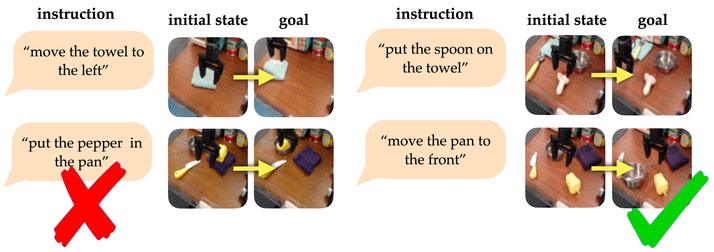
# Policy Adaptation via Language Optimization

## Decomposing Tasks for Few-Shot Imitation

RAIL — InterACT Laboratory — Berkeley UNIVERSITY OF CALIFORNIA — BAIR BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

## Motivation

Goal-conditioned policies robustly generalize across tabletop manipulation skills and don't require task annotations. Language is a more natural form of task specification for humans, though it is harder to learn. How can we get the benefits of goal-conditioned learning when following language instructions?

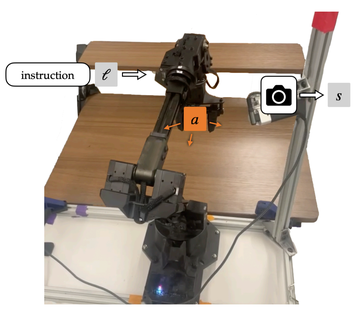


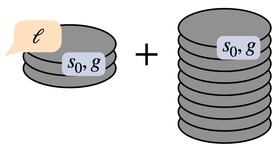

**Language**: less supervision, hard to learn

**Goals**: hindsight relabeling, better generalization

## Setup

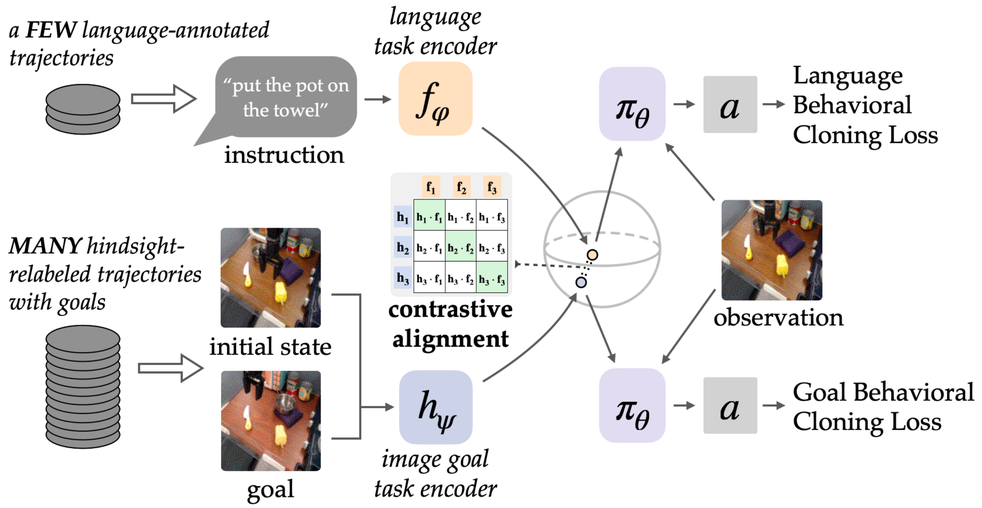
We perform 6 DoF control of a WidowX arm for language-conditioned tabletop manipulation.




We have a few language-annotated trajectories along with many **unlabeled** trajectories. We need a **semi-supervised** approach to use both sources of data.

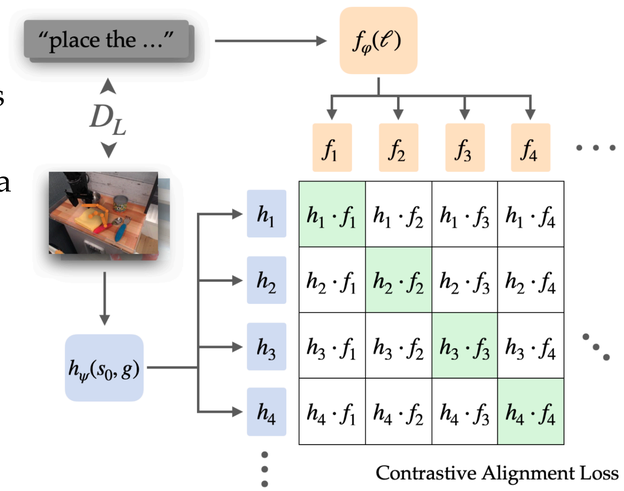
## Approach




### Goal Representations for Instruction Following (GRIF)

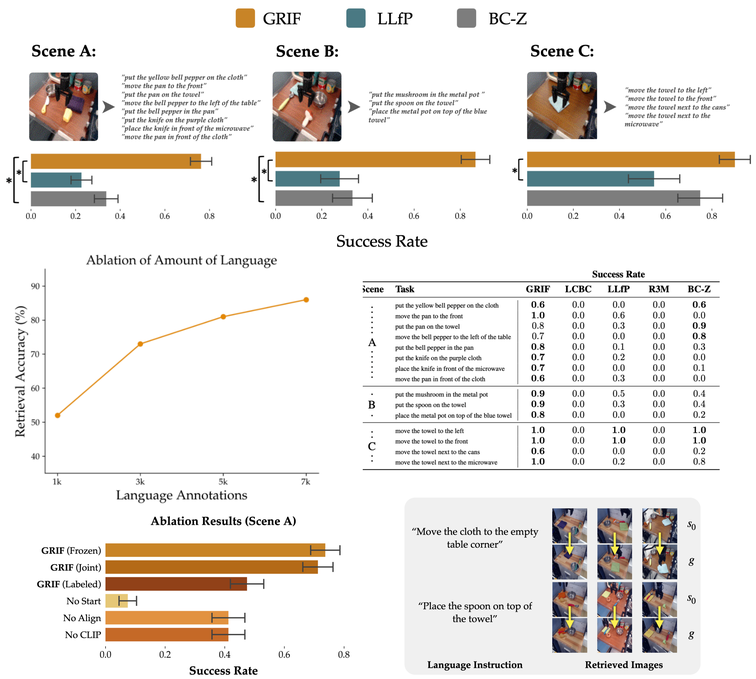
## Aligned Task Representations

Our approach aligns (start state, goal) pairs in the scene with language tasks using a **contrastive loss**. We learn a policy on top of these aligned representations using behavioral cloning.




With **hindsight relabeling**, we can use both unlabeled trajectories and language-annotated trajectories to train our policy. We also modify the **CLIP** architecture to leverage pre-trained language knowledge.

## Results

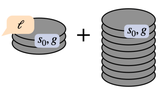



| | Task | GRIF | LCBC | LLfP | R3M | BC-Z |
|---|---|---|---|---|---|---|
| A | put the yellow bell pepper on the cloth | **0.6** | 0.0 | 0.0 | 0.0 | **0.6** |
| | move the pan to the front | **1.0** | 0.0 | 0.6 | 0.0 | 0.0 |
| | put the pan on the towel | 0.8 | 0.0 | 0.3 | 0.0 | **0.9** |
| | move the bell pepper to the left of the table | 0.7 | 0.0 | 0.0 | 0.0 | **0.8** |
| | put the bell pepper in the pan | **0.8** | 0.0 | 0.1 | 0.0 | 0.3 |
| | put the knife on the purple cloth | **0.7** | 0.0 | 0.2 | 0.0 | 0.0 |
| | place the knife in front of the microwave | **0.7** | 0.0 | 0.0 | 0.0 | 0.1 |
| | move the pan in front of the cloth | **0.6** | 0.0 | 0.3 | 0.0 | 0.0 |
| B | put the mushroom in the metal pot | **0.9** | 0.0 | 0.5 | 0.0 | 0.4 |
| | put the spoon on the towel | **0.9** | 0.0 | 0.3 | 0.0 | 0.4 |
| | place the metal pot on top of the blue towel | **0.8** | 0.0 | 0.0 | 0.0 | 0.2 |
| C | move the towel to the left | **1.0** | 0.0 | **1.0** | 0.0 | **1.0** |
| | move the towel to the front | **1.0** | 0.0 | **1.0** | 0.0 | **1.0** |
| | move the towel next to the cans | **0.6** | 0.0 | 0.0 | 0.0 | 0.2 |
| | move the towel next to the microwave | **1.0** | 0.0 | 0.2 | 0.0 | 0.8 |

Success Rate

We outperform baseline and ablation approaches on **diverse unseen language instructions** in different scenes. In particular GRIF is better able to **ground** language instructions compared to baselines while performing well at manipulation.

## Summary




We train policies on **language**-annotated and **goal**-relabeled trajectories

**Contrastive** language-goal task alignment enables robust grounding

Our approach outperforms baselines on **unseen** instructions