Long-term Planning with Implicit 3D Feature Tensors

Outline

- ❑ **Motivation**
- ❑ Current Approaches
- ❑ Proposed Approach & Progress
- ❑ Future Work

Motivation

Most of the complex tasks are a composition of several basic skills

Motivation

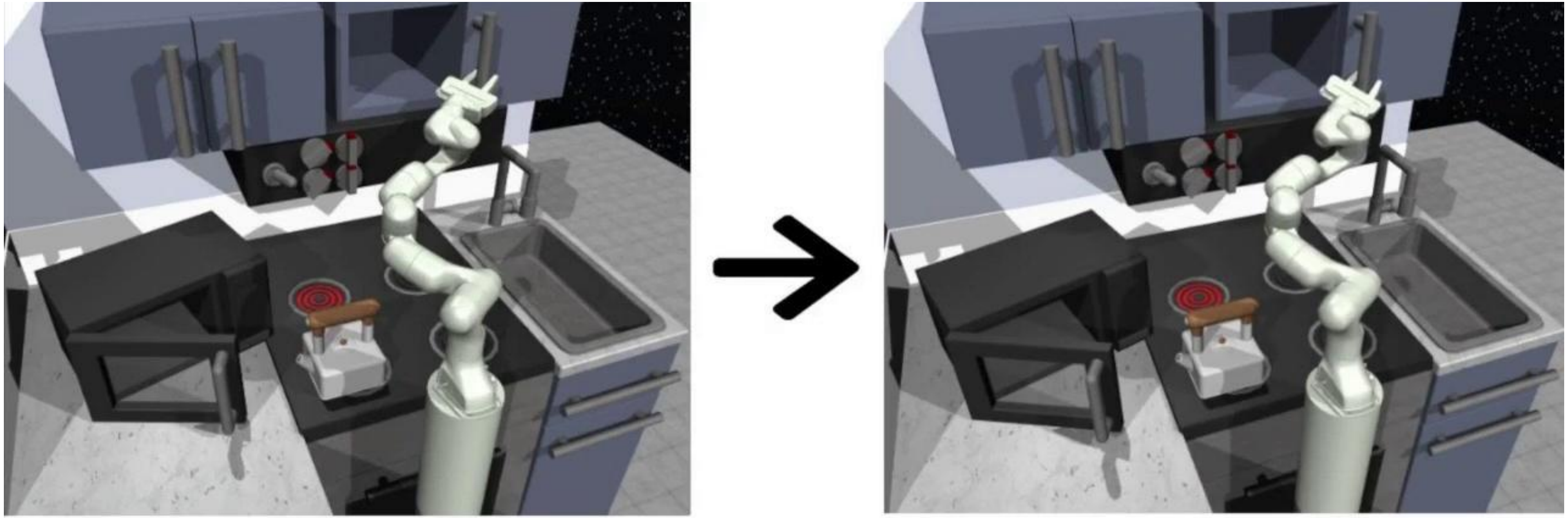
Efficiently learn long-horizon goal-conditioned policies

- Challenges
    - Model-free RL algorithms struggle to perform well on temporally extended tasks
        - Sample inefficient - need lot of data
    - Planning methods requires a good representation of states and goals
        - How to deal with discontinuous switches in the local dynamics of consecutive tasks?
    - Hierarchical RL methods can theoretically scale well to perform long-horizon tasks but struggle with exploration, skill segmentation and reward definition

Outline

- ❑ Motivation
- ❑ **Current Approaches**
- ❑ Proposed Approach & Progress
- ❑ Future Work

Relay Policy Learning *(Gupta, Kumar, Lynch, Levine, and Hausman, CoRL 2019)*



Long Horizon Goal

RPL Policy

Relay Policy Learning *(Gupta, Kumar, Lynch, Levine, and Hausman, CoRL 2019)*

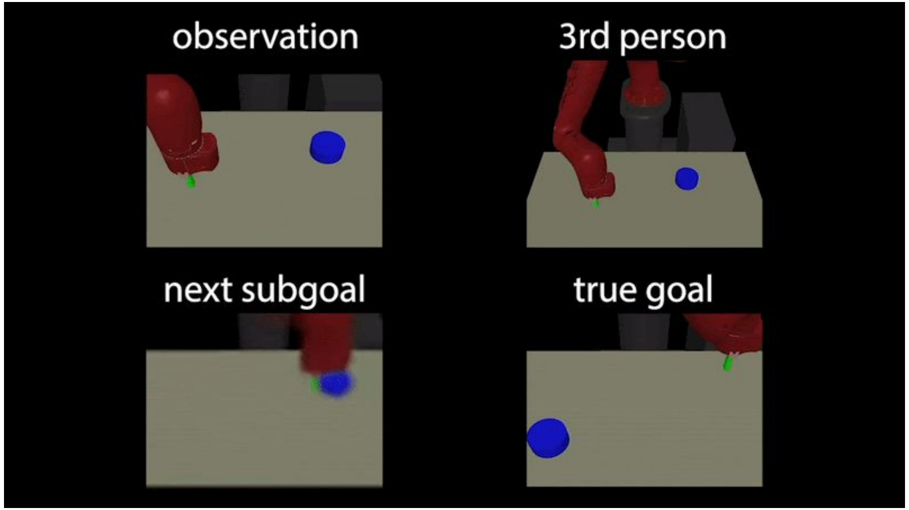
Hierarchical Learning Framework

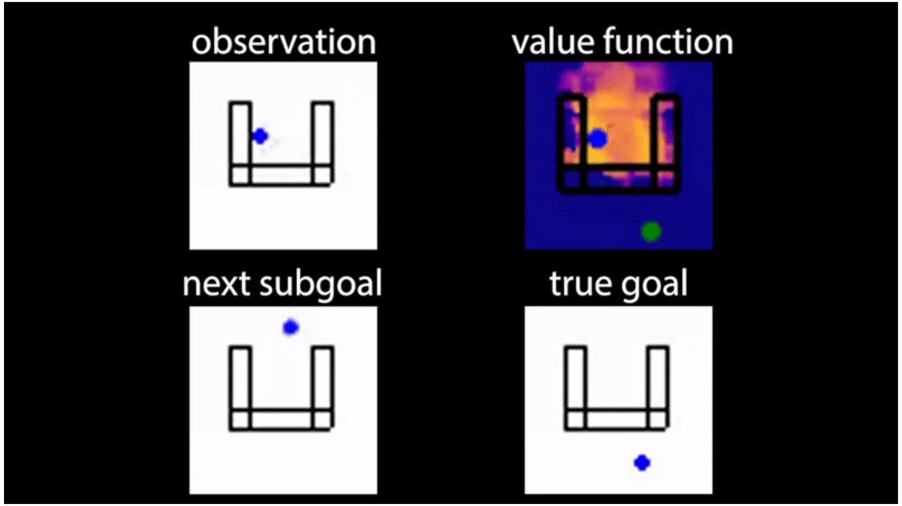
1. Collect an unstructured pool of semantically meaningful demonstrations
2. Re-label the goals in these collected trajectories
3. **Relay Imitation Learning** : Use imitation learning to learn an initial low-level and high-level policies
4. **Relay Policy Learning** : Collect on-policy experiences and update the policy via policy gradient updates
5. Fine-tune on a number of high-level goals and distill these policies into a single multi-goal policy

Limitations

1. No visual input - 30-dimensional state space which consists of positions of the arm and the objects in the scene
    a. Does not account for variation in size and orientation of the objects being manipulated

Planning with Goal-Conditioned Policies *(Nasiriany et. al., NeuRIPS 2019)*





Planning with Goal-Conditioned Policies *(Nasiriany et. al., NeuRIPS 2019)*

Key Idea: Plan over abstraction of states - useful for high-dimensional inputs. Planner focuses on which states to reach whereas goal-conditioned policies figure out how to reach that state.

Latent Embeddings for Abstracted Planning (LEAP)

1. Planner plans intermediate subgoals in a low-dimensional latent space for a given start and goal state.
   a. Low-dimensional latent space of valid states is captured through a VAE trained over a dataset of randomly collected states from the environment.
   b. Each consecutive pair of subgoals is given a feasibility score. The planner **maximizes the overall feasibility score.**
2. The goal-conditioned policy then tries to reach the first subgoal. After t1 time steps, the policy re-plans and repeats.

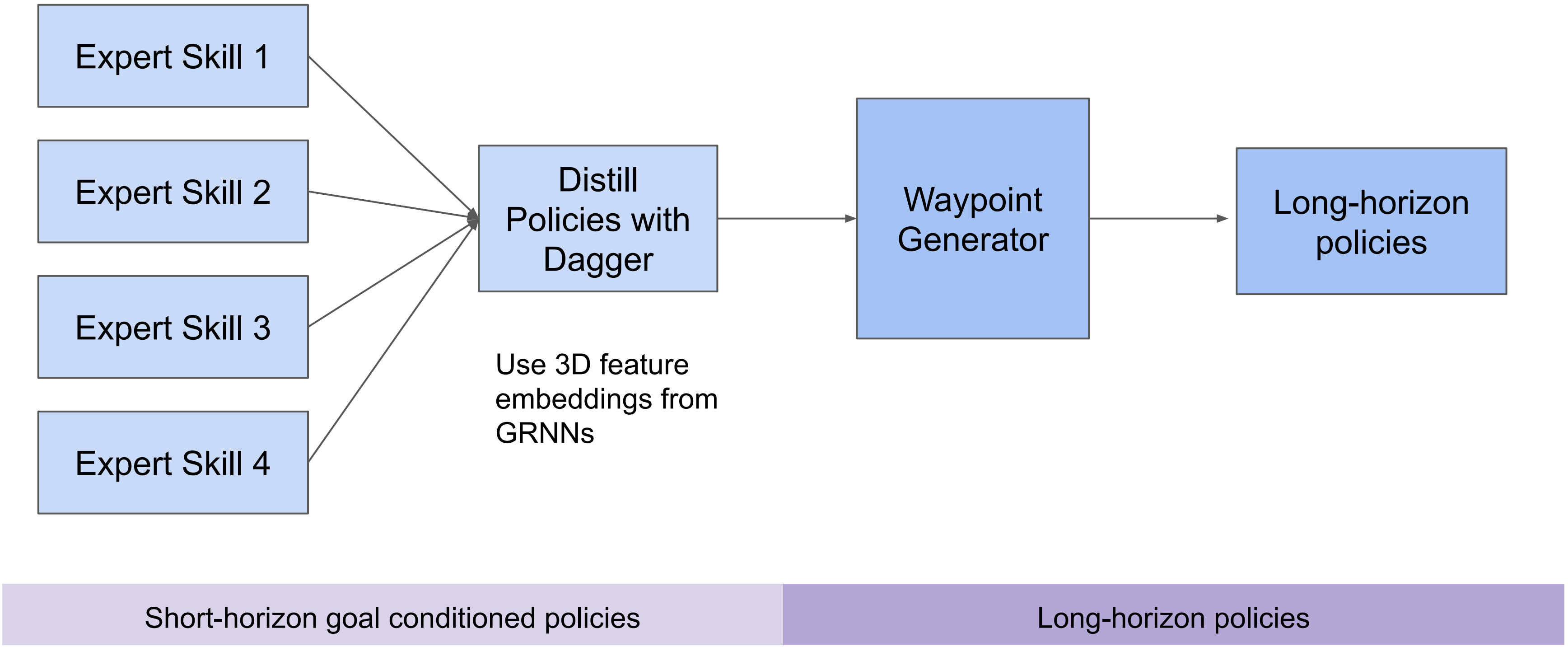
Limitations

1. Experiments were performed on relatively simple tasks and environments. It is not clear if this method will generalize well to more complex environments and will do well on more complex temporally extended tasks.
2. Input is unstructured.
   a. Might not generalize well to different object configurations
   b. Might not generalize across multiple cameras

Outline

- ❏ Motivation
- ❏ Current Approaches
- ❏ **Proposed Approach & Progress**
- ❏ Future Work

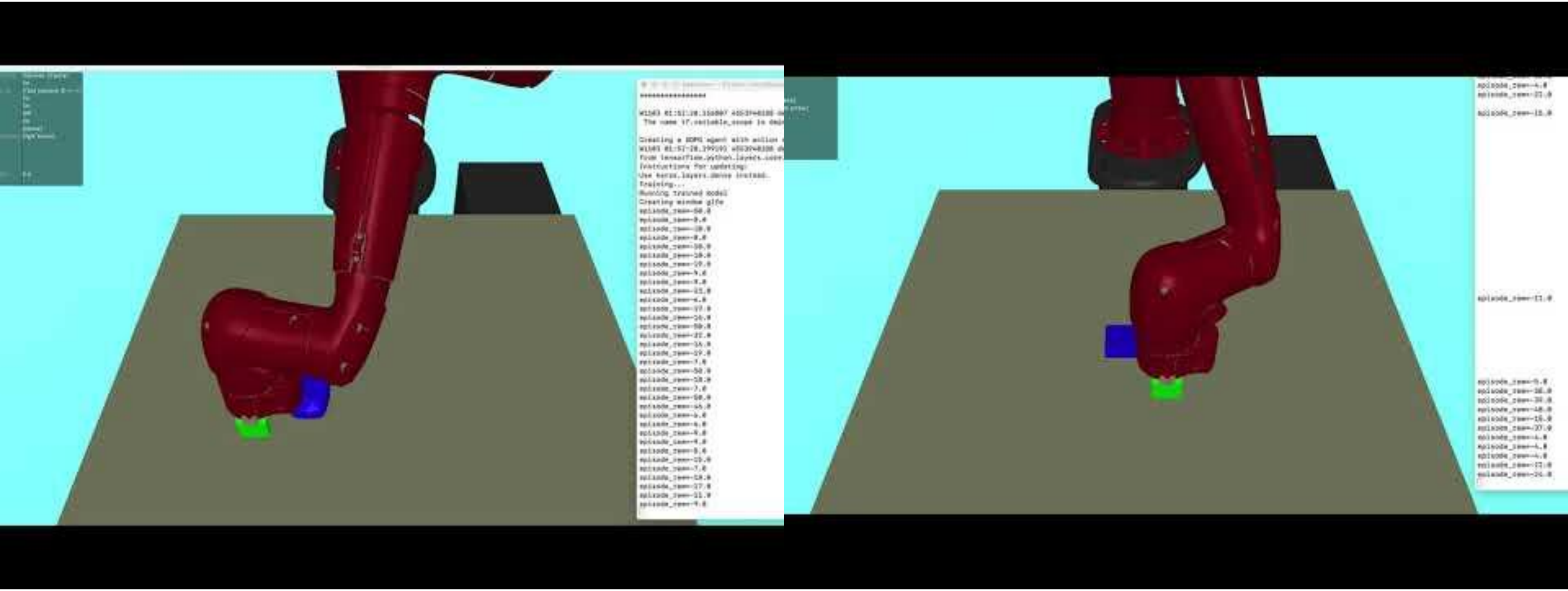
Proposed Approach - High-level Overview



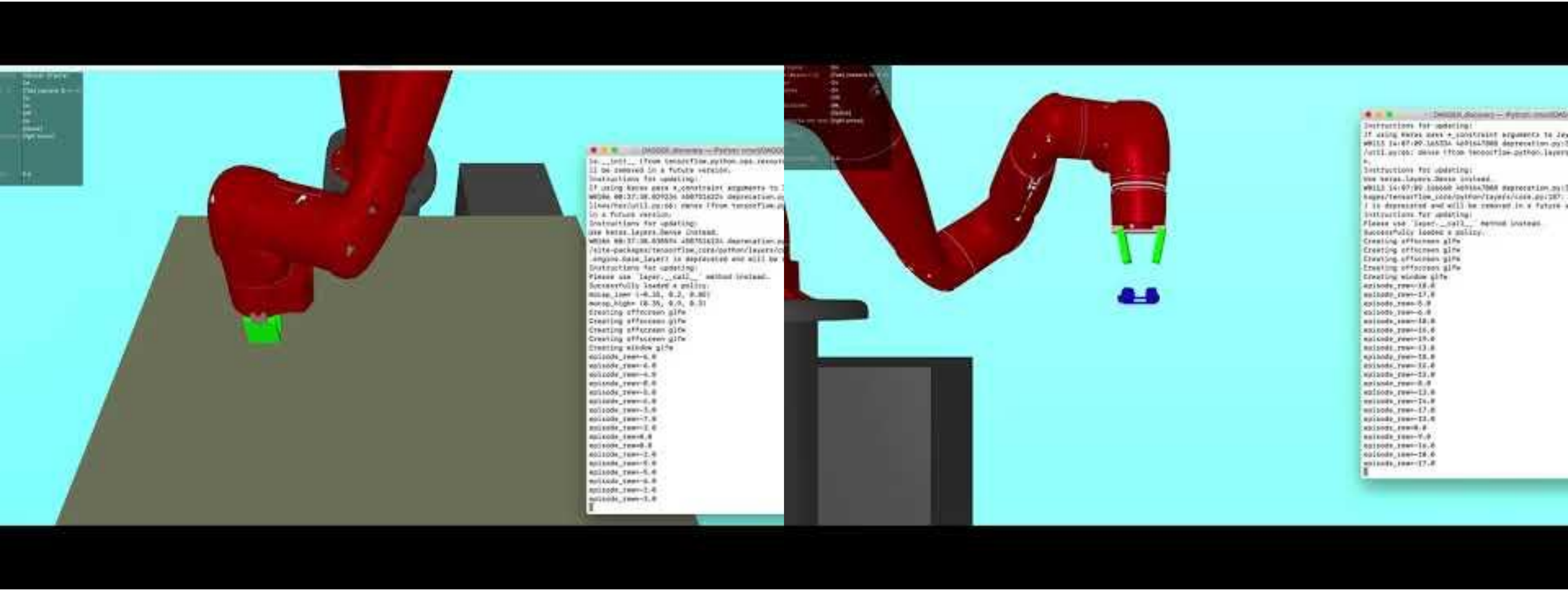
Our Progress

- Working on Mujoco Sawyer arm environment
- 10 different objects to be considered for manipulation
- Trained goal-conditioned experts to manipulate above objects on the following tasks
    - Push
    - Grasp
    - Reach
    - Place
    - Rotate

Results (Push)

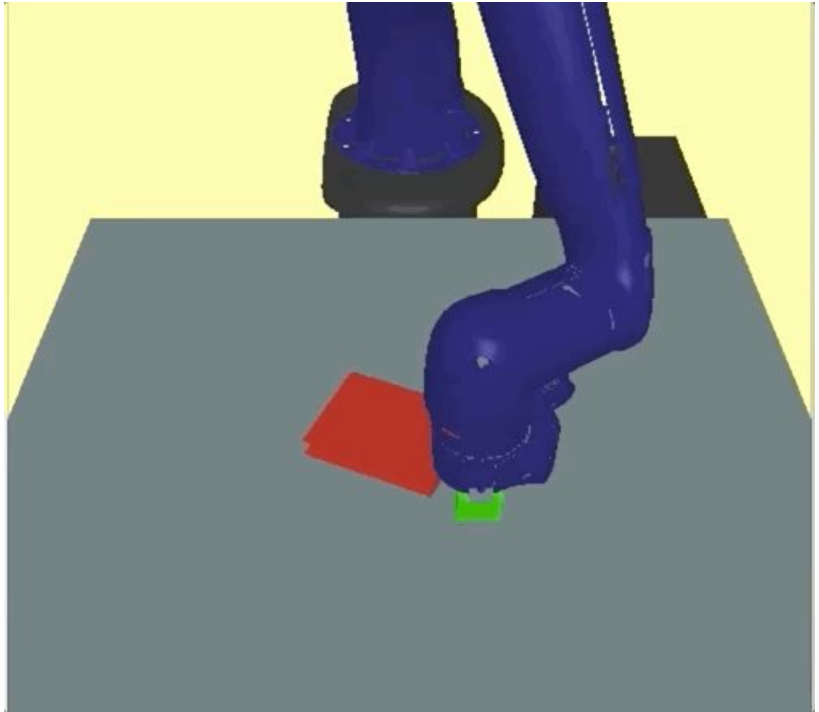


Result (Reach, Pick and Place)



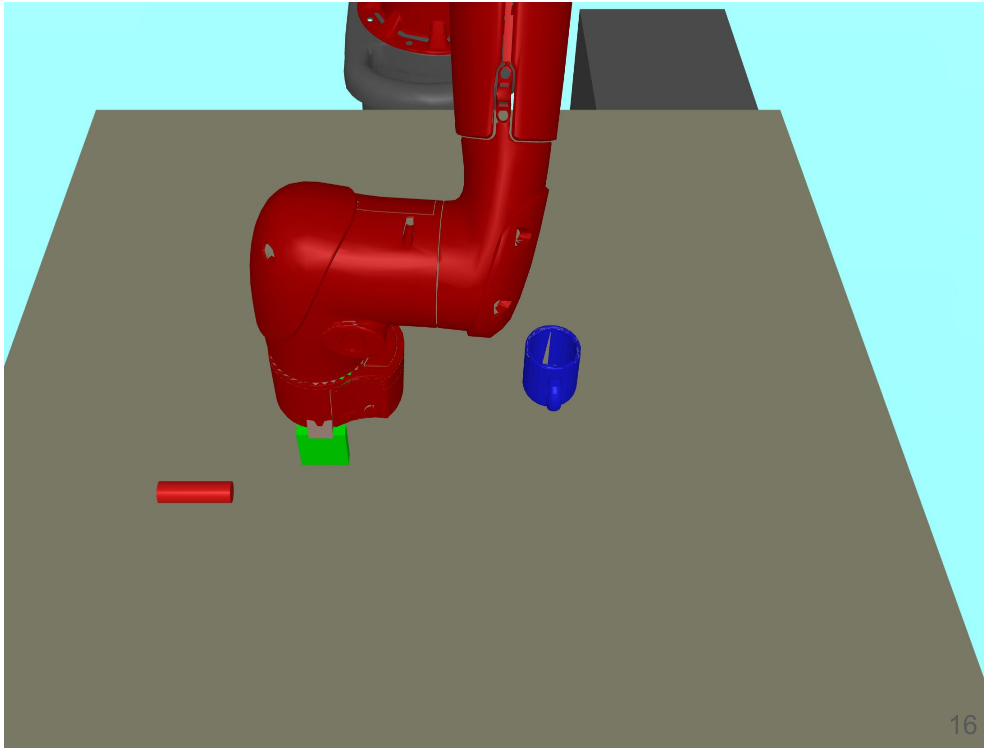
Our Progress

- Distilled policies with Dagger(Dataset Aggregation algorithm) - able to achieve around ~70% goal reach accuracy
  - Baseline (without 3D feature embeddings) has comparable performance



Our Progress

- Long-horizon tasks
  - Stacking of blocks
  - Transfer a pen to a mug



Future Directions

- Speed- up Dagger algorithm with multi-gpus
- Implement a waypoint generator that generates waypoints for long-horizon goals
- Fine-tuning skills for long-term tasks
- Plan generation (required for stacking)

Thank you!

Questions?