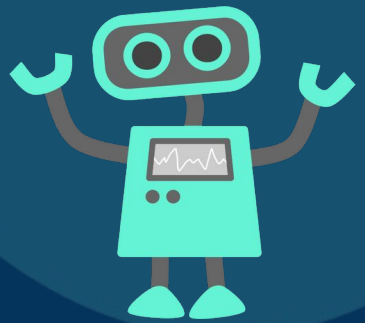
December 8th, 2021

SWARM ROBOTICS

CSE446: TEAM 2

MICHAEL MU, ROMIKA SAIRAM & SACHIN GEORGE



OUR MAIN TOPICS TODAY

Project Overview

Background

Design Process

Demo

Findings

Summary

PROJECT OVERVIEW

- Combination of swarm robotics and reinforcement learning
- Multiple agents are controlled by the same reinforcement learning algorithm
- Agents must cooperate with each other to find a target and move it to a specified location
- Compared the performance of different algorithms

Image retrieved from *The New Stack*

BACKGROUND

- Inspired by the SWARM-BOTS project from 2002
- Search and retrieval is useful after disasters
- DQN - value based method
- A2C - actor-critic method
- Reinforcement learning with swarm robotics is a relatively unexplored area

Youtube: swarm-bots pulling a child
https://www.youtube.com/watch?v=CJOubyiITsE

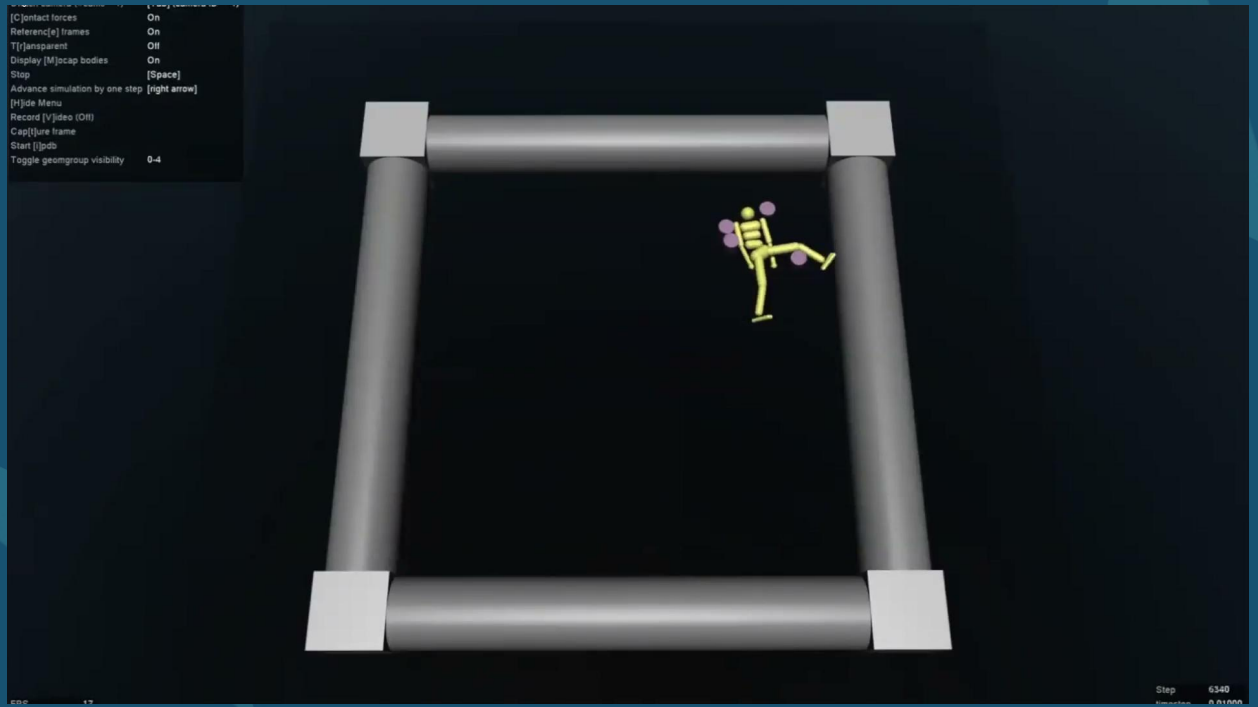
ENVIRONMENT DESIGN

- Environment: custom designed MuJoCo simulation
  - Cooperative multi-agent
  - Partially observable
  - Decentralized
  - Once found, position and velocity of agents and targets are provided to agents
- Rewards:
  - Small negative for accomplishing nothing
  - Reward for finding and moving towards target
  - Positive or negative reward based on distance of target to a specified location

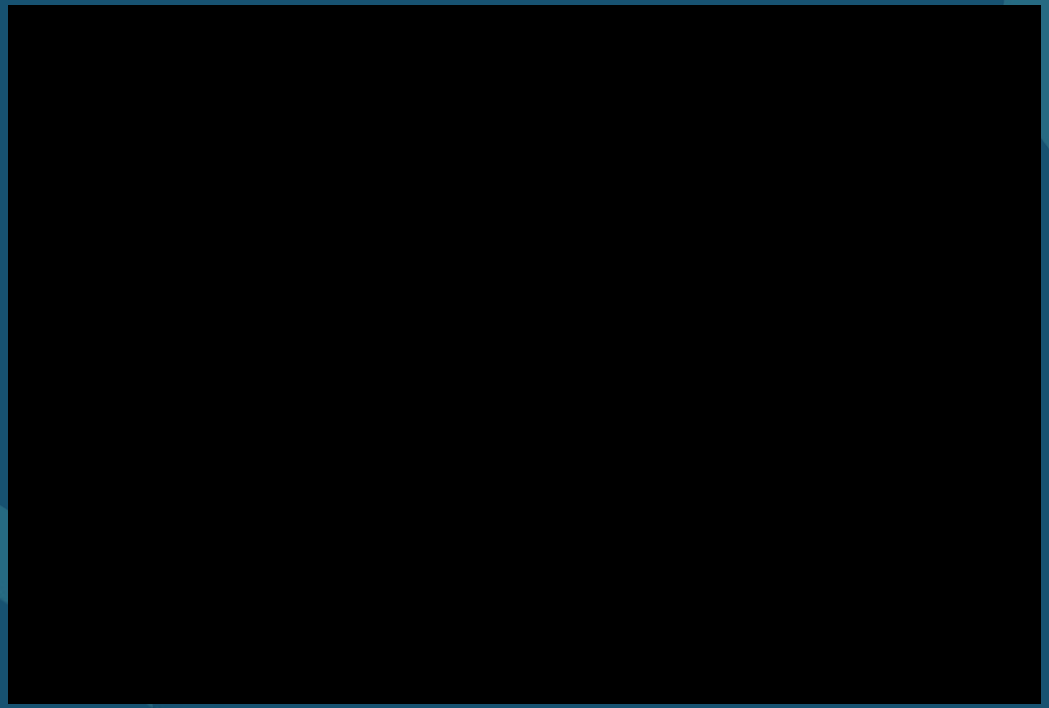
AGENT DESIGN

- Agents:
  - Placed in random location near the starting area
  - Implemented with DQN and A2C algorithms
  - Goal: Find target
- Goal:
  - Find the target/victim
  - Navigate to the target
  - Push the target to a specified destination

DEMO [A2C]


[C]ontact forces On
Referenc[e] frames On
T[r]ansparent Off
Display [M]ocap bodies On
Stop [Space]
Advance simulation by one step [right arrow]
[H]ide Menu
Record [V]ideo (Off)
Cap[t]ure frame
Start [i]pdb
Toggle geomgroup visibility 0-4
Step 6340

DEMO [DQN]



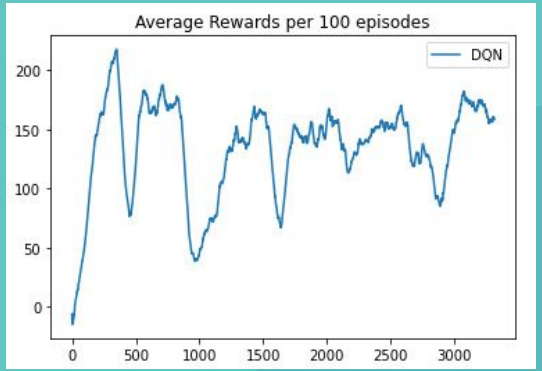
TRAINING RESULTS

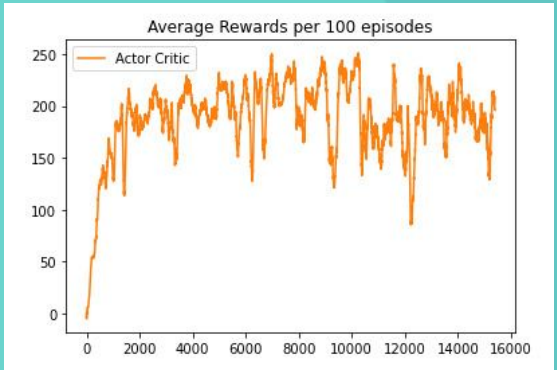
DQN

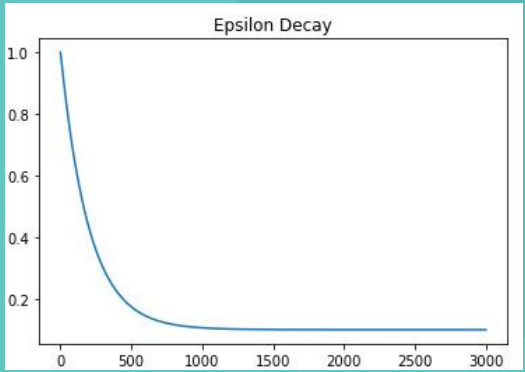
- Faster convergence
- Performance drops off after convergence

A2C

- Slower convergence
- Performance oscillates upwards

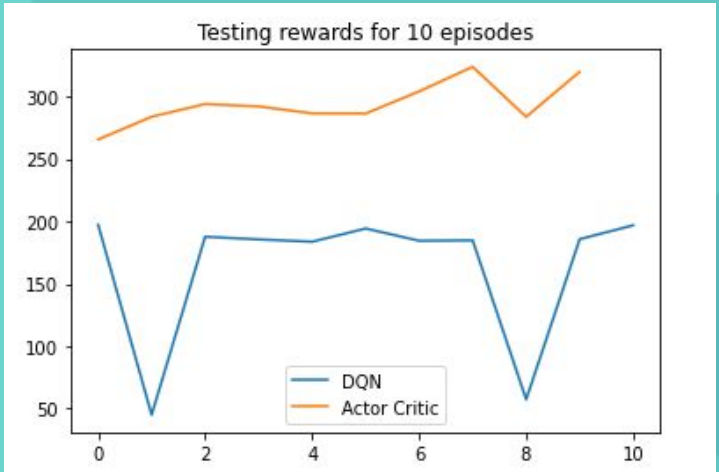






OPTIMAL EVALUATION

- Final A2C performed better and was able to synthesize a plan
- DQN was less prone to complete failure



SUMMARY

- A2C backed agent was able to accomplish its task very often but was less reliable and more prone to complete failure
- DQN works well for location finding and general initial exploration, but deriving a long-term plan in a continuous environment with random motions is unlikely
- Decentralized multi-agents must learn to deal with more noise in their rewards from the actions of other agents

CONTRIBUTION SUMMARY

Michael Mu:

Sachin George:

Romika Sairam:

Thank you for your time and attention!