Multi-task Reinforcement Learning for Physical Reasoning

Gino Kwun, Anupam Patil, Eric Liu, Shardul Datar, Rahul Nair, Ayan Bhowmick

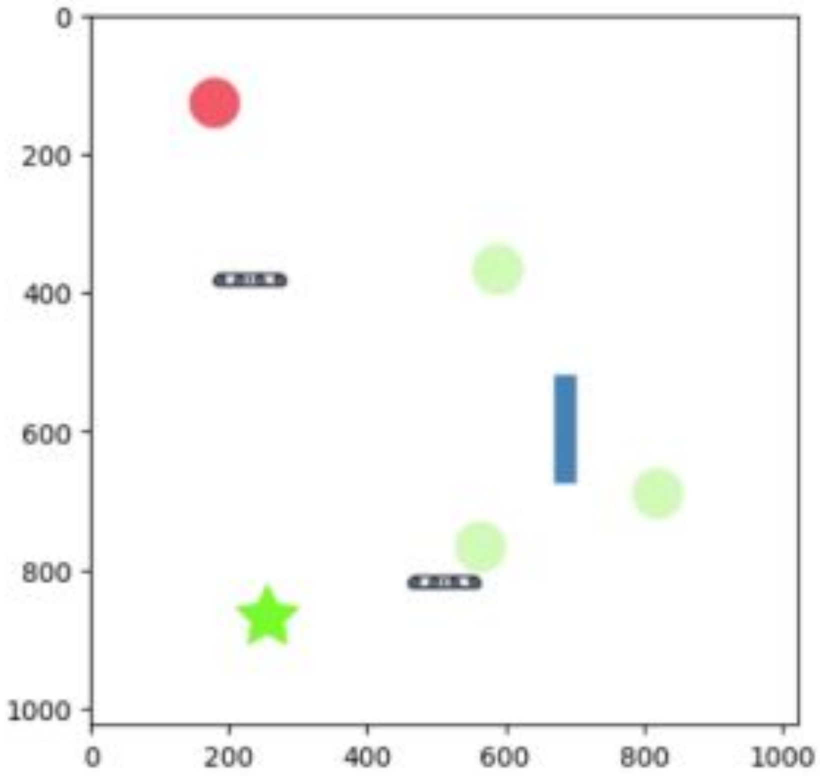
TA: Ayush Jain

Introduction

- Reinforcement learning (RL) has demonstrated remarkable success in various domains, but its application to physical reasoning remains a challenging frontier.

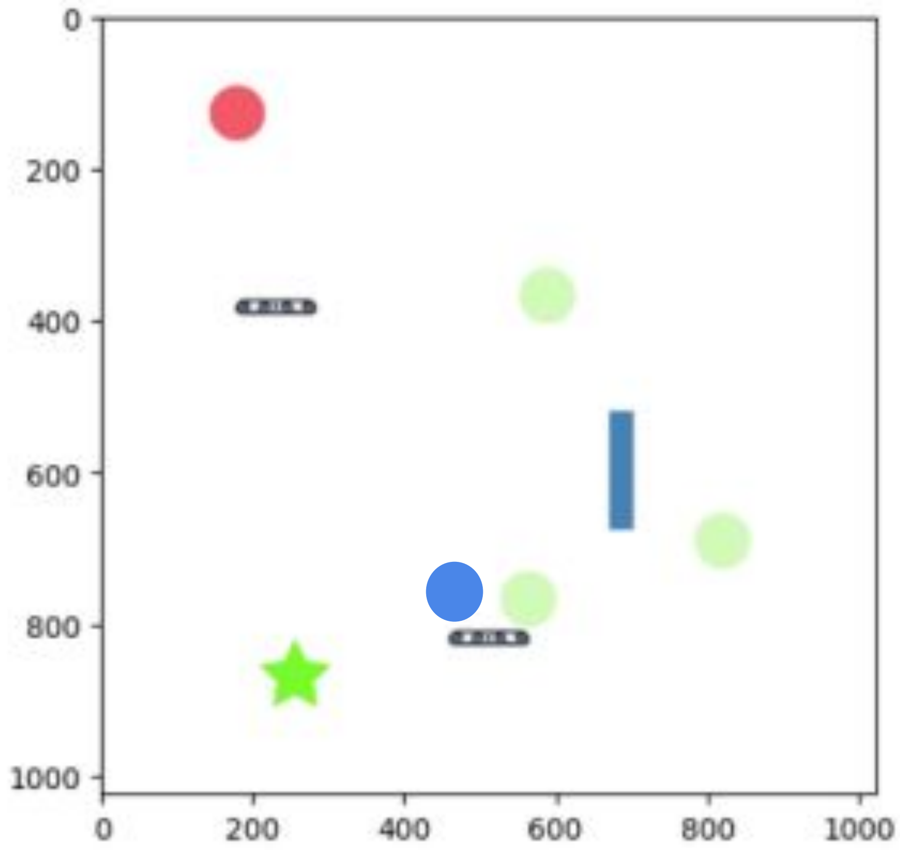
- Project focus: training a multi-task reinforcement learning agent for physical reasoning using the CREATE (Chain Reaction Tool Environment) environment.
  - The ability to reason about physical interactions and dynamics is crucial for tasks such as robotic manipulation, autonomous navigation, and game playing.
  - Multi-task learning is useful for generalizing a model to multiple environments. Training tasks together brings benefits from the shared useful information across different tasks and often achieves higher performance than single-task learning.

CREATE: Chain Reaction Tool Environment



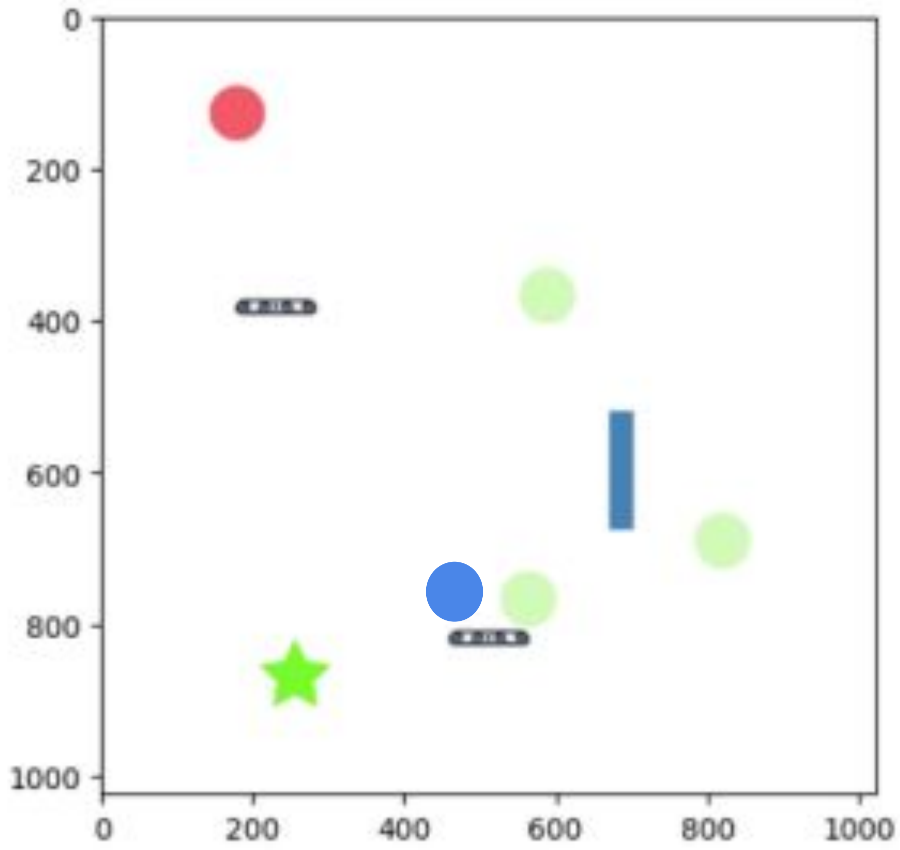
- The CREATE (Chain Reaction Tool Environment) provides a comprehensive simulation platform for physical reasoning tasks.
- The environment consists of essential components such as a source ball (red), a target ball (black), and a target goal (green star).
- These elements are complemented by a variety of interactive objects including walls, floors, belts, see-saws, trampolines, buckets, cannons, funnels, ladders, and obstacle balls. Each object introduces unique dynamics and interactions, challenging the reinforcement learning agent to navigate the target ball towards the goal using the available assets.

CREATE: Chain Reaction Tool Environment



- Essential components:
  - source ball (red)
  - target ball (blue/black)
  - target goal (green star)
- Variety of interactive objects (tools):
  - Walls, belts, see-saws, trampolines, buckets, cannons, funnels, ladders, and obstacle balls.
  - Each object introduces unique dynamics and interactions.
- Goal:
  - Navigate the target ball towards the goal using the available tools.

CREATE: Chain Reaction Tool Environment



- Observation Space
  - 84 x 84 x 3 image
- Action Space: adding a tool to the environment (Dict)
  - Index of the tool (Discrete)
  - Position (x, y) of the tool (Box)

CREATE: Introducing Randomness

To incorporate randomness in the existing environments and introduce new levels of difficulty, we have prepared four versions of each environment, depending on the noise added to the coordinates of the target, goal, secondary goal and the objects in the environment. These are as follows:-

DEF_NOISE:- The least amount of noise or randomness. Set as 0.1 in the file create_level.py.

MID_NOISE:- Higher amount of noise or randomness. Set as 0.2 in the file create_level.py.

HIGH_NOISE:-Higher compared to MID_NOISE. Set as 0.3 in file create_level.py.

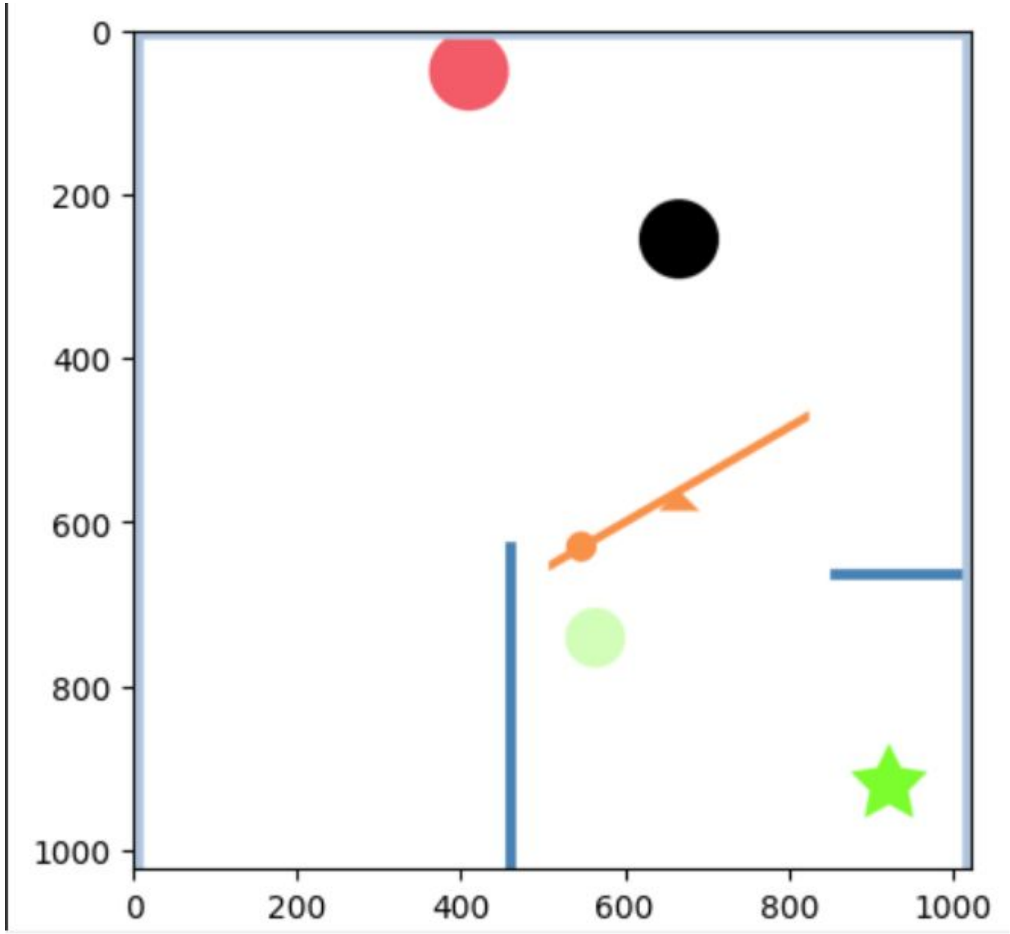
LARGE_NOISE:- Highest amount of noise. Changes position of objects drastically. Set as 0.5 in file create_level.py.

These types of noises also represent the difficulty level of the environments with DEF_NOISE being Level 1, MID_NOISE being Level 2 and so on. Uniform function is used to introduce randomness to the positions.

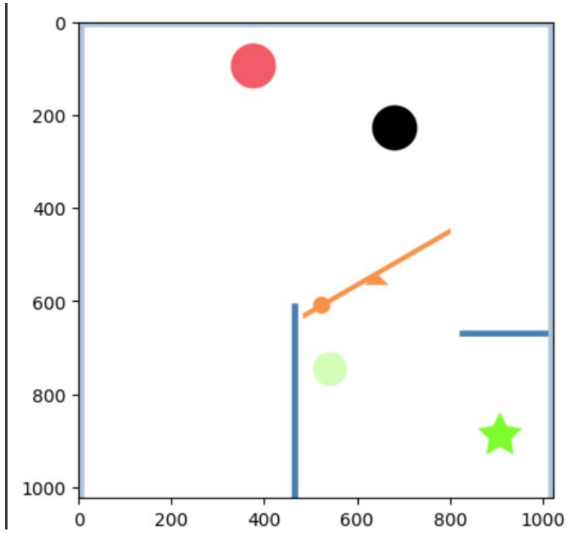
CREATE: Introducing Randomness

- To incorporate randomness in the existing environments and introduce new levels of difficulty, we have prepared four versions of each environment:
  - **DEF_NOISE**: The least amount of noise or randomness.
  - **MID_NOISE**: Higher amount of noise or randomness.
  - **HIGH_NOISE**: Higher than MID_NOISE.
  - **LARGE_NOISE**: Highest amount of noise. Changes position of objects drastically.

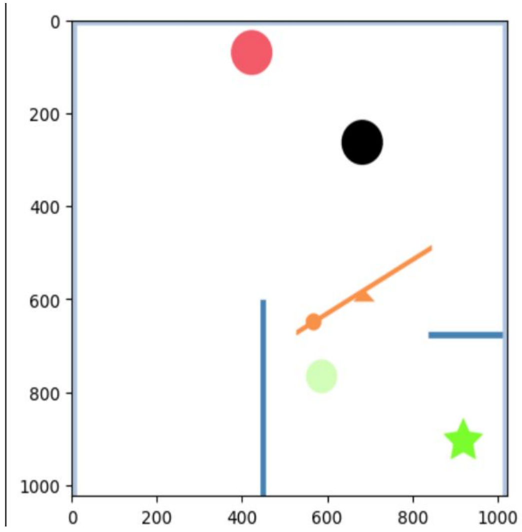
- Types of noises represent the difficulty level of the environments with DEF_NOISE being Level 1, MID_NOISE being Level 2 and so on.
  - Uniform function is used to introduce randomness to the positions.



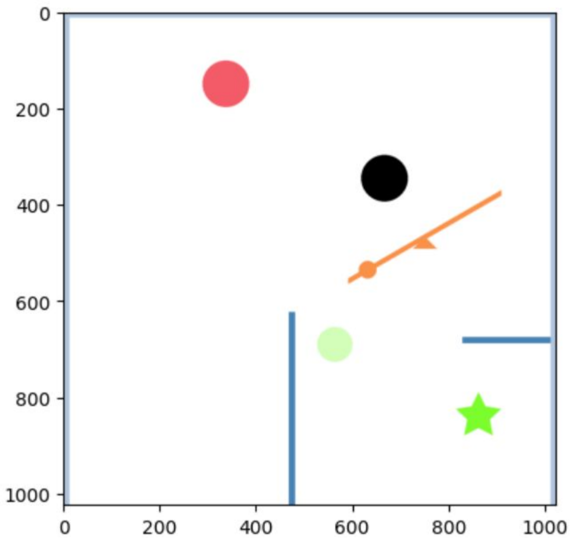
ORIGINAL SEESAW ENV



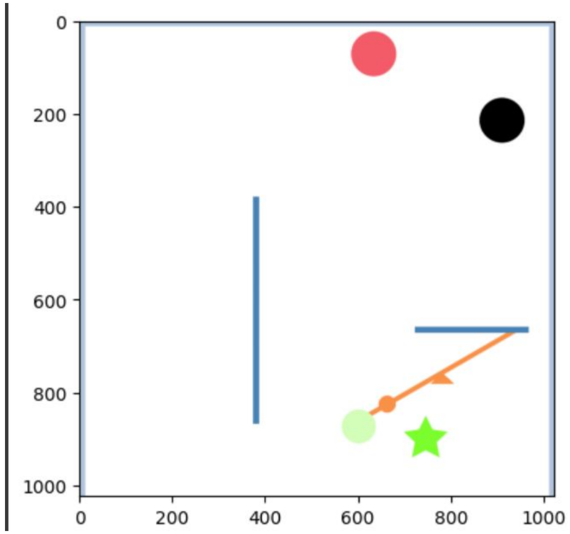
SEESAW ENV WITH DEF_NOISE



SEESAW ENV WITH MID_NOISE



SEESAW ENV WITH HIGH_NOISE

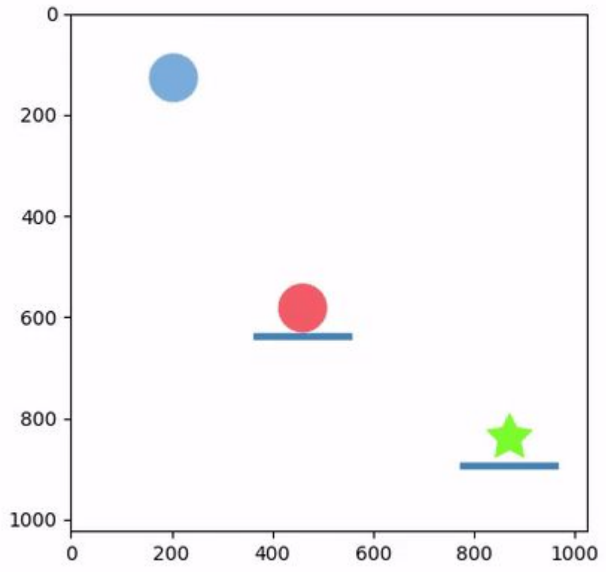


SEESAW ENV WITH LARGE_NOISE

Approach

- Build upon new-actions-rl codebase developed by Ayush Jain, the TA in charge of our project.
  - Takes advantage of multi-core processing techniques.
  - Randomize/create new environments.

- Failure: Stable-Baselines 3
  - Unable to handle CREATE's unique action space



Implementation

- Basic Multi-task RL Agent
  - 2 similar environments
    - Performs pretty well across both environments
  - 3 different environments
    - Performs less well
    - Tends to overfit to a subset of the environments

Multi-task RL for Similar Environments

Bucket level

Push level

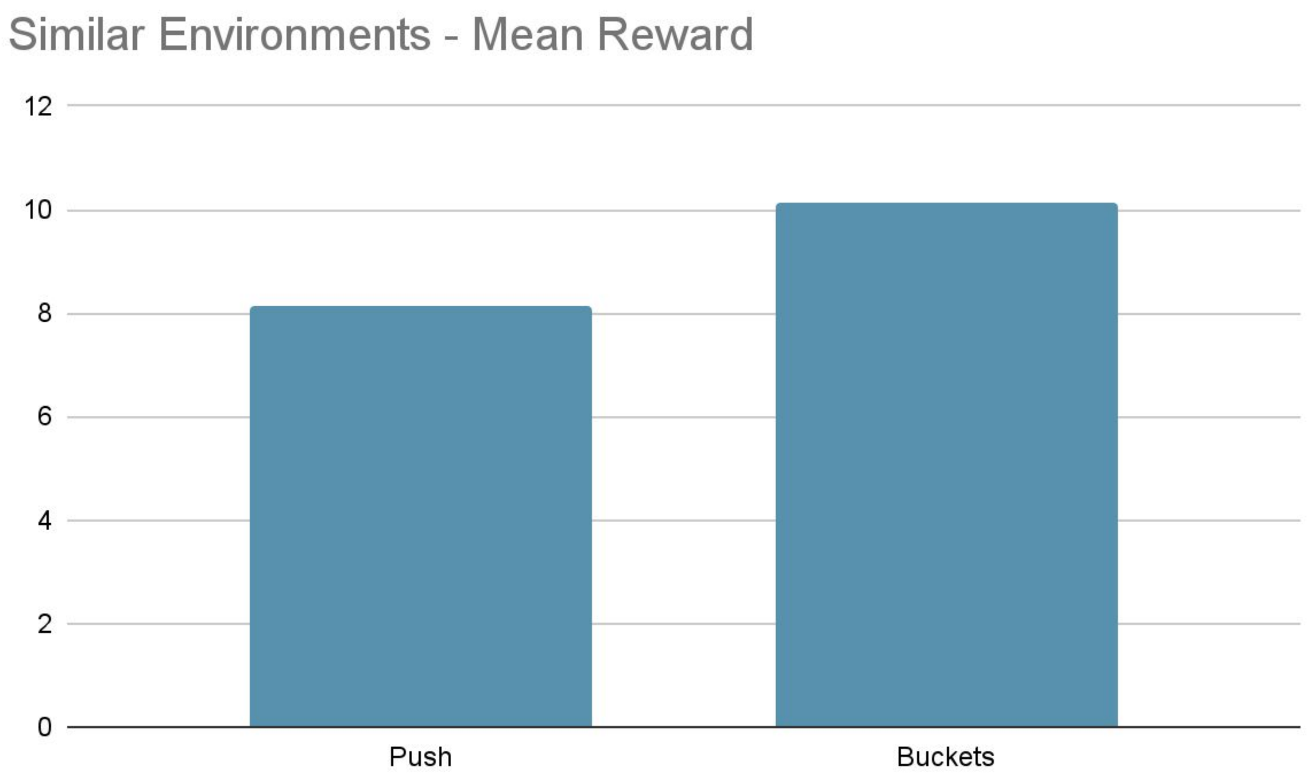
Multi-task RL for Different Environments

Push level

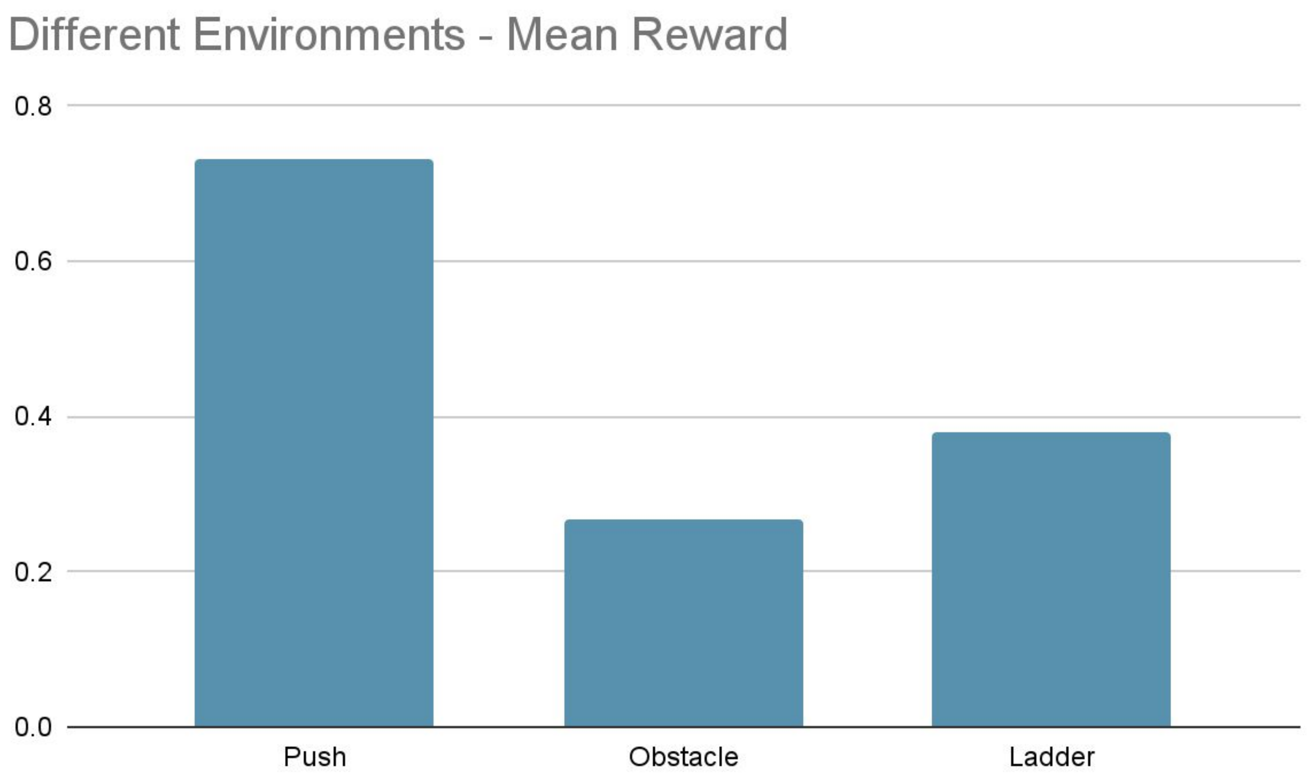
Ladder level

Obstacle level

Results



Results

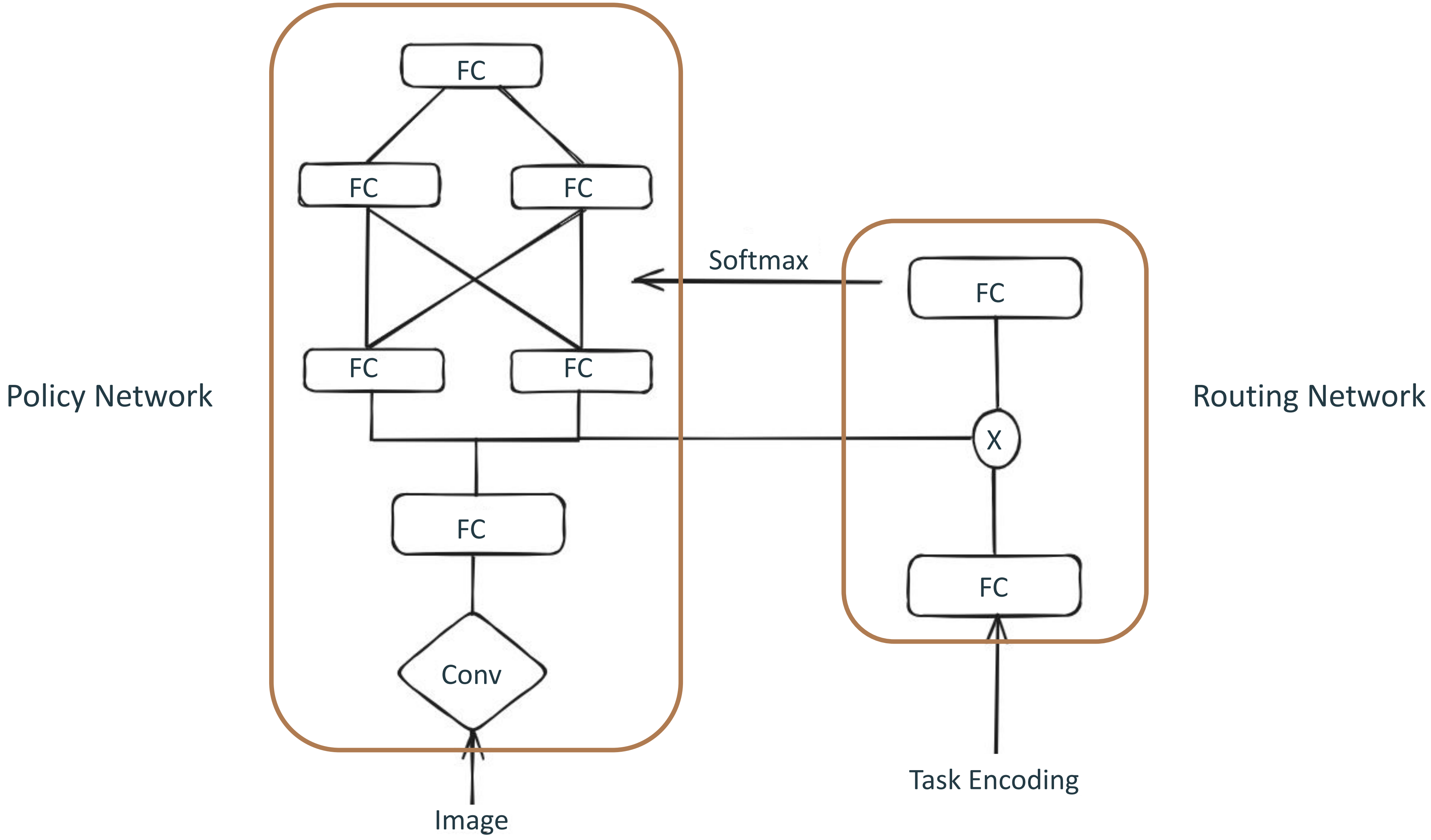


Different Environments - Mean Reward
0.8
0.6
0.4
0.2
0.0
Push
Obstacle
Ladder

Implementation

- More advanced approaches:
  - Gradient Surgery
    - Alleviate conflict between two opposing environments.
    - Currently tuning on CARC.
  - Soft modularization
    - Policy network with multi-layers and multi-modules of fully connected networks
    - Introduce additional routing network with multi-task embeddings
      - Similar to attention
    - Currently tuning on CARC

Soft Modularization Model



Future Work

- Explore impact of training multiple environments at once.
  - Measure similarity of environments. What makes environments similar, and does this reduce training time? Can we reduce training time of two very different environments by also including a “middle-point” environment?

- Expand to training on more “difficult” environments.
  - Combine different environments into one new environment.
  - Introduce new environments and new interactive objects.

- Continue exploring more advanced multi-task RL techniques.
  - Distral
  - Popart and Impala

Conclusion

In conclusion, our project has demonstrated the potential of multi-task reinforcement learning for physical reasoning in the CREATE environment. By leveraging environment variations and innovative algorithmic approaches, we have made significant strides towards developing a robust and adaptable reinforcement learning agent. The challenges encountered and insights gained throughout this project will serve as valuable lessons for future research endeavors in the field. We are excited about the possibilities that lie ahead and remain committed to pushing the boundaries of reinforcement learning for real-world applications.

Conclusion

- Our project has demonstrated the potential of multi-task reinforcement learning for physical reasoning in the CREATE environment.
    - Made progress towards developing a robust and adaptable reinforcement learning agent.
    - Works well for similar environments.

- Explored more advanced multi-task RL techniques and will finish constructing final models for the final report
    - Random Environment + Soft Modularization + Gradient Surgery

Q & A

Thanks for listening!

Works Cited

1. Jain, Ayush, Andrew Szot, and Joseph J. Lim. "Generalization to new actions in reinforcement learning." *arXiv preprint arXiv:2011.01928* (2020)
2. Teh, Yee, et al. "Distral: Robust multitask reinforcement learning." *Advances in neural information processing systems* 30 (2017)
3. Vithayathil Varghese, Nelson, and Qusay H. Mahmoud. "A survey of multi-task deep reinforcement learning." Electronics 9.9 (2020): 1363.
4. Yu, Tianhe, et al. "Gradient surgery for multi-task learning." *Advances in Neural Information Processing Systems* 33 (2020): 5824-5836.
5. Yang, Ruihan, et al. "Multi-task reinforcement learning with soft modularization." Advances in Neural Information Processing Systems 33 (2020): 4767-4777.