

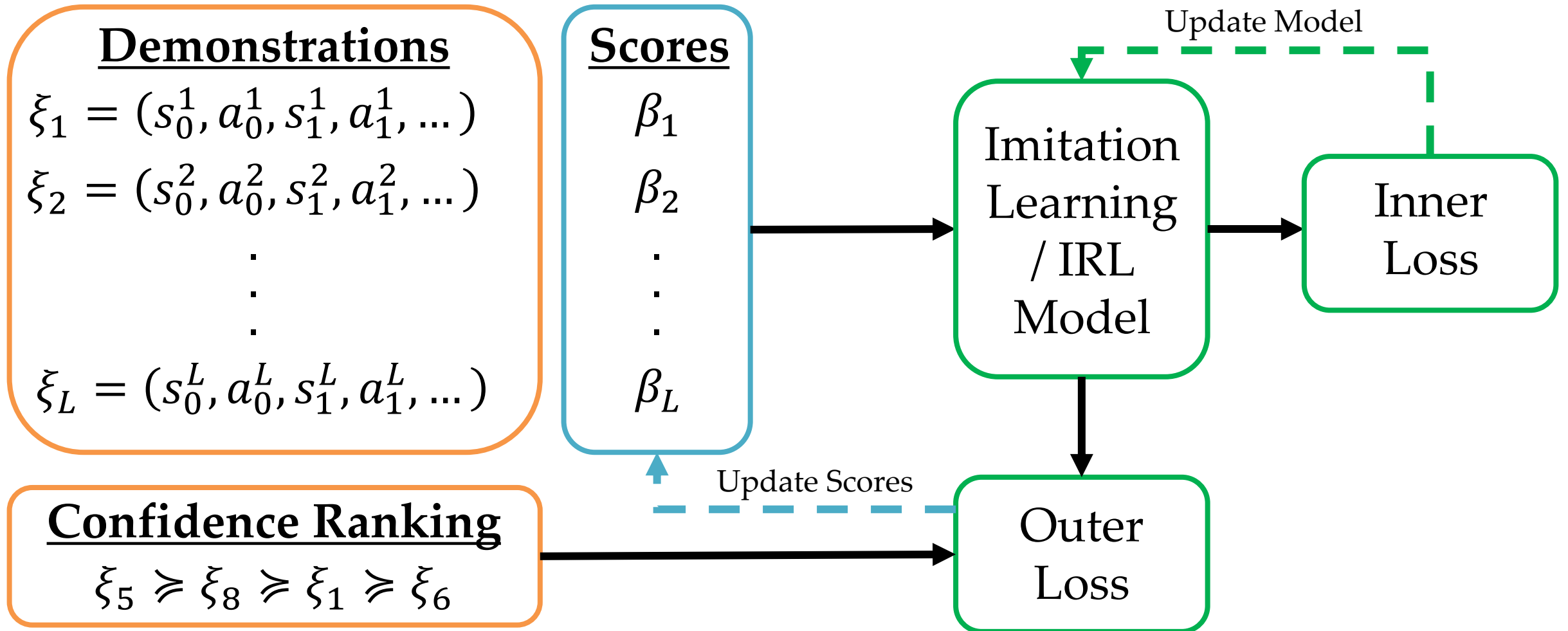
Robot Learning

Learning from human demonstrations

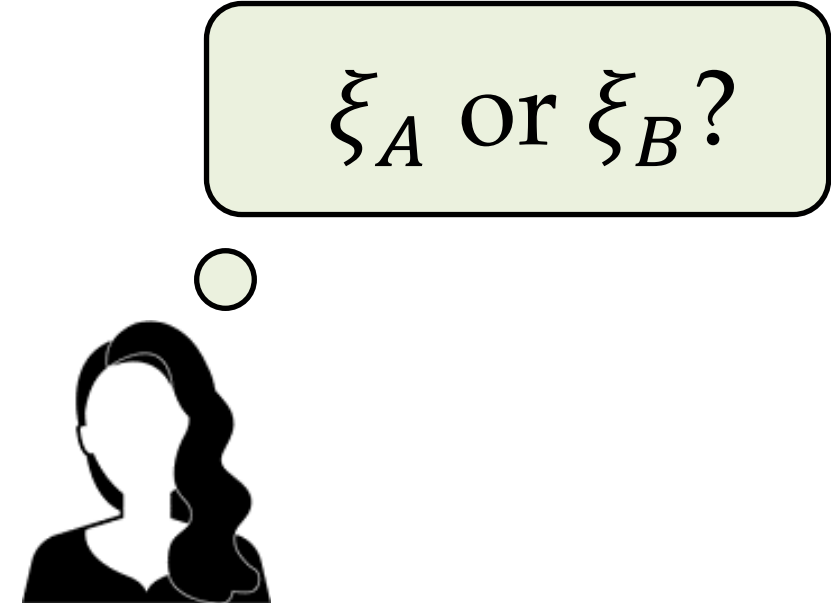
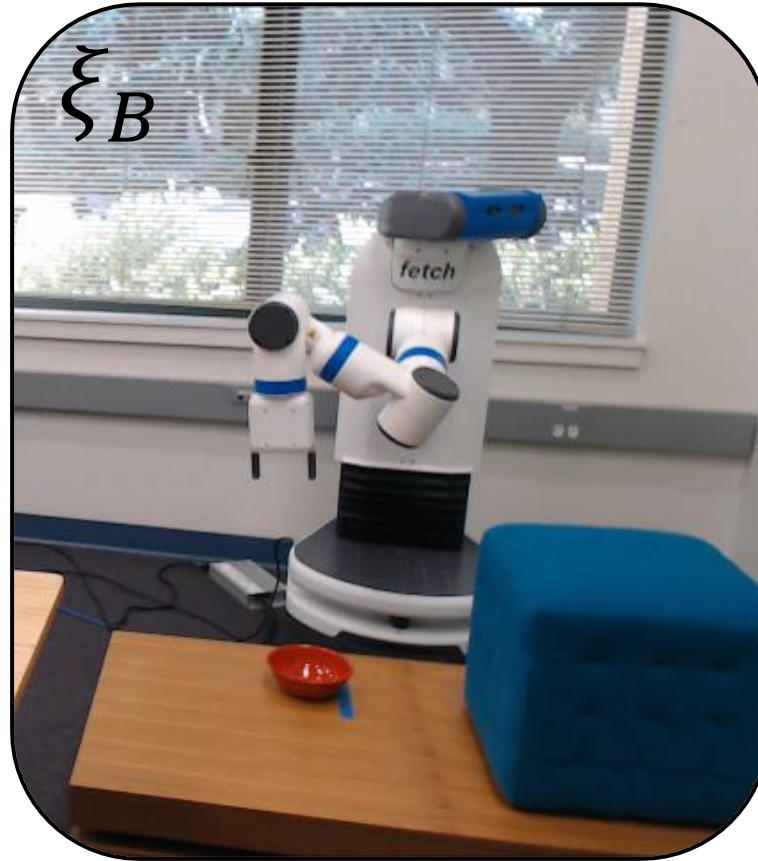
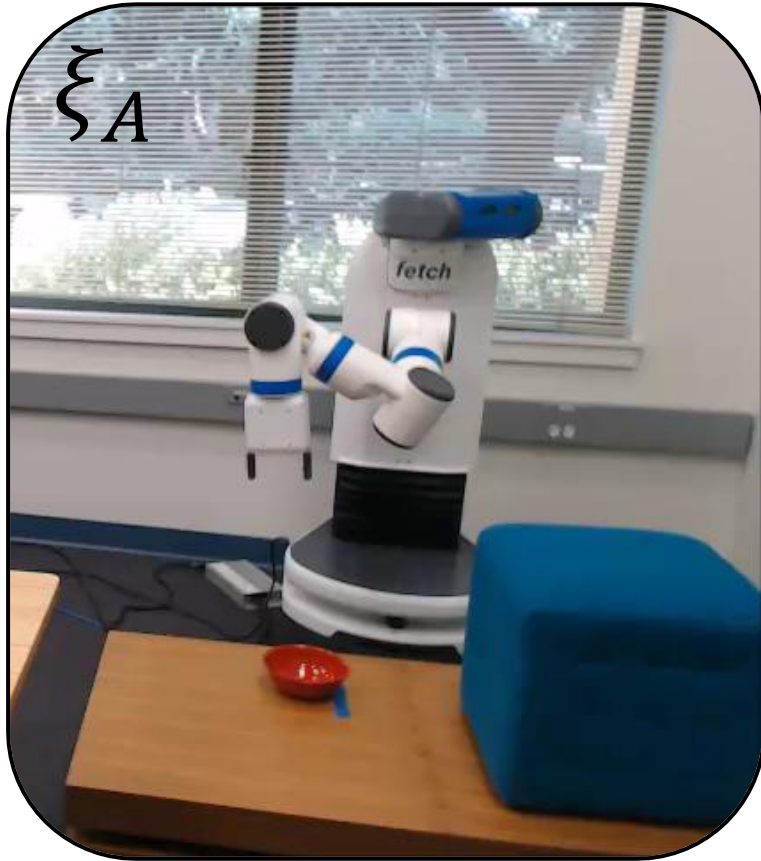
Last time...

- Learning from human feedback
 - Suboptimal demonstrations
 - Pairwise comparisons
 - Comparative language
 - Reinforcement learning from human feedback (RLHF)

Confidence-aware imitation learning

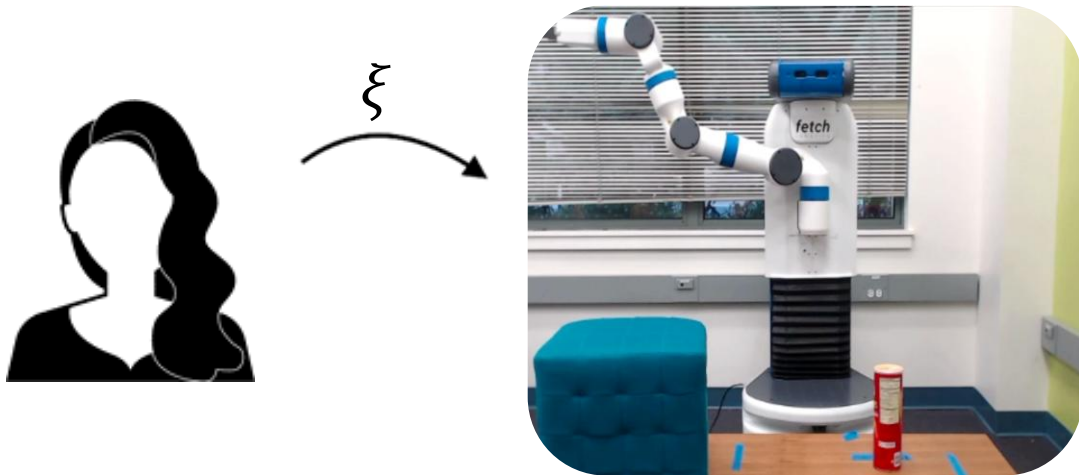


Incorporating Comparisons

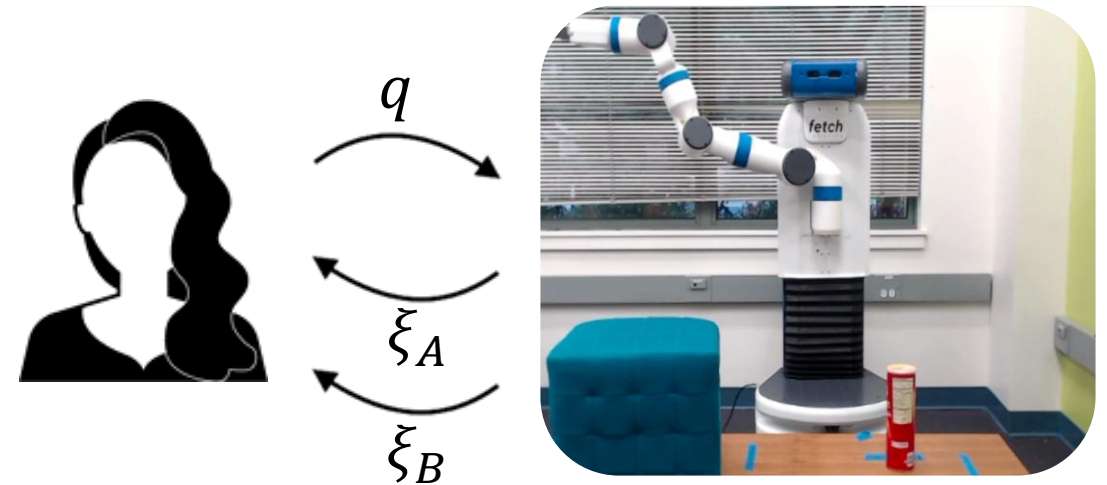


Choosing Queries

Demonstrations

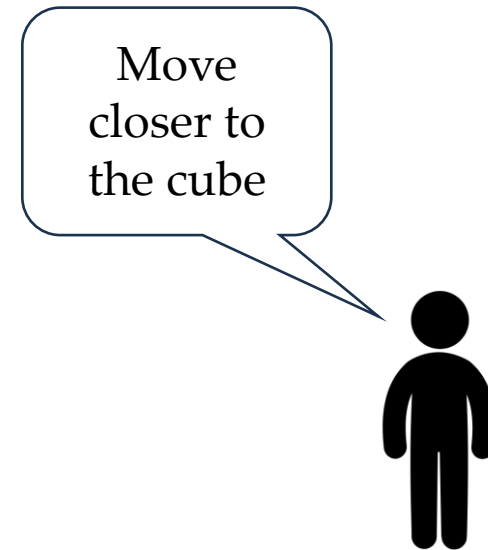
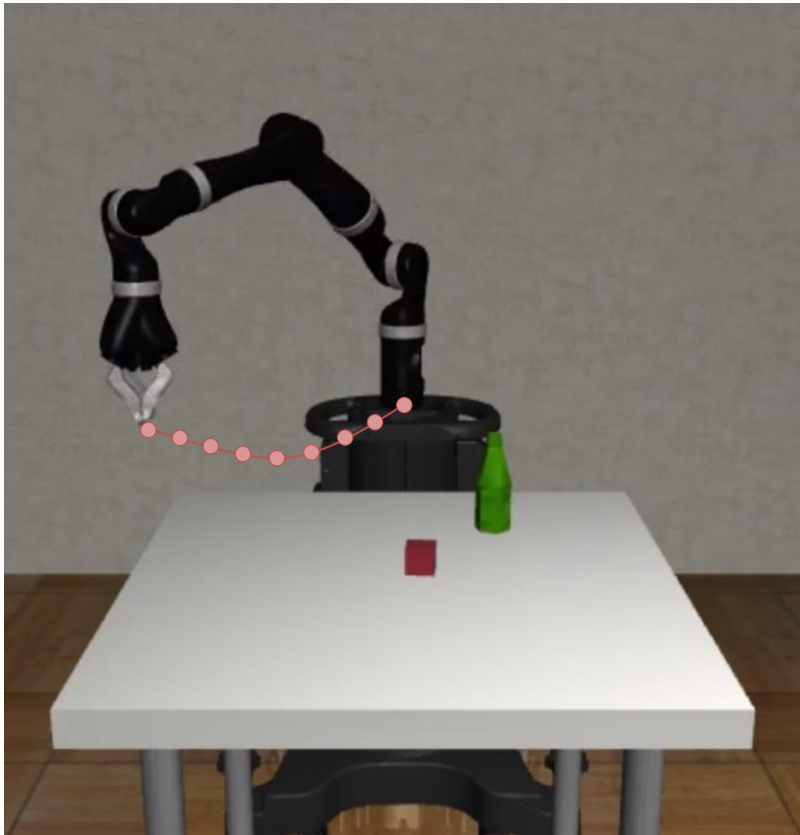


Comparisons

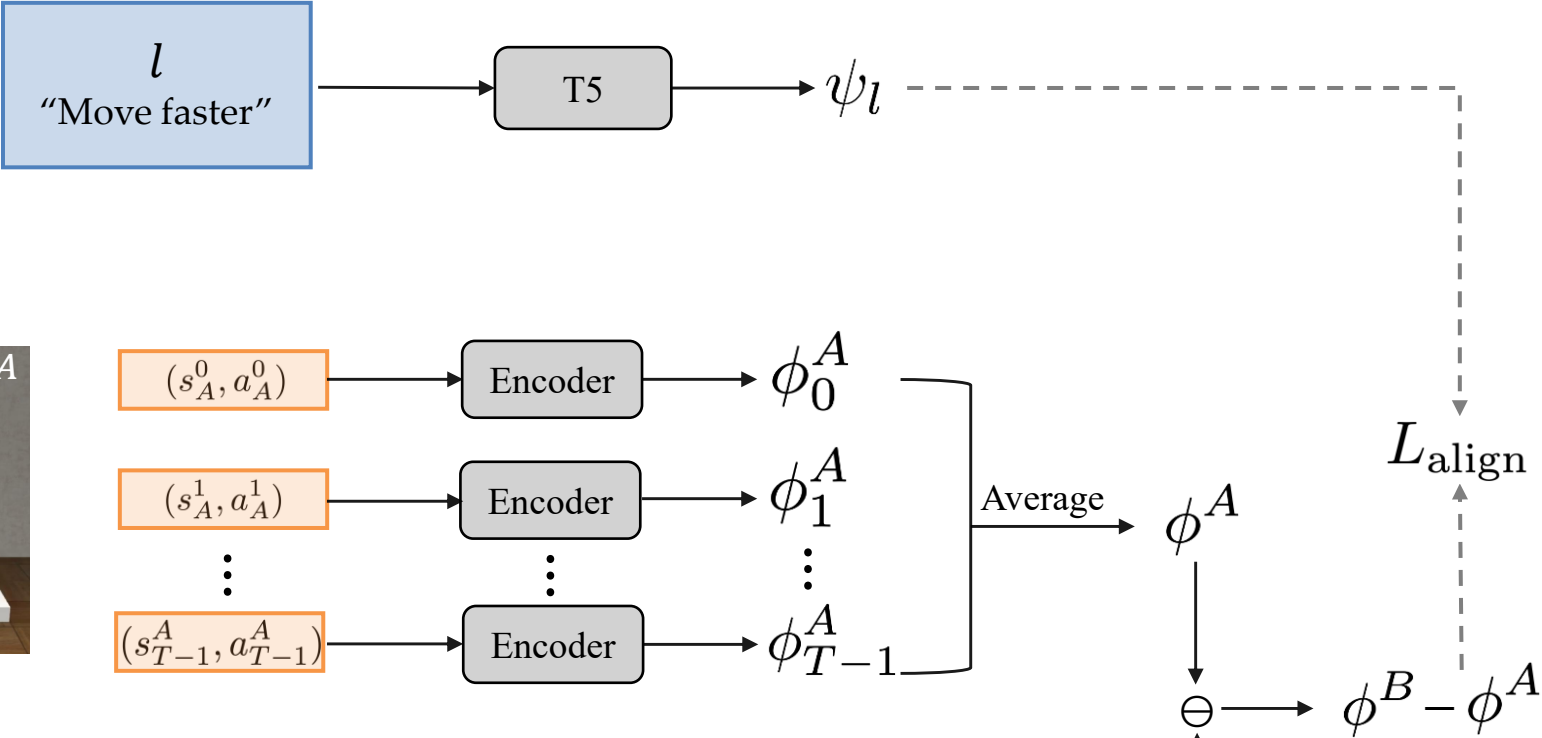


The robot can query the user with the query that will give the **most information**.

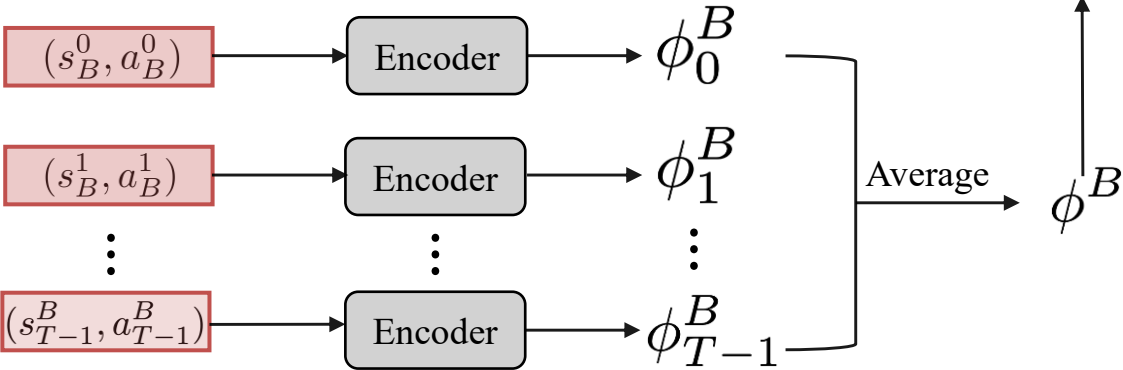
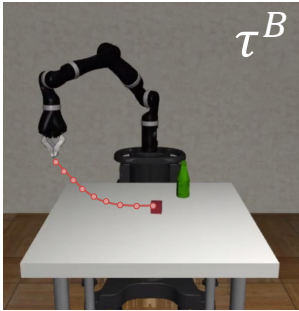
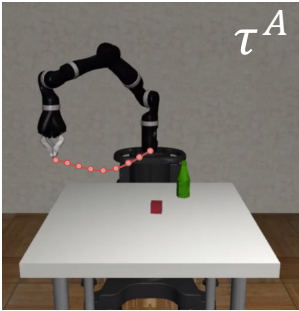
Comparative Language Feedback



Language feedback

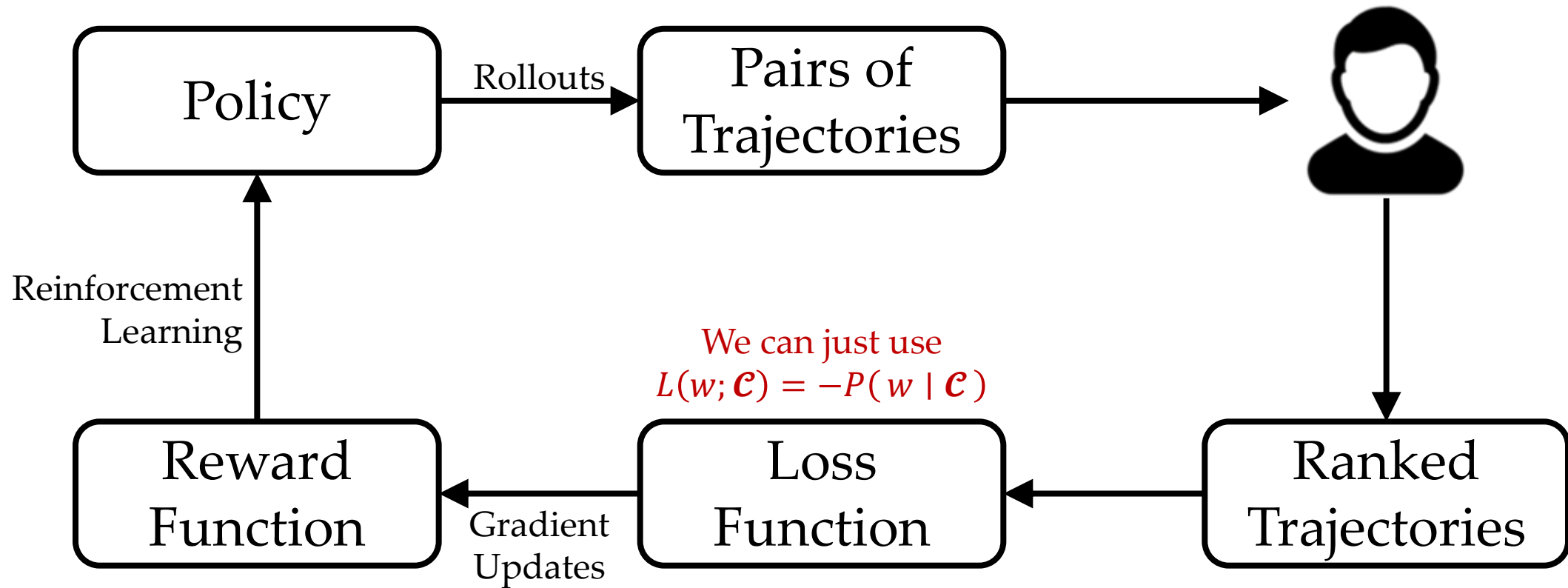


Trajectory pair



	Accuracy
Robosuite	84.9%
Meta-World	82.9%

RLHF



Today...

- Learning from human demonstrations

Wait, that's IRL, no?

Today...

- Learning from **human** demonstrations

Wait, that's IRL, no?

A person wearing a white lab coat is standing in a laboratory or office environment. They are positioned behind a desk or workbench. The background shows a modern interior with large windows and a brick wall. A white text box with a black border is overlaid on the image, containing the text "10 robot teleop demos takes 12 minutes".

**10 robot teleop
demos takes 12 minutes**

Today...

- Learning from **human** demonstrations
 - Retrieval-based methods
 - Latent action models
 - Mixing human and robot data

Challenges

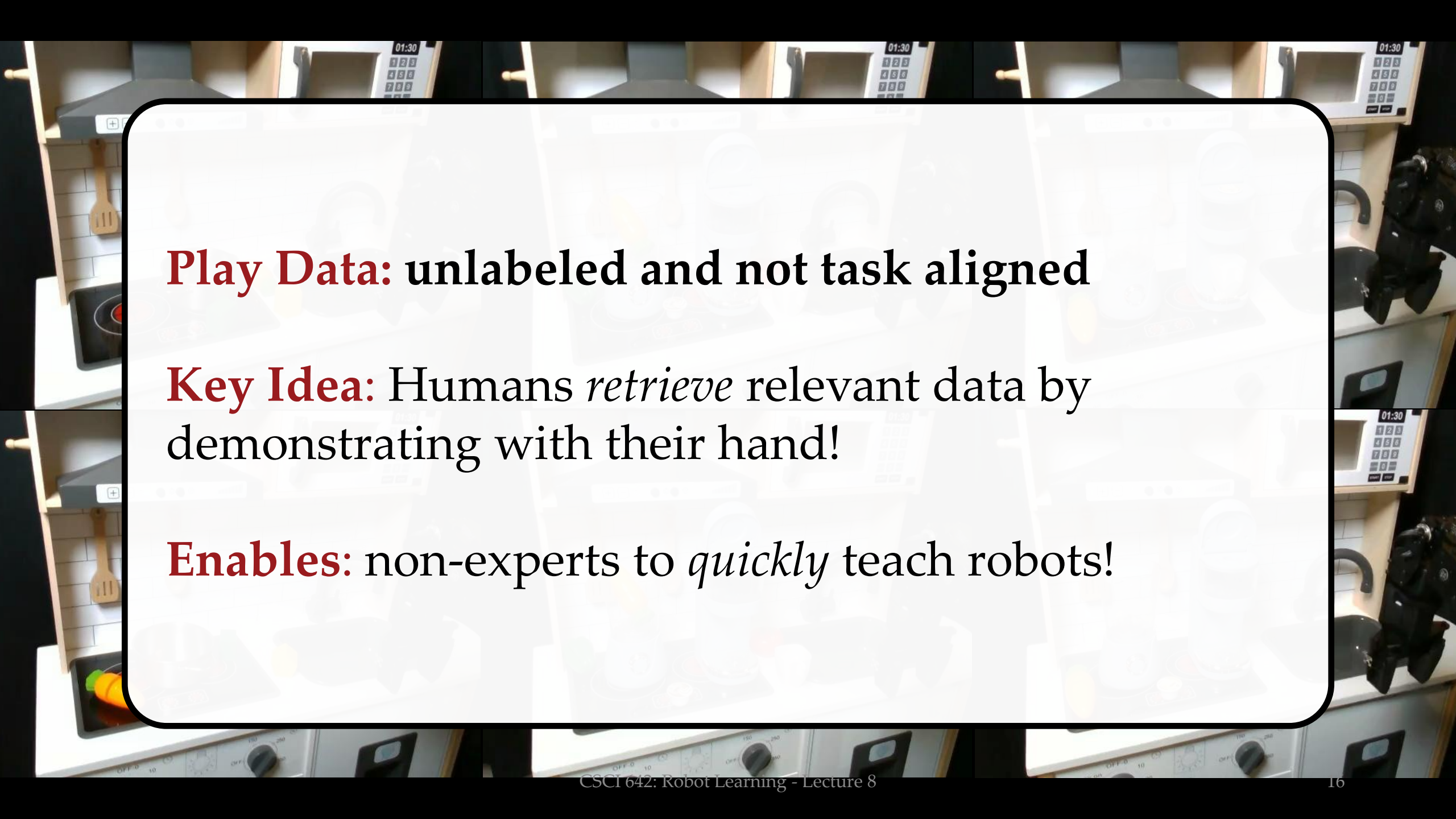
- Human demonstrations do not contain action labels.
- Humans and robots have very different morphologies.
 - Shape
 - Constraints
 - Sensors

Today...

- Learning from **human** demonstrations
 - Retrieval-based methods
 - Latent action models
 - Mixing human and robot data

Play Data



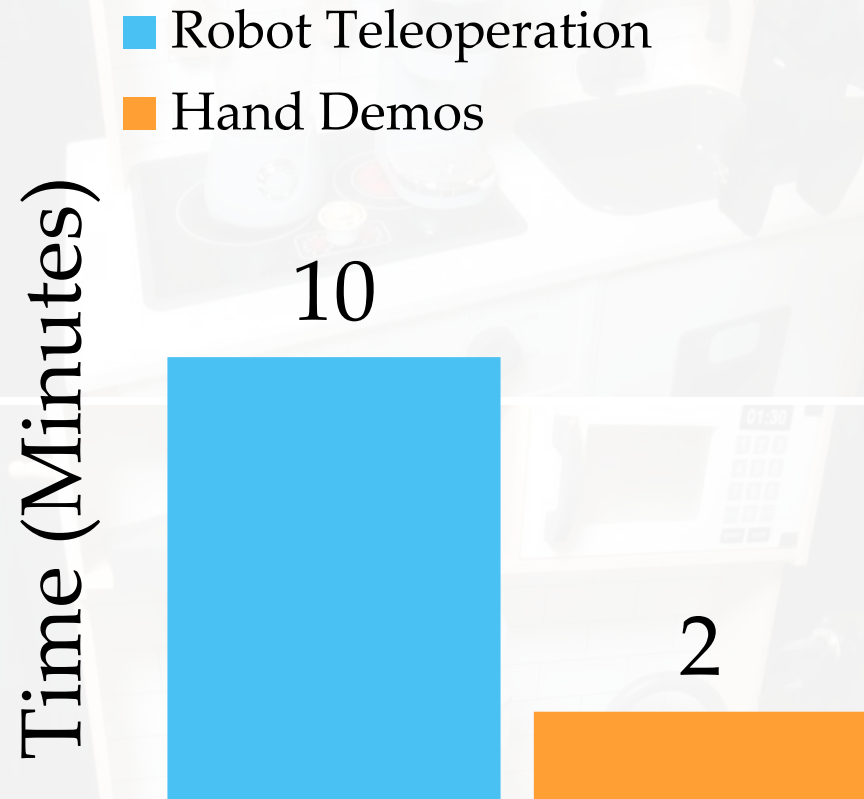


Play Data: unlabeled and not task aligned

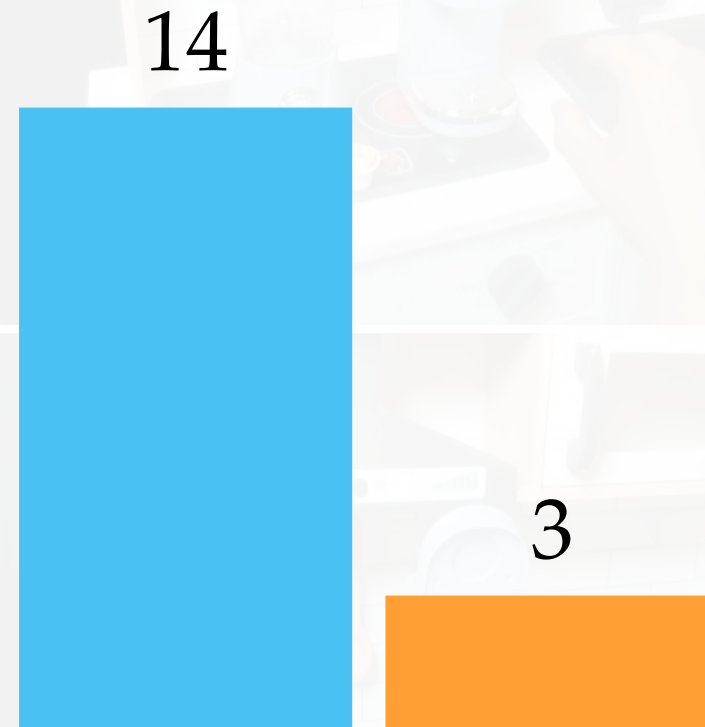
Key Idea: Humans *retrieve* relevant data by demonstrating with their hand!

Enables: non-experts to *quickly* teach robots!

Hand demos take 4x less time to collect!



User 1



User 2

HAND Me the Data

Filter Trajectories

Hand Demo τ_{hand}



Offline Play Dataset $\mathcal{D}_{\text{play}}$



Filter for visual similarity



Gripper/Hand Tracking Pipeline



Prompt: "Point at center of hand"



HAND Me the Data

Filter Trajectories

Hand Demo τ_{hand}



Offline Play Dataset $\mathcal{D}_{\text{play}}$



Filter for visual similarity

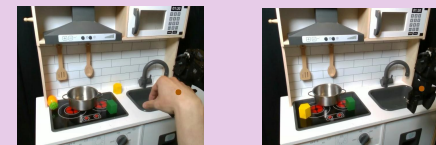


Gripper / Hand Tracking
Prompt: "Point at center of hand"

 Molmo



 CoTracker3



Hand Paths



Subsequence
Dynamic-
Time
Warping

Cost: 2



Cost: 10



Cost: 3



Parameter-Efficient Fine-tuning

Transformer Policy 

π_{base}

Add + Layer Norm

Bottleneck

Feed Forward Layer

Add + Layer Norm

Bottleneck

Multi-head Attention

LoRA Layers 

$\oplus$

W

W_{up}

W_{down}

$\oplus$

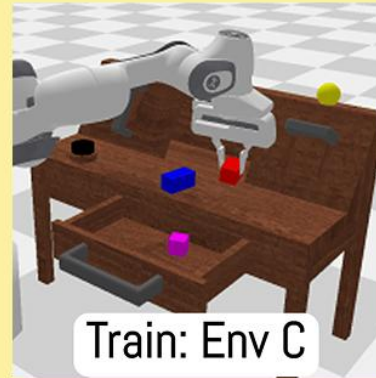
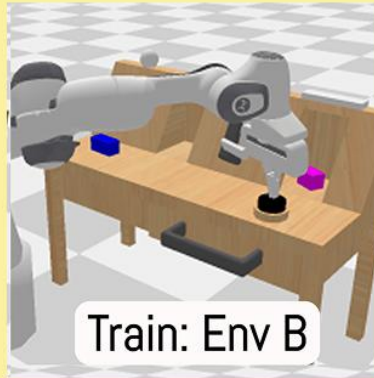
W

W_{up}

W_{down}

Experiments in CALVIN

Play Data



Real World Evaluation Tasks

Reach Block



Push Button



Close Microwave



Retrieval Quality

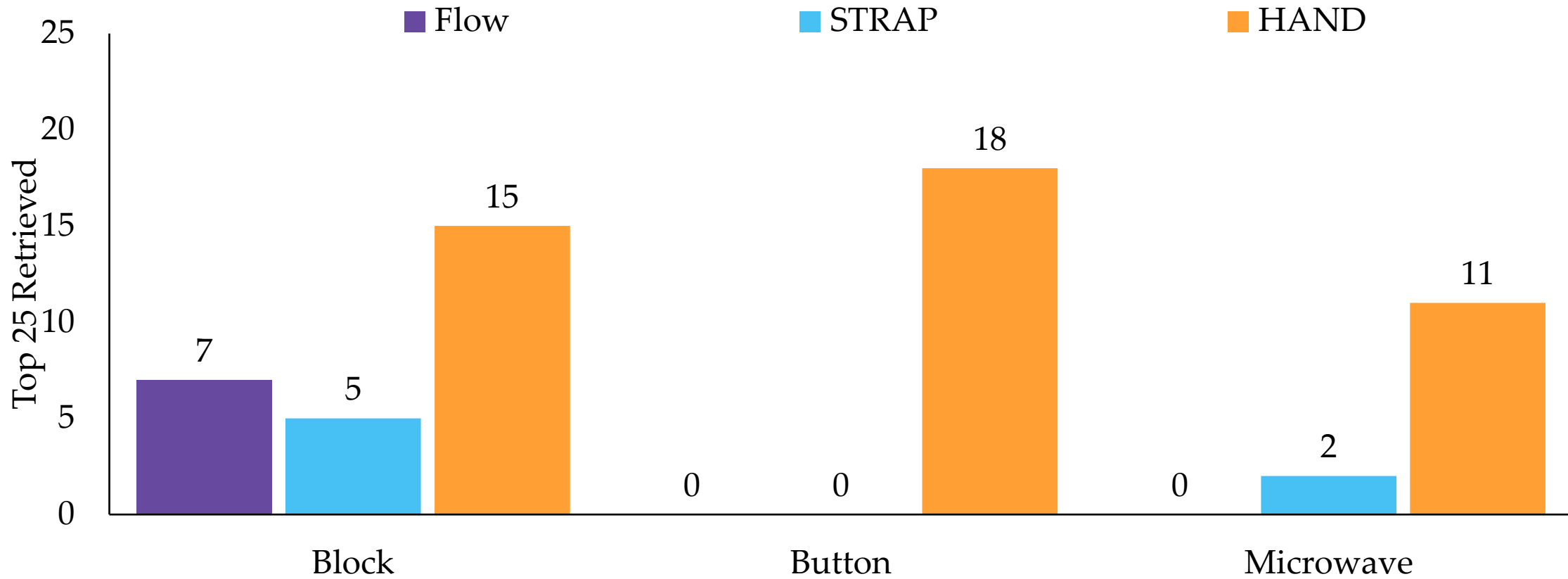
Reach Block



STRAP

HAND

Retrieval Quality



Out-of-Domain Retrieval

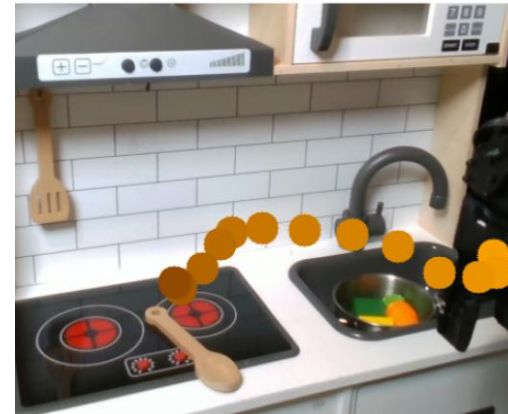
Hand Demo



STRAP



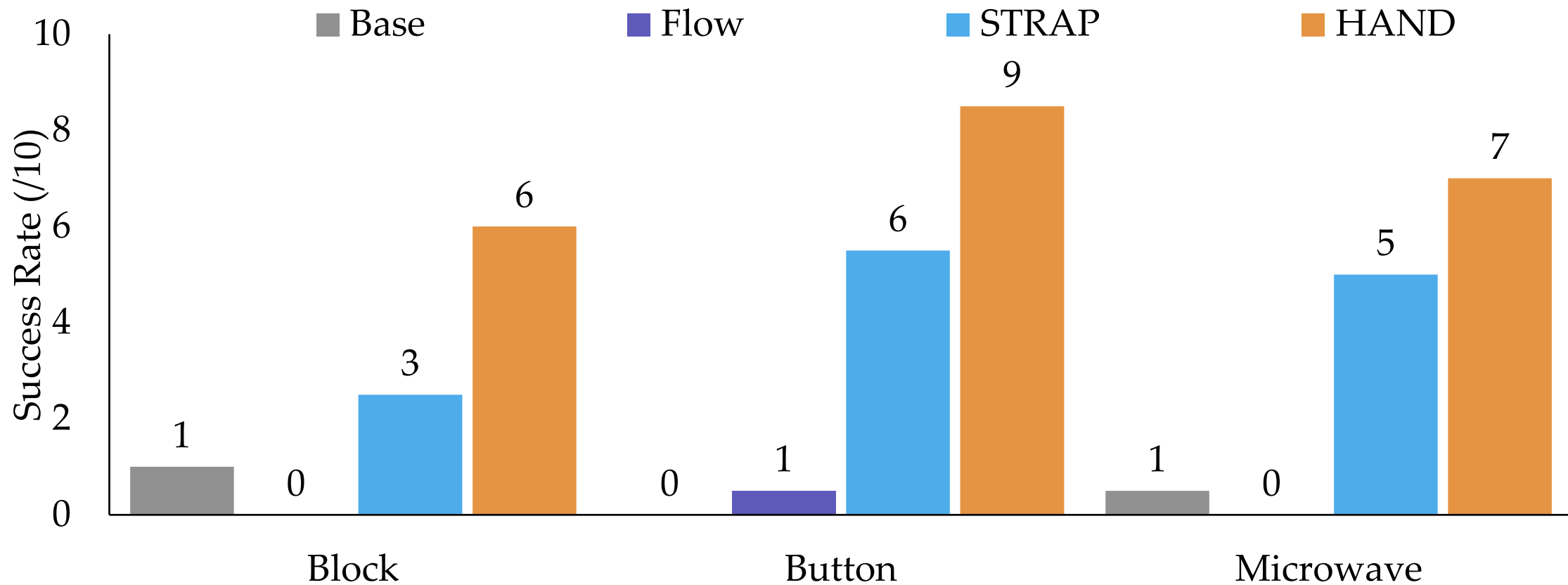
HAND (-VF)



HAND



Policy Performance



HAND Me the Data



Collect One Hand Demo
(~15 seconds)

Retrieval + Fine-Tune
(~2 minutes)

Evaluate
(~ 15 seconds)





HAND Me the Data:

Fast Robot Adaptation via Hand Path Retrieval

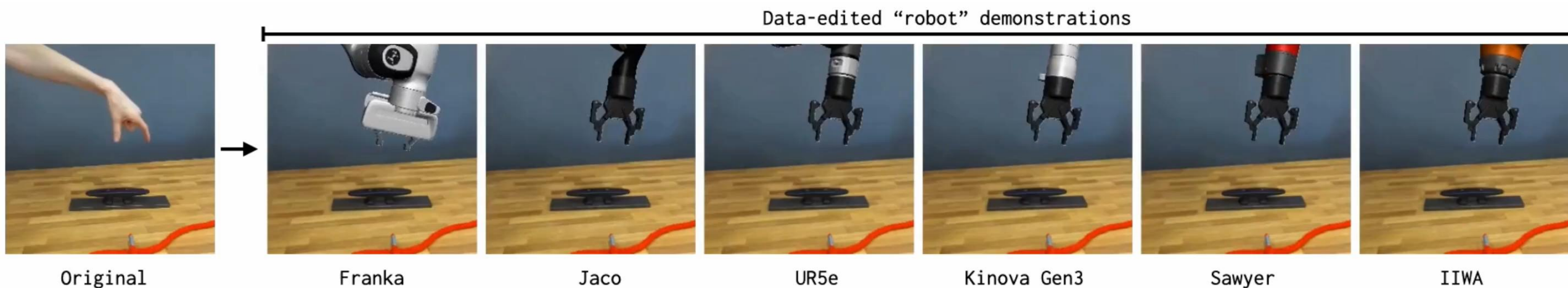
Website



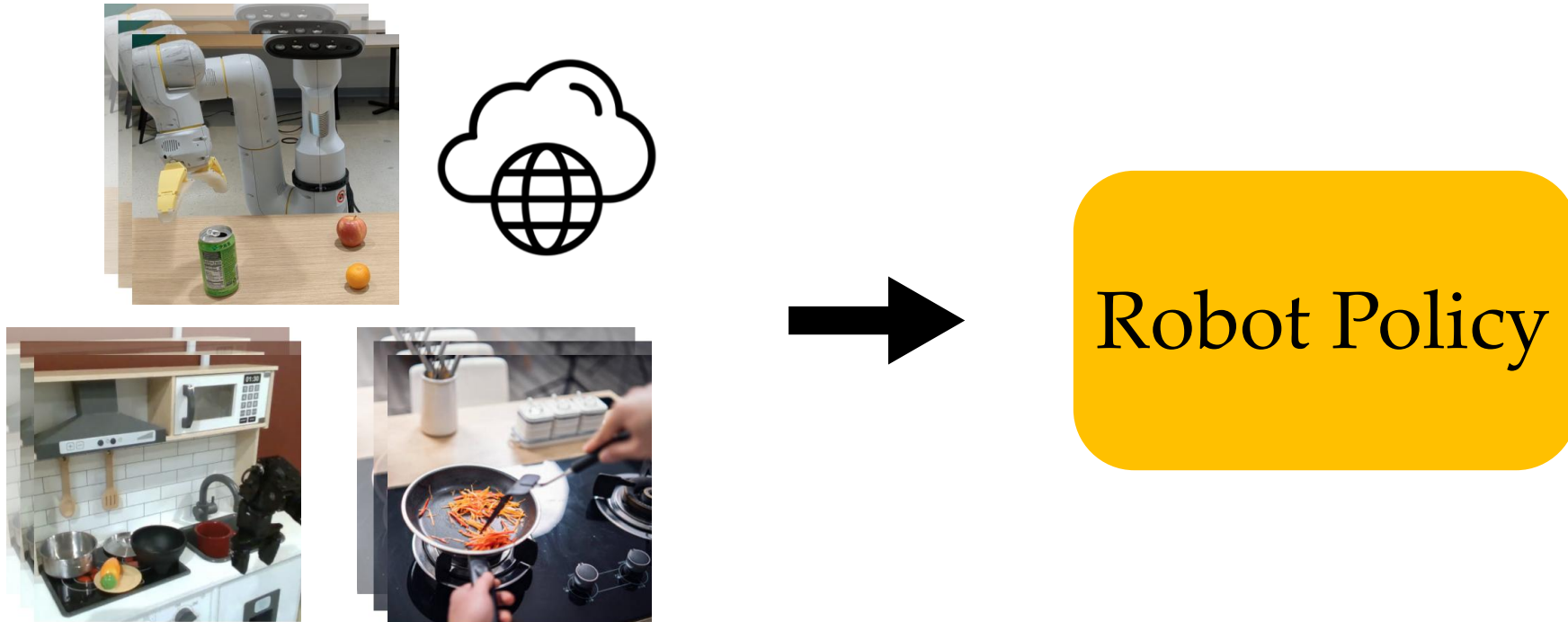
Today...

- Learning from **human** demonstrations
 - Retrieval-based methods
 - Latent action models
 - Mixing human and robot data

Human-to-Robot Transformation

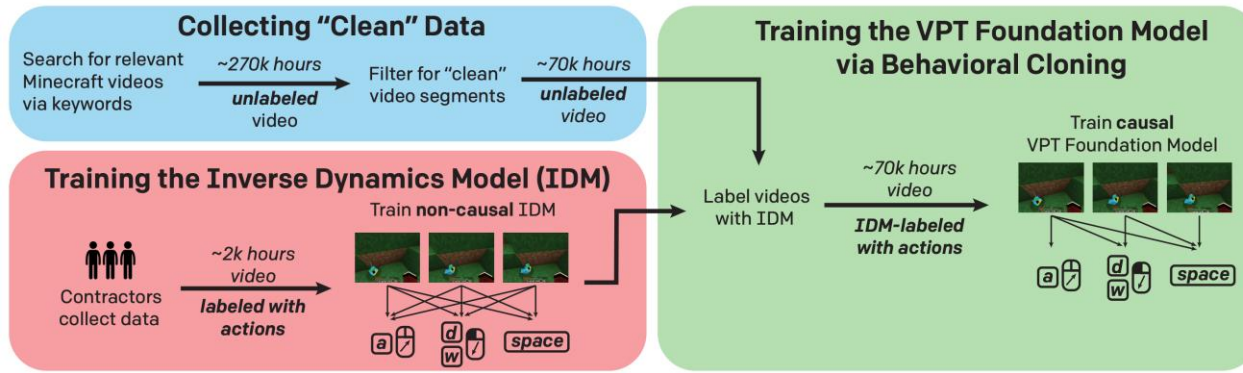


Learning from unlabeled observations

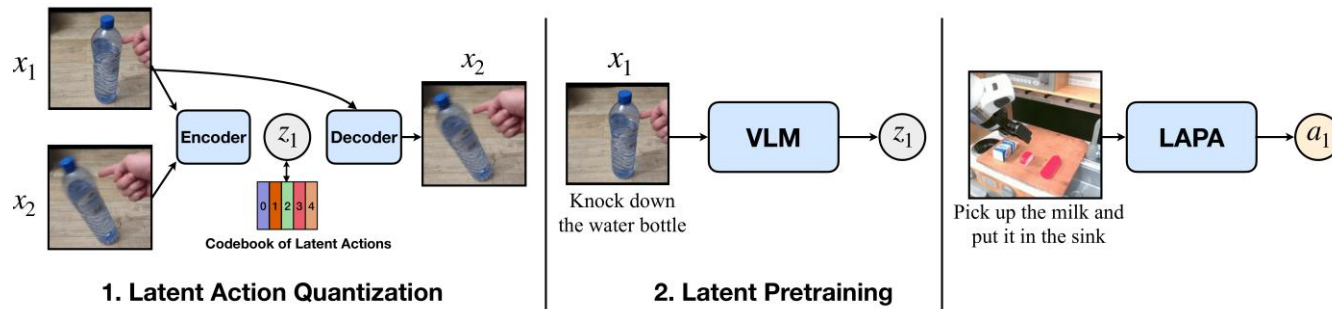


Missing action and reward labels!

Learning from Videos



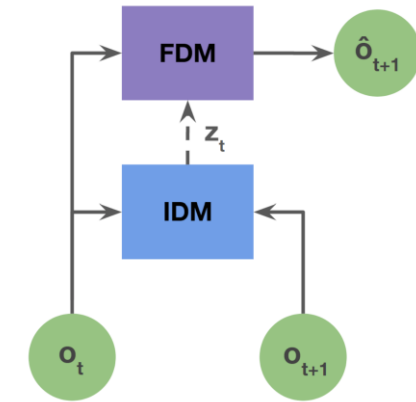
VPT [Baker et al., NeurIPS 2022]



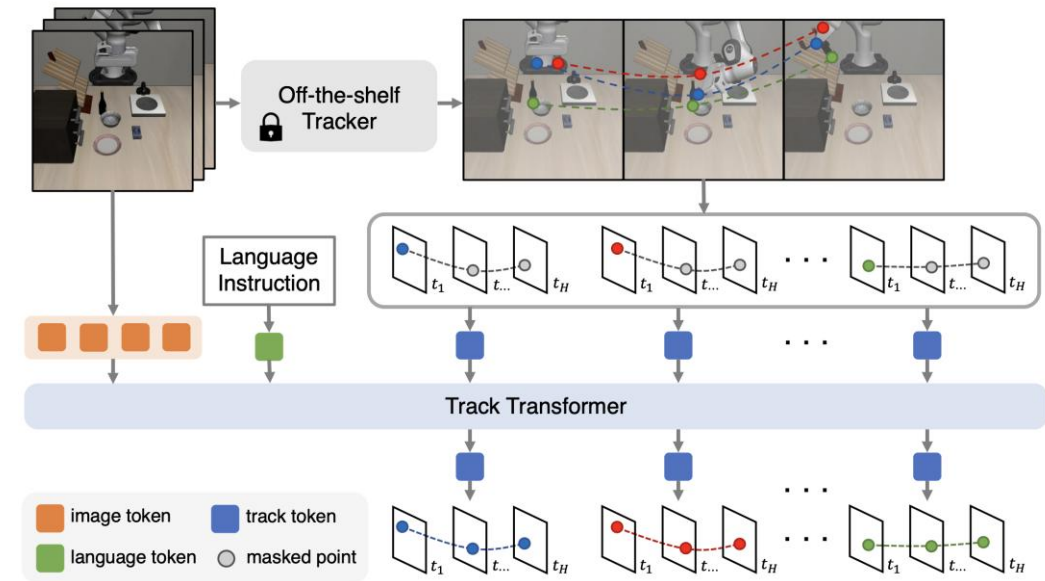
Latent Action Pretraining

Action Finetuning

LAPA [Ye et al., ICLR 2024]



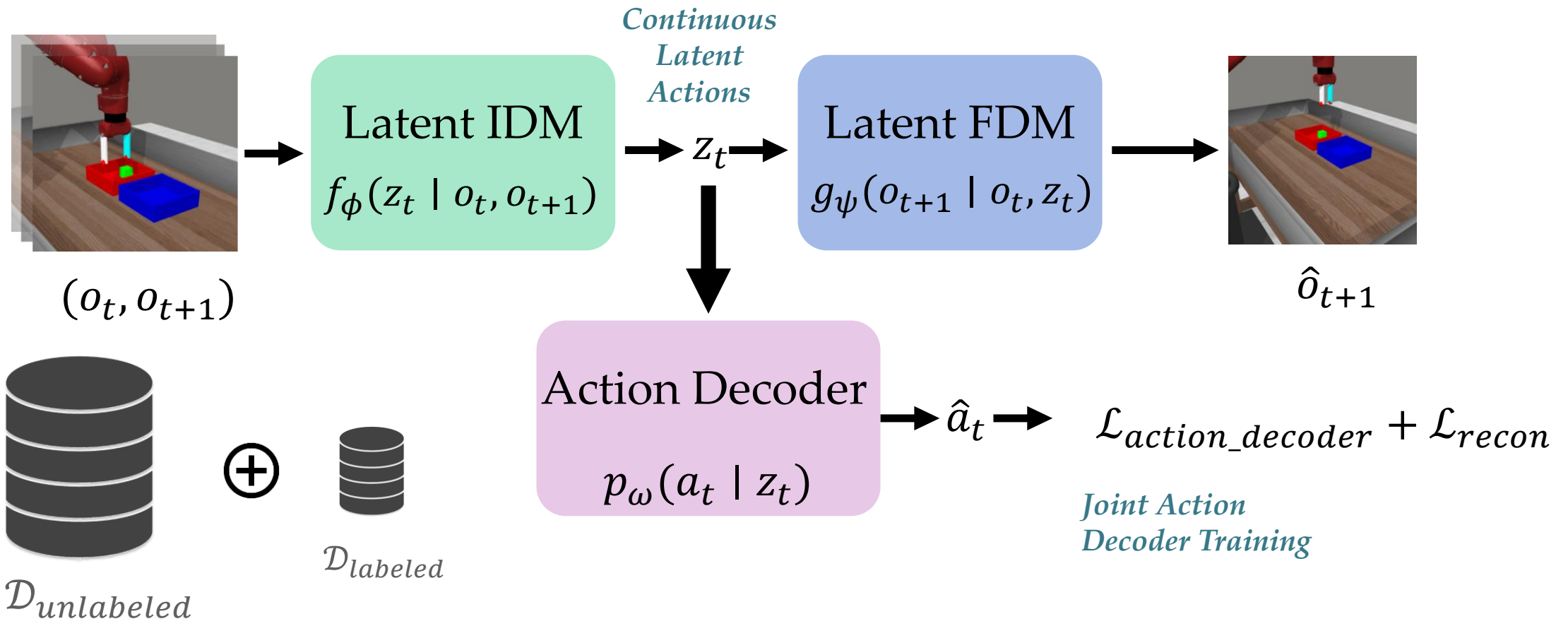
LAPO [Schmidt et al., ICLR 2024]



ATM [Wen et al., RSS 2024]

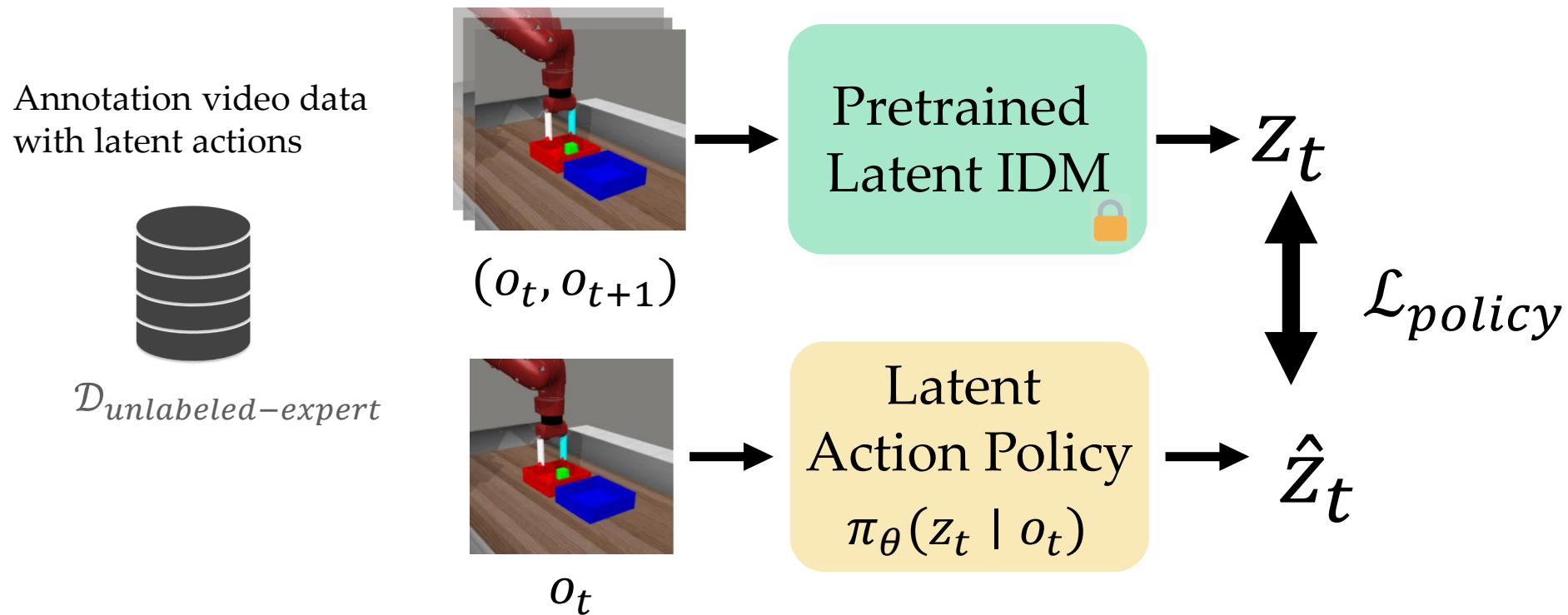
CLAM: Continuous Latent Action Models

Stage 1: CLAM Pretraining



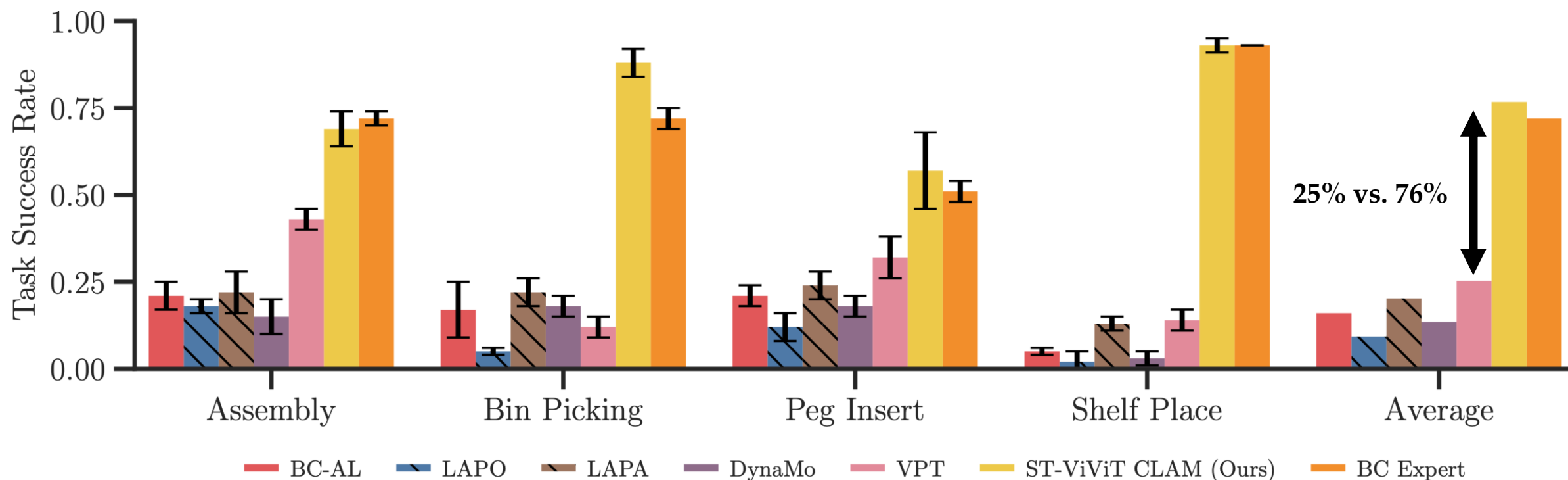
CLAM: Continuous Latent Action Models

Stage 2: Latent Action Policy Training



Results

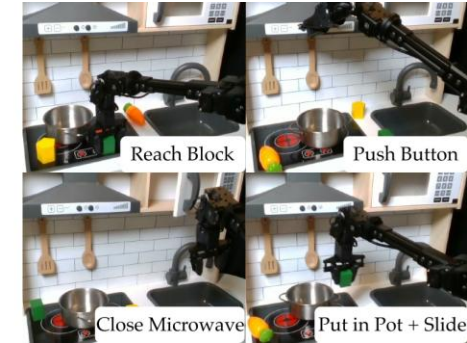
CLAM improves over best baseline (VPT) by more than 2x on average normalized return in DMControl and 2-3x success rate in MetaWorld



*Hatches denote methods that use discretized latent actions

Real WidowX Robot Experiments

- CLAM scales to policy learning on a WidowX robot arm
- $\mathcal{D}_{labeled}$ comprises of *task-agnostic, mixed expertise play* data
- **VPT** is the closest baseline, but struggles from labeled data
- **CLAM** achieves best performance



	Block	Button	Microwave	Slide Pot
BC-AL	0/10	0.5/10	1/10	0/10
LAPA	2/10	3/10	3/10	0/10
VPT	2.5/10	4/10	5/10	2/10
ST-CLAM	7/10	8.5/10	8/10	4/10

Reach Green Block



Close Microwave



Put Block in Pot and Slide Right

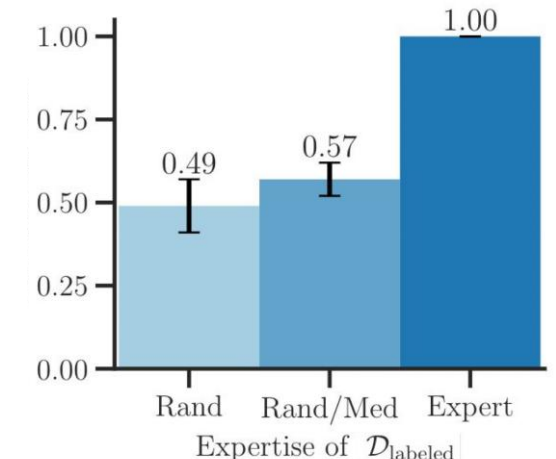
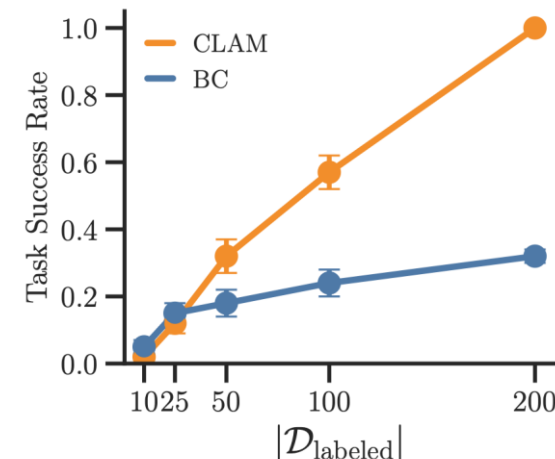
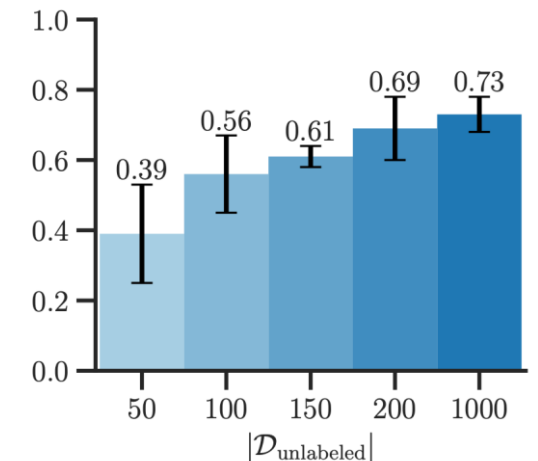
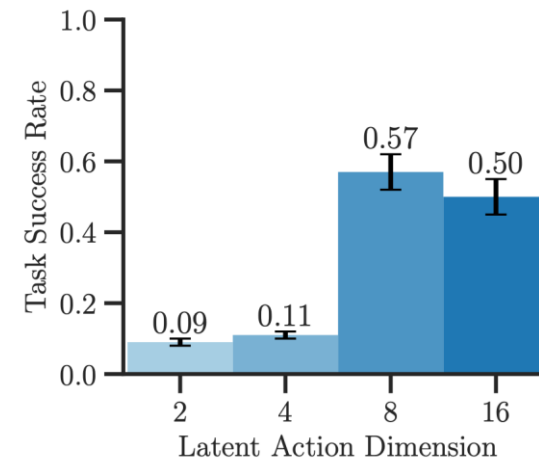


Continuous latent actions and *joint action decoding* significantly improves performance for continuous control problems

	Assembly	Bin Pick	Peg Insert	Shelf Place
Disc., $\neg$ JT	0.15 ± 0.03	0.12 ± 0.02	0.18 ± 0.04	0.19 ± 0.03
Disc., JT	0.14 ± 0.04	0.14 ± 0.03	0.17 ± 0.03	0.16 ± 0.04
Cont., $\neg$ JT	0.28 ± 0.04	0.18 ± 0.03	0.21 ± 0.08	0.26 ± 0.06
Cont., JT	0.69 ± 0.05	0.82 ± 0.04	0.57 ± 0.11	0.88 ± 0.02

CLAM Design Choices

- Latent action dimension affects model expressivity
- CLAM scales with $|\mathcal{D}_{unlabeled-expert}|$
- CLAM scales with $|\mathcal{D}_{labeled}|$ resulting a more accurate action decoder
- CLAM is robust to different expertise of action-labeled data



Today...

- Learning from **human** demonstrations
 - Retrieval-based methods
 - Latent action models
 - Mixing human and robot data