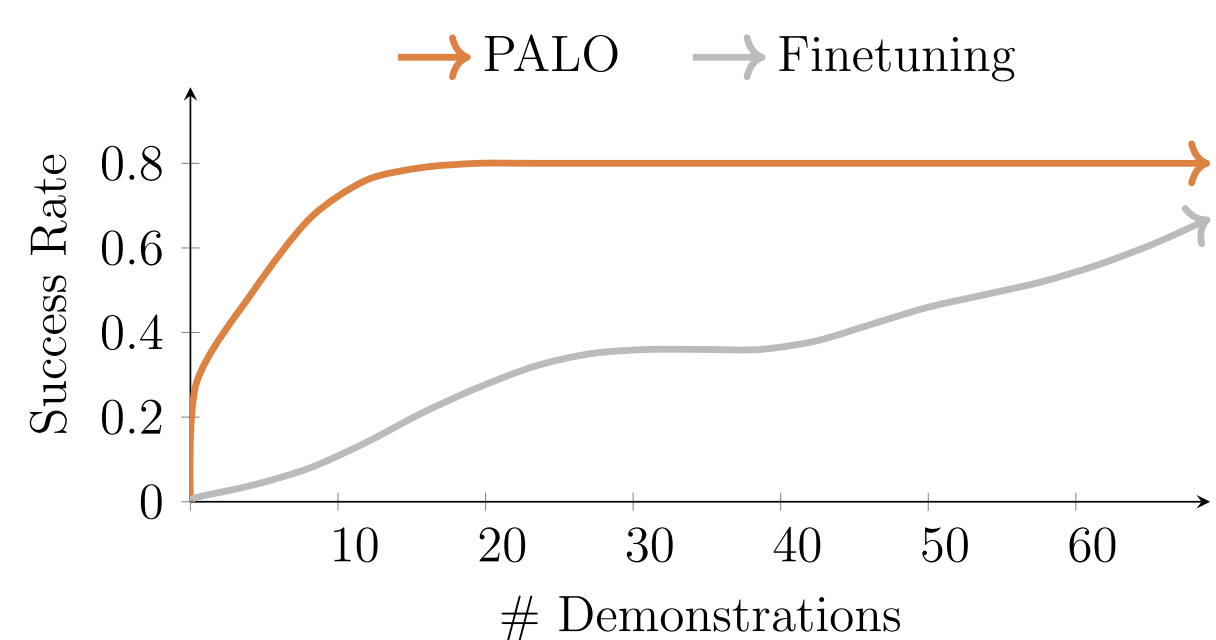
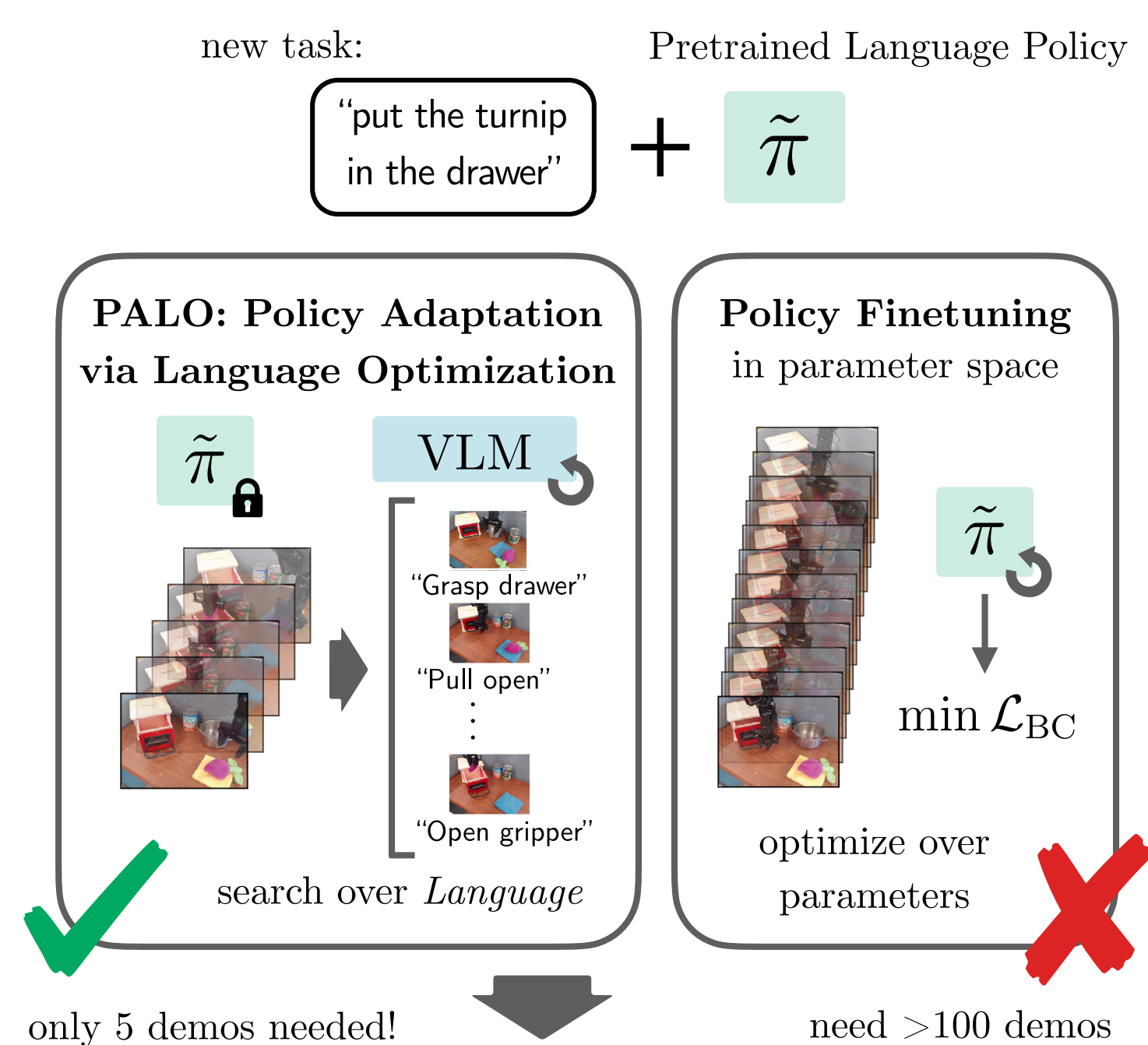


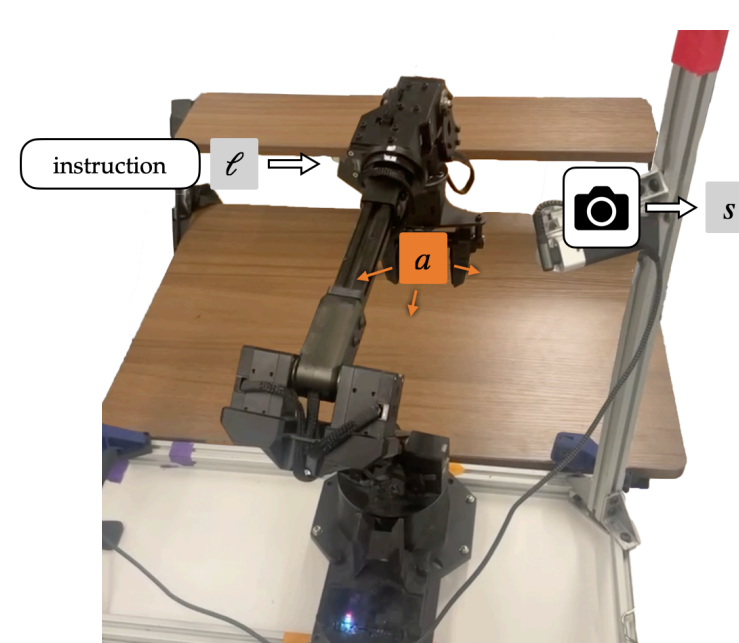
Motivation

- Learned language-conditioned robot policies struggle to adapt to new manipulation tasks even when pre-trained across diverse instructions
- How can we adapt a pre-trained generalist policy to **new tasks from a few demonstrations**?
- **Insight:** exploit semantic structure of new tasks by decomposing them into language that the pre-trained policy is more familiar with using a VLM



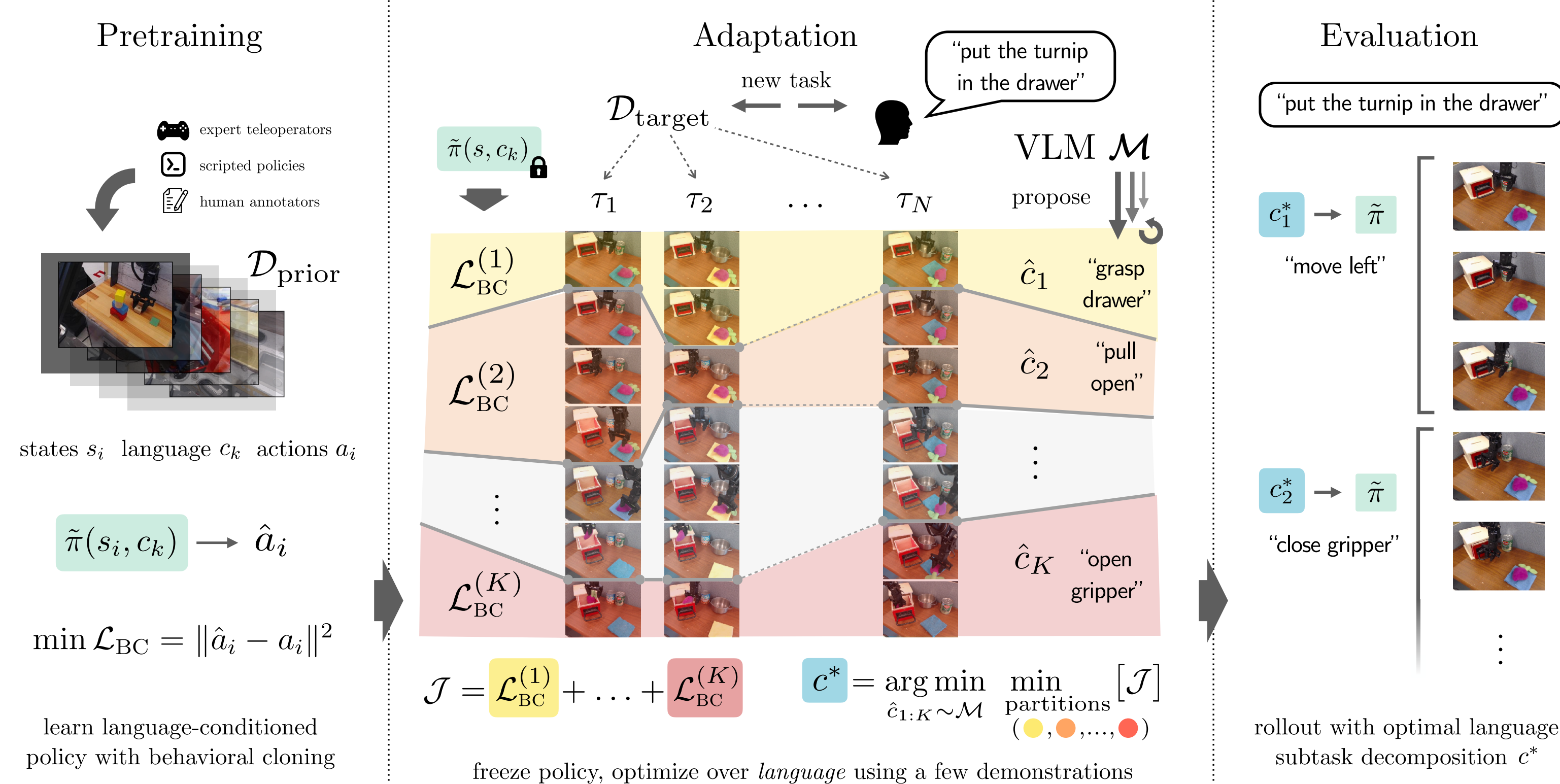
Setup

We conduct experiments with a 6 DoF WidowX arm with a pretrained language conditioned policy.



The training Bridge-v2 dataset consists of short-horizon manipulation tasks with language annotations, augmented with heuristics and language model rephrasing

Approach



Policy Adaptation via Language Optimization

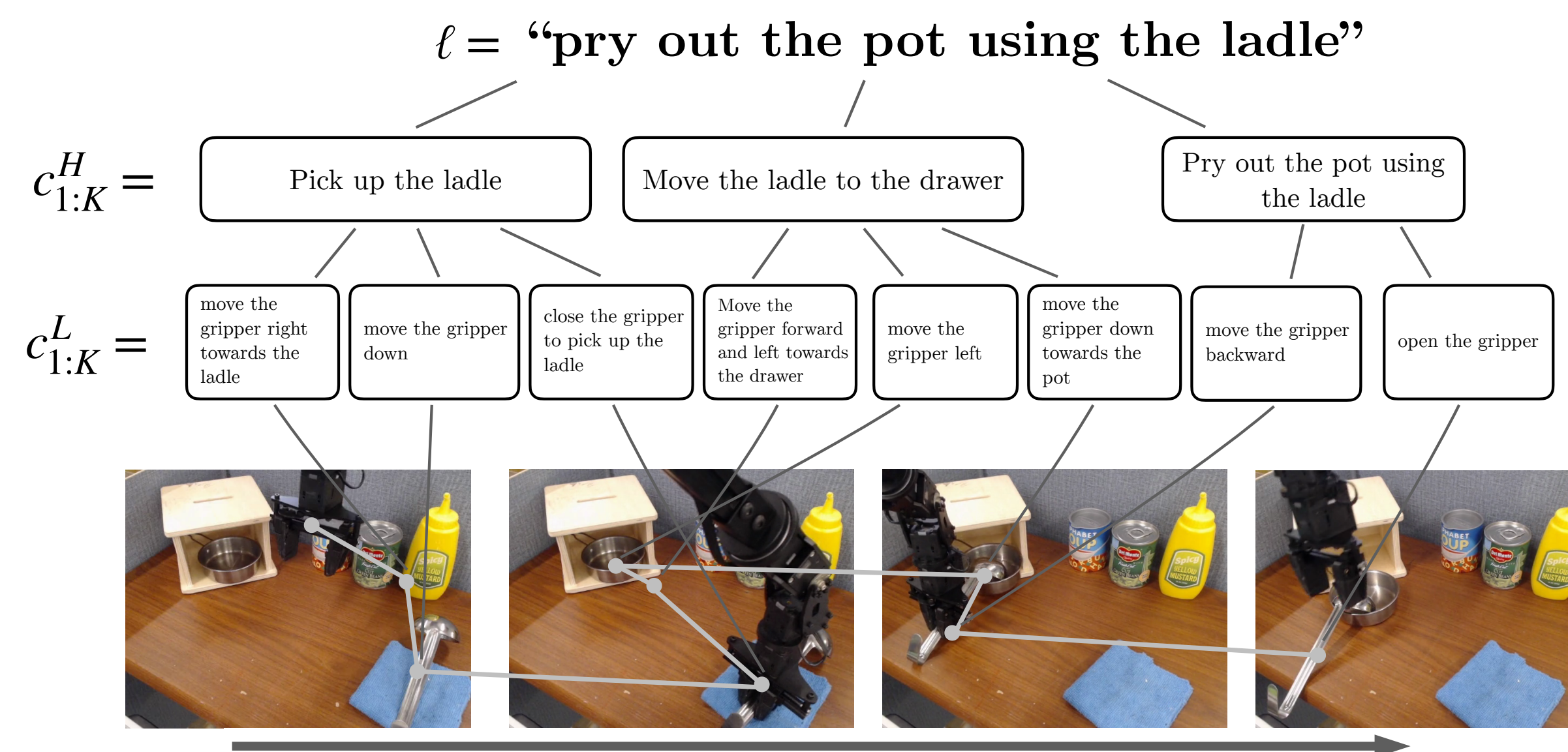
PALO selects the best task decomposition from several feasible ones proposed by a VLM (GPT-4o) based on MSE, and executes the decomposition during inference.

Theorem 3.7. The (out-of-distribution) regret of PALO on ρ_{target} can be bounded as:

$$R_{\pi_{\beta}}(\pi_{\text{PALO}}; \rho_{\text{target}}) \leq R_{\pi_{\beta}}(\hat{\pi}; \rho_{\text{prior}}) + \mathbb{E}[D_{\text{TV}}(p_{\text{target}}(c_{k_t}), p_{\text{prior}}(c_{k_t}))] + (2D_{\text{KL}}[p(c_{1:K}), p_{\mathcal{M}}])^{1/2} + \frac{\sqrt{M} + \sqrt{n \log(Mn)}}{n} + 1/M + 1/K + N^{-2/K} \quad (8)$$

where π_{PALO} is from Algorithm 1, $\hat{\pi}(s_t, \ell)$ is trained on $\mathcal{D}_{\text{prior}}$, and $t \sim \text{Unif}(1 \dots H)$.

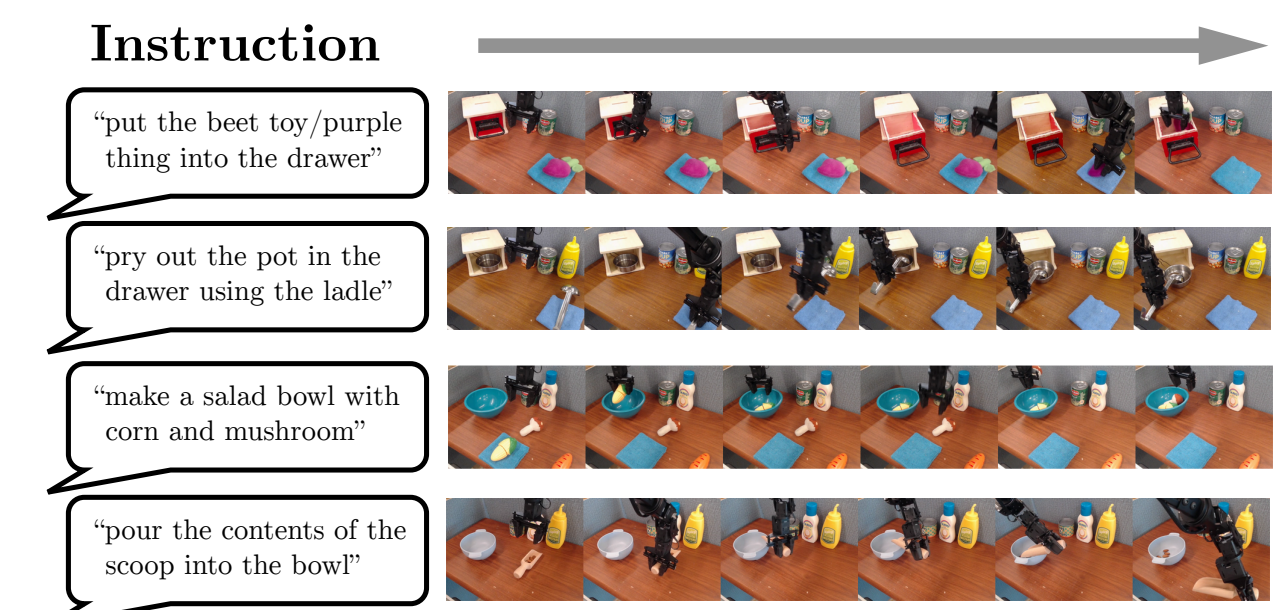
Hierarchical Language Policy



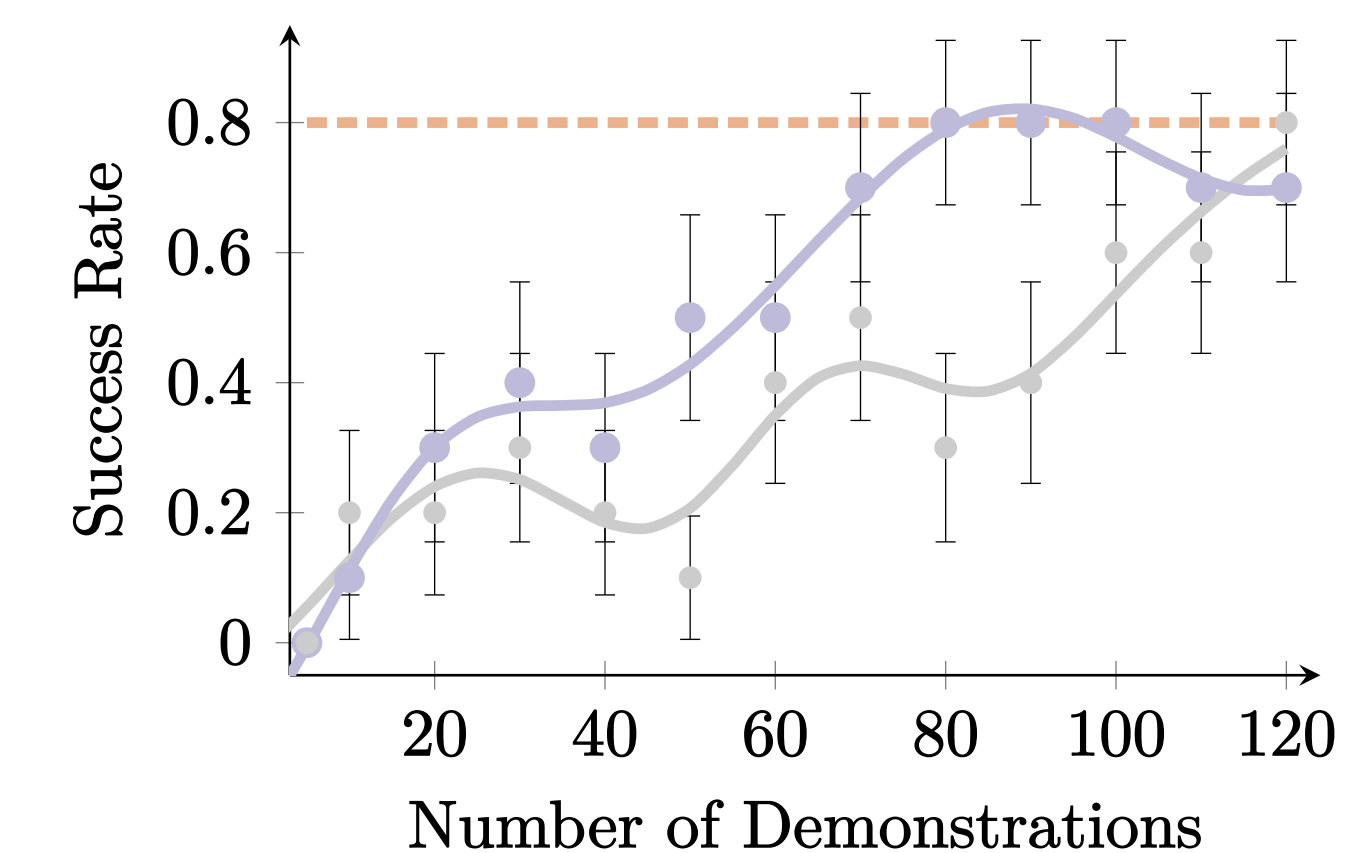
During pre-training, we optimize the policy with the joint encodings from both the high- and low-level task decompositions

Results

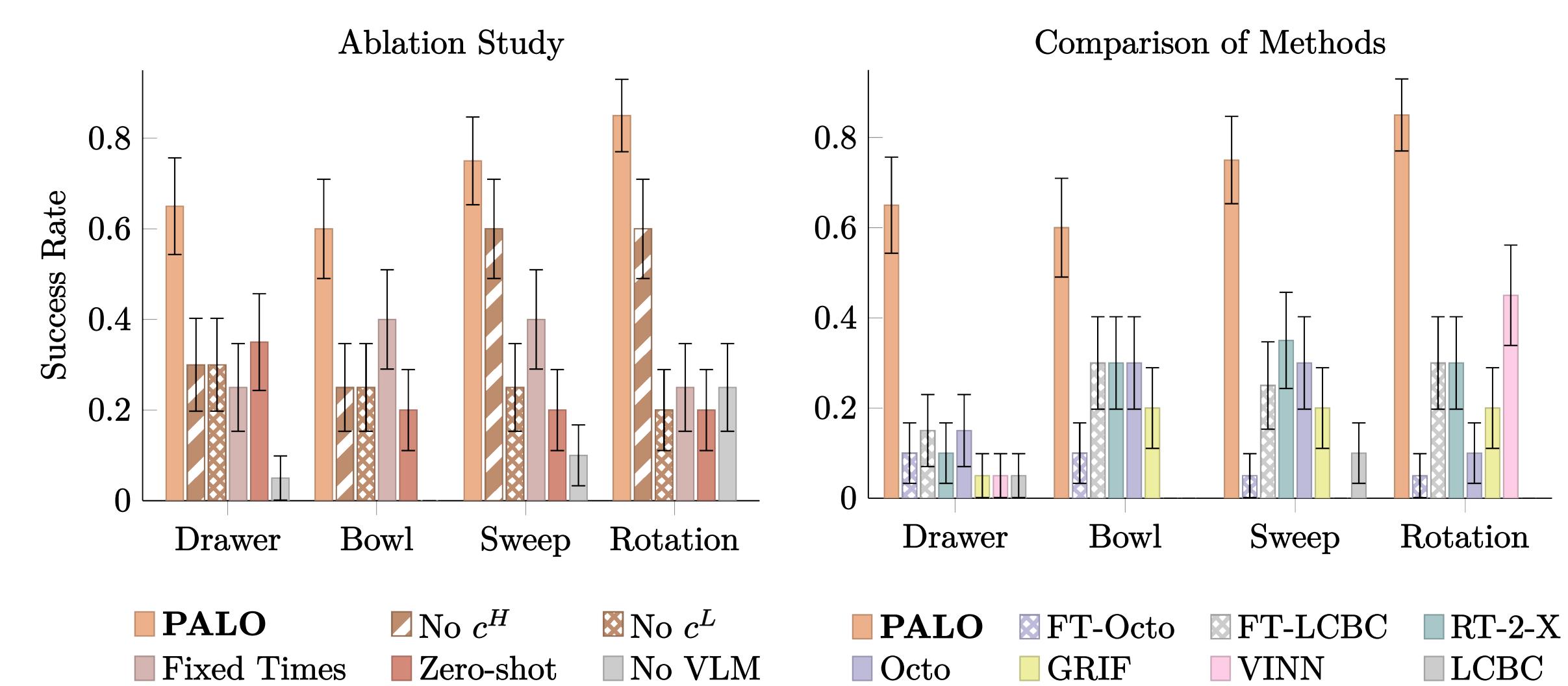
PALO adapts to **diverse, unseen, long-horizon tasks** better than existing fine-tuning and nonparametric approaches



Scaling of PALO and Finetuning Approaches



- PALO needs ≤ 5 demos for 80% success
- Fine-tuning needs > 80 demos for similar performance



Summary

- PALO enables adaptation to **unseen, long-horizon tabletop manipulation tasks** specified with language
- PALO uses the **semantic understanding of a VLM** to enable better adaptation of a pretrained generalist policy to new tasks
- PALO needs **15x less data to adapt to new tasks** compared to fine-tuning on the expert data