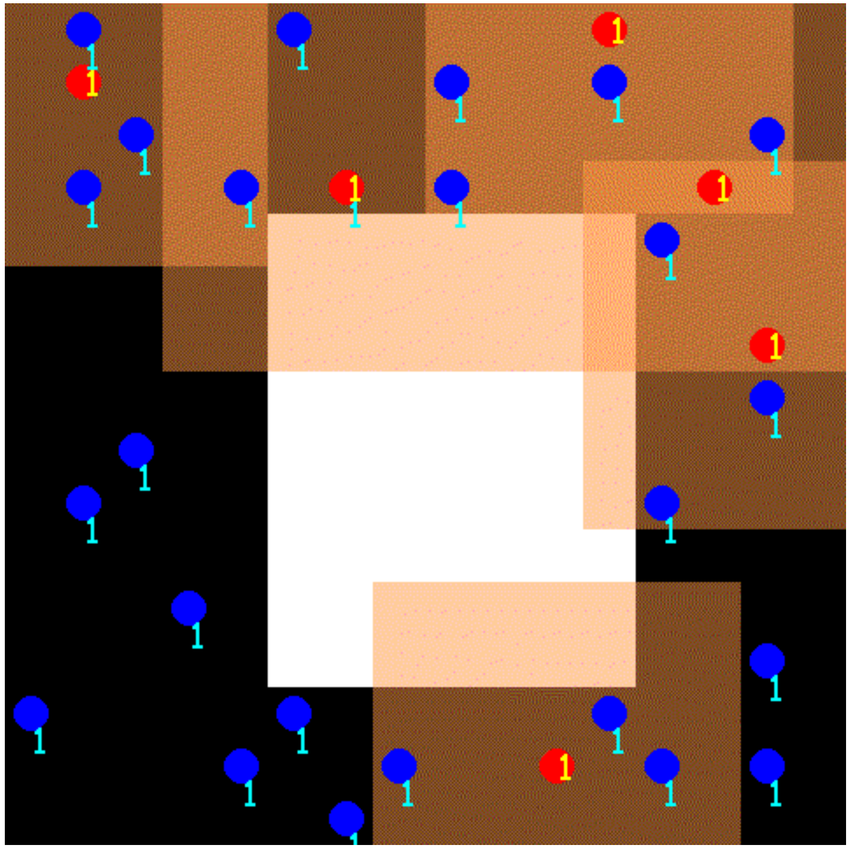
Batch-size = 32, Discount factor = 1, Eps range: 1-0.1
Episodes: 500, Optimizer: RMSProp(lr = 0.00025, alpha = 0.95), Replay Memory Size: 1000000
Target network updated after 5000 iterations

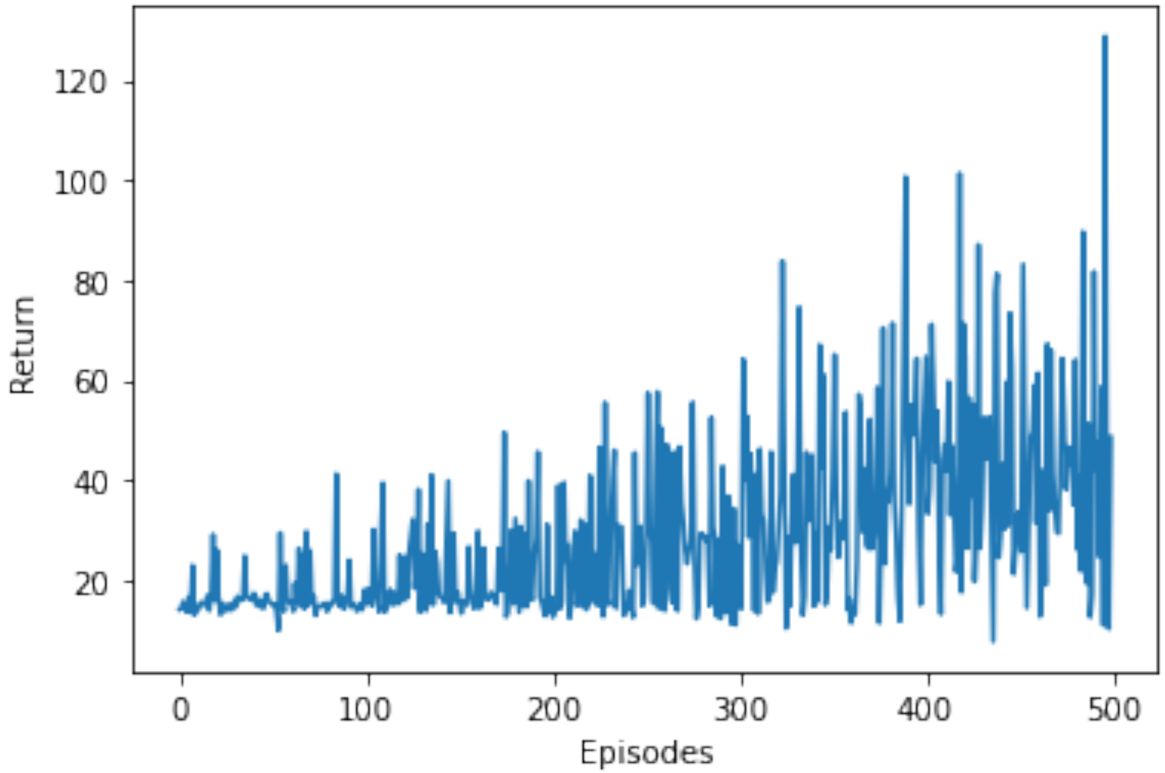
Results (Evaders Frozen)

Results (Evaders Frozen)

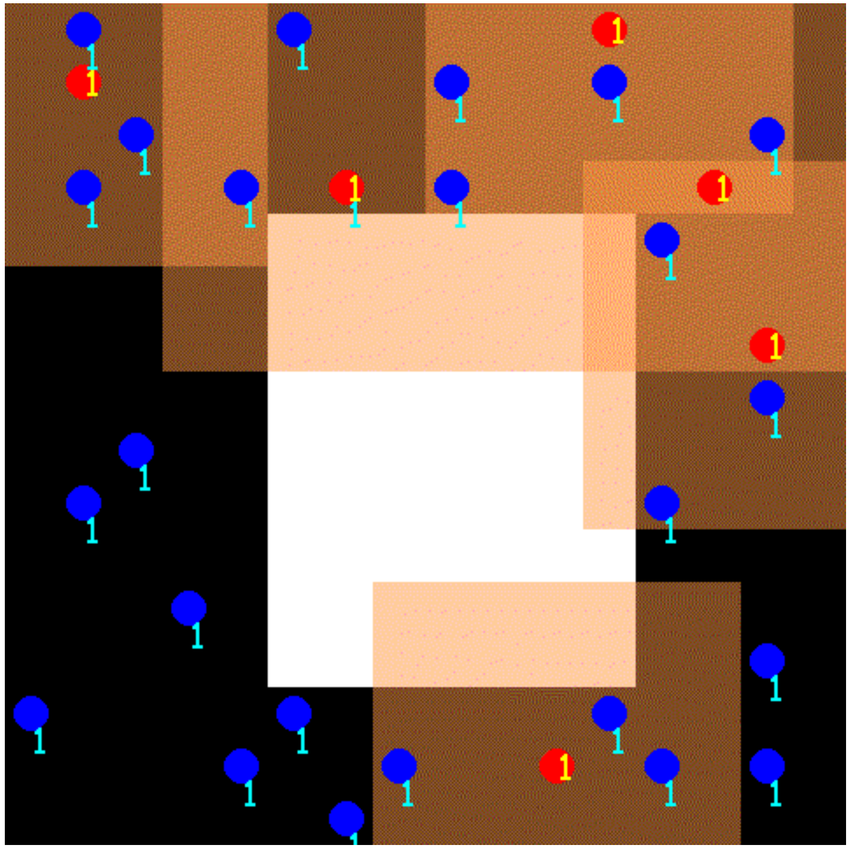


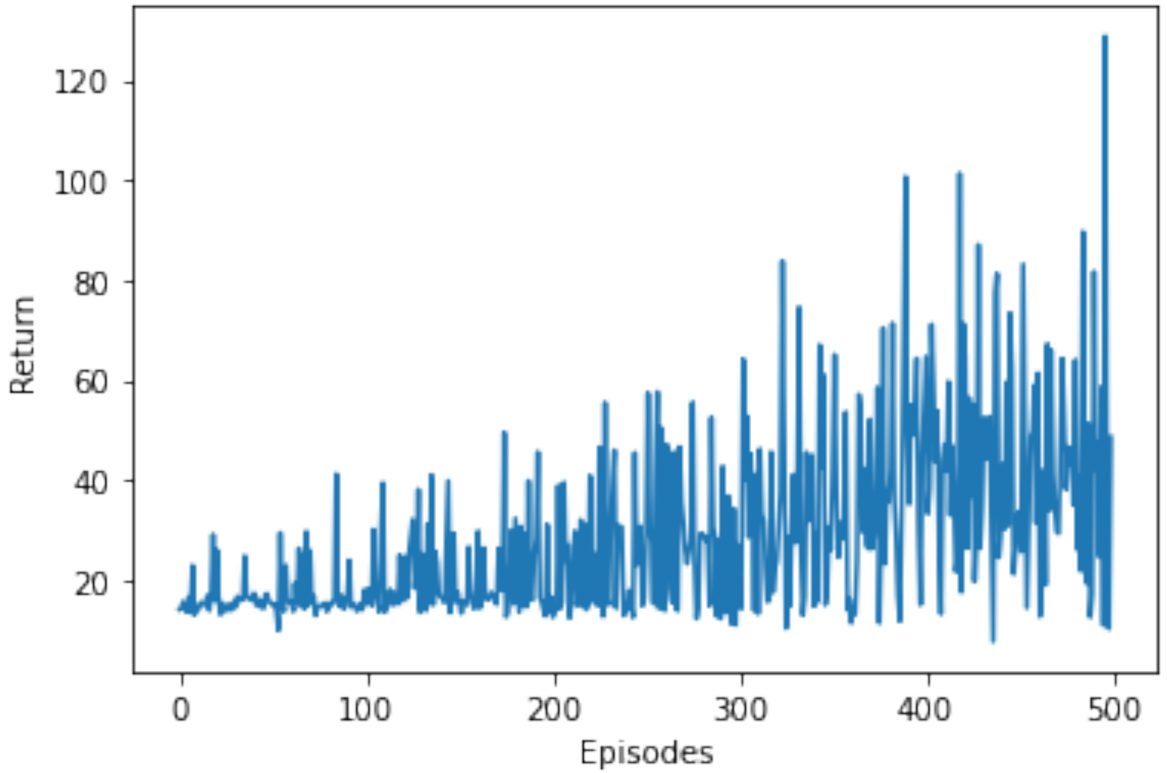
Results (Evaders Frozen)





Results (Evaders Frozen)





Batch-size = 32, Discount factor = 1, Eps range: 1-0.1
Episodes: 500, Optimizer: RMSProp(lr = 0.00025, alpha = 0.95), Replay Memory Size: 1000000
Target network updated after 5000 iterations

- **Problem:**

  Multi-agent deep RL for pursuit game

- **Problem:**

  Multi-agent deep RL for pursuit game

- **Applied:**

  Independent DQN.

- **Problem:**

  Multi-agent deep RL for pursuit game

- **Applied:**

  Independent DQN.

- **Key Learning:**

  It is very difficult to train independent DQN. Lots of effort in hyperparameter tuning is required even to get simple results.

- J. Terry, B. Black, N. Grammel, M. Jayakumar, A. Hari, R. Sullivan, L. S. Santos, C. Dieffendahl, C. Horsch, R. Perez-Vicente, and others, "Pettingzoo: Gym for multi-agent reinforcement learning," in Advances in Neural Information Processing Systems, vol. 34, pp. 15032-15043, 2021.
- Mnih, V., Kavukcuoglu, K., Silver, D. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015). DOI: https://doi.org/10.1038/nature14236

- J. Terry, B. Black, N. Grammel, M. Jayakumar, A. Hari, R. Sullivan, L. S. Santos, C. Dieffendahl, C. Horsch, R. Perez-Vicente, and others, "Pettingzoo: Gym for multi-agent reinforcement learning," in Advances in Neural Information Processing Systems, vol. 34, pp. 15032-15043, 2021.
- Mnih, V., Kavukcuoglu, K., Silver, D. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015). DOI: https://doi.org/10.1038/nature14236

Thank you!

mcbhatt2@illinois.edu