



National
Science
Foundation



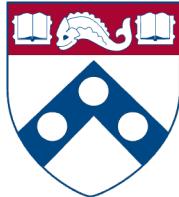
NEC Laboratories
America
Relentless passion for innovation

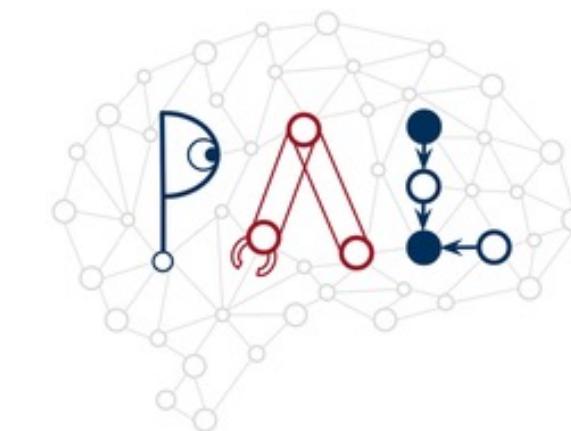


Goal-Based Task Specification For Robots in Vision, Language, and More

Dinesh Jayaraman

Assistant Professor, CIS & ESE, UPenn

 Penn Engineering | GRASP Laboratory
General Robotics, Automation, Sensing & Perception Lab



Perception,
Action, &
Learning Group

Towards General-Purpose Robots: Key Problems

- Failures In High-Information Settings

Task Difficulty and Information

Information \approx Unpredictability

Low information



High information



Controlled factory
automation settings



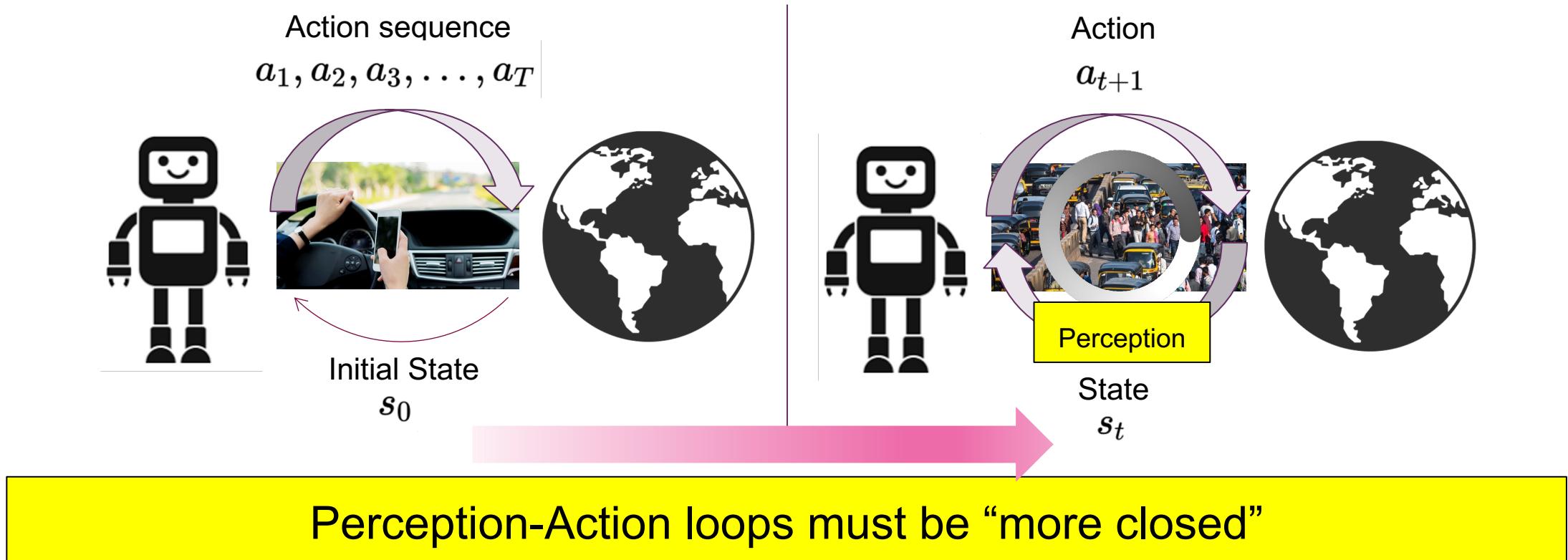
Fully observed and
well-modeled

Homes, offices, hospitals, small-
scale manufacturing ...



Unknown / stochastic dynamics, partial
observation, resource constraints ...

Perception Delivers Information



Towards General-Purpose Robots: Key Problems

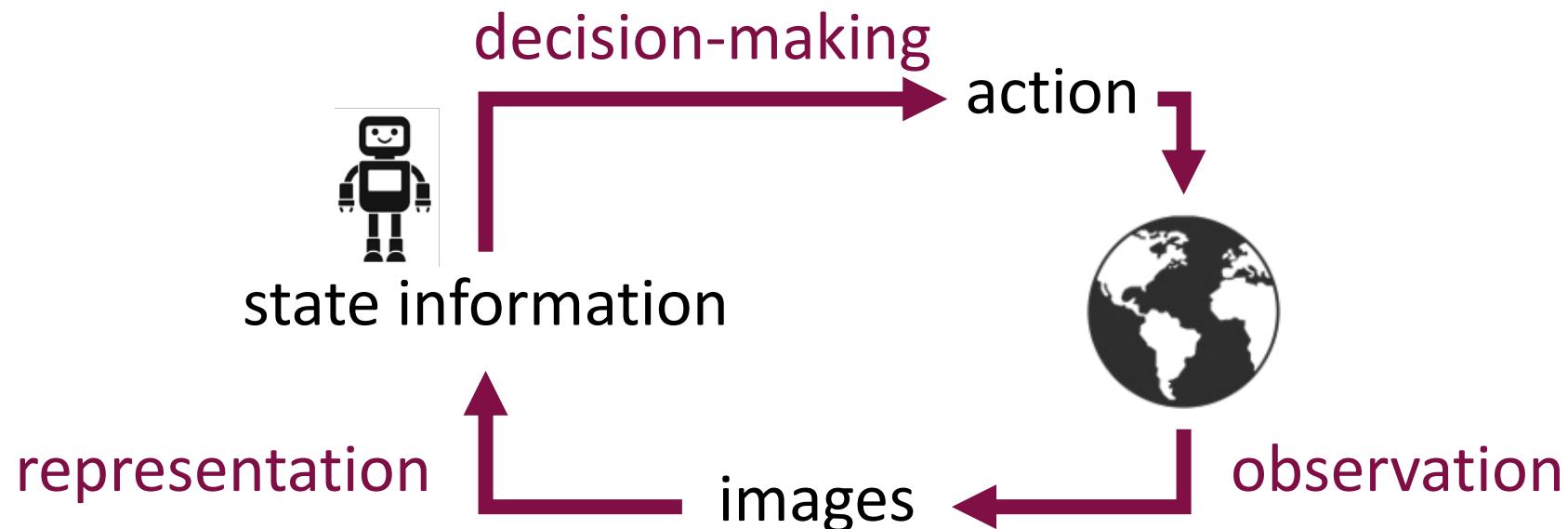
- Failures In High-Information Settings

Perception should carry high-throughput, low-latency information.

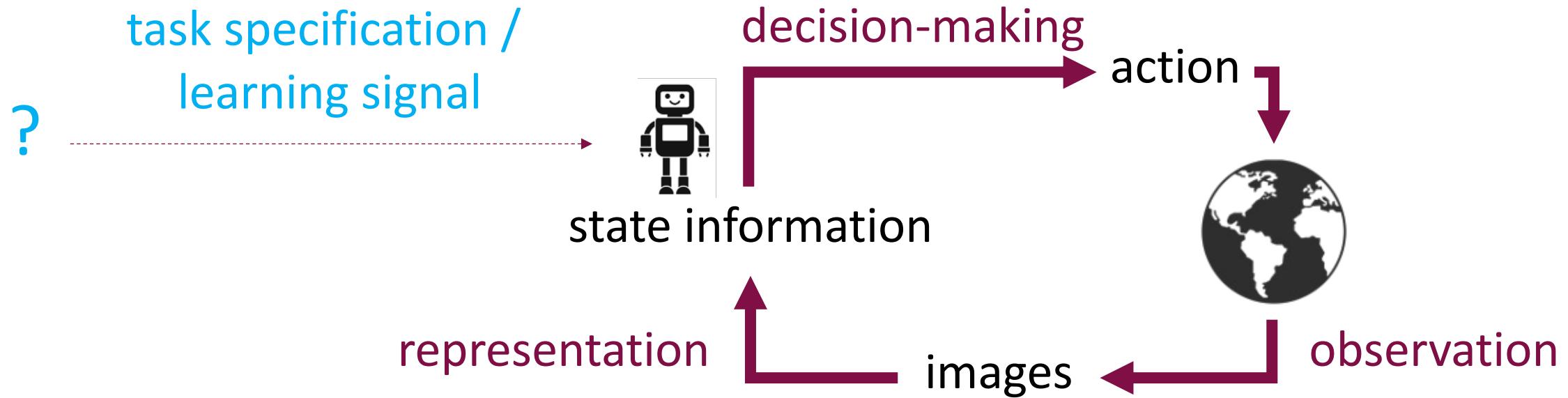
Vision-Based
Controllers

- Acquiring New Skills Specialized to Each Use Case

Learning?



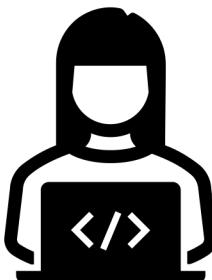
Perception-Action-Learning Loop



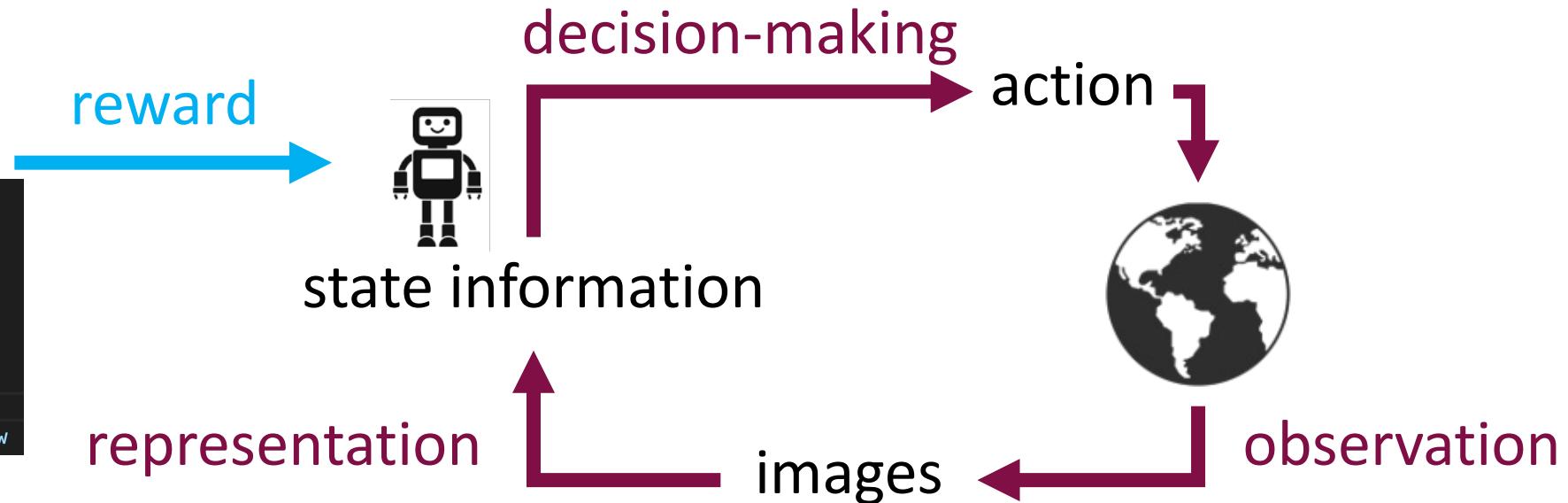
How to get learning signals to flexibly specify new skills for these large learned components in the controller?

Dense Rewards as Task Specifications

Painstaking manual
reward specification



```
total_rew = 0
if stage == 1:
    total_rew += door_close_rew
elif stage == 2:
    total_rew += insert_key_rew
elif stage == 3:
    total_rew += turn_key_rew
else:
    total_rew += turn_key_back_rew
```



Expertise-intensive, inaccessible to a lay user

Often relies on true state information

Task-specific robot-experience-intensive

Does not scale to large numbers of skills

Towards General-Purpose Robots: Key Problems

- Failures In High-Information Settings

Perception should carry high-throughput, low-latency information.

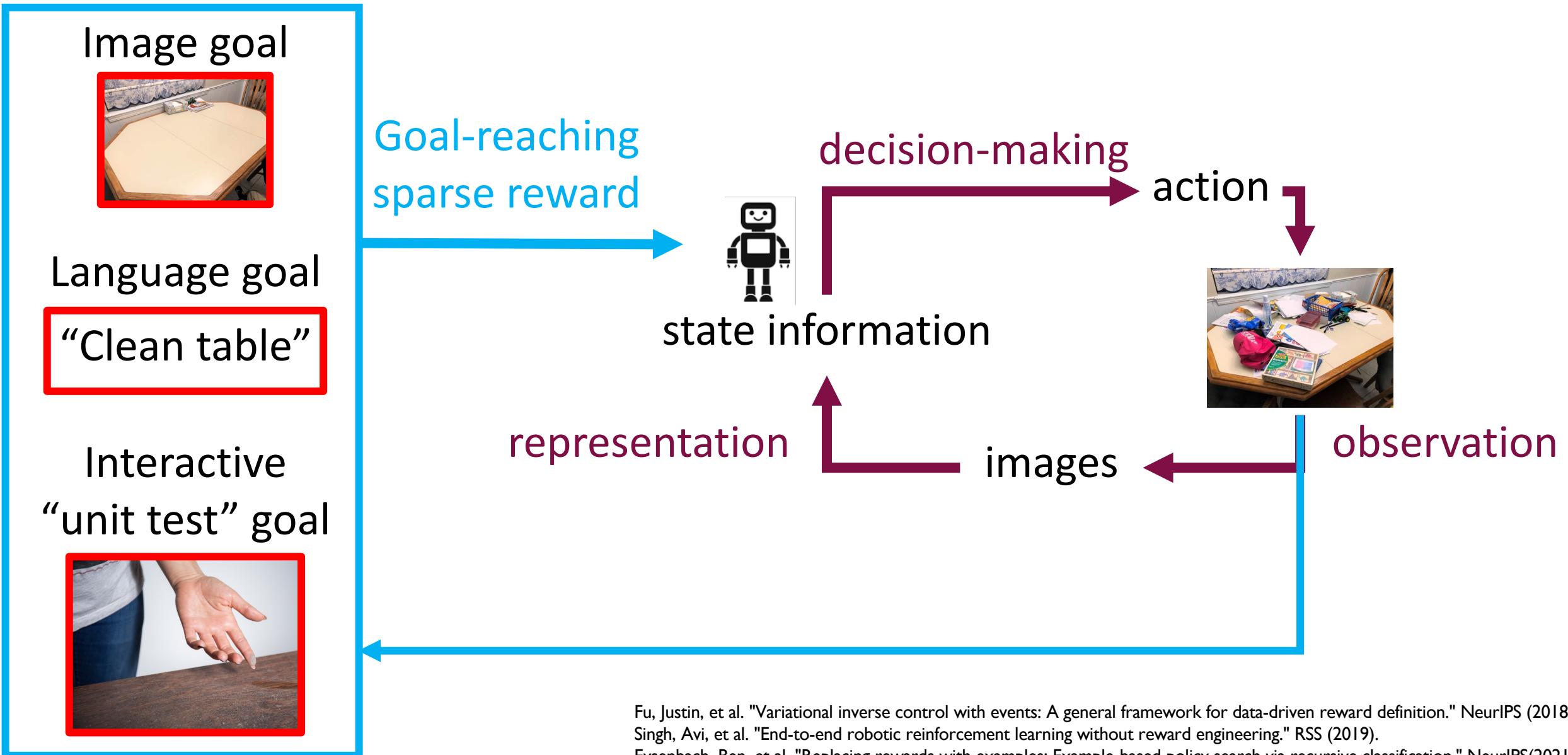
Vision-Based
Controllers

- Acquiring New Skills Specialized to Each Use Case

Versatile human-robot interfaces for task specification / teaching.

Multimodal Goals For
Robot Learners

Goal Specifications for Vision-Based Robot Learning



Talk Outline

- Language and Image-Based Goal Specifications
 - Ma et al, VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training. ICLR 2023
 - Ma et al, Language-Image Representations and Rewards for Robotic Control (under review)
- Physical Objects as Goal Specifications
 - Huang et al, Training Robots to Evaluate Robots: Example-Based Interactive Reward Functions for Policy Learning. CORL 2022
- Exploration to Discover Goal-Based Skills
 - Hu et al. Planning Goals for Exploration. ICLR 2023

Talk Outline

- Language and Image-Based Goal Specifications
 - Ma et al, VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training. ICLR 2023
 - Ma et al, Language-Image Representations and Rewards for Robotic Control (under review)
- Physical Objects as Goal Specifications
 - Huang et al, Training Robots to Evaluate Robots: Example-Based Interactive Reward Functions for Policy Learning. CORL 2022
- Exploration to Discover Goal-Based Skills
 - Hu et al. Planning Goals for Exploration. ICLR 2023

Interactively Perceiving Task Rewards For Training RL Agents

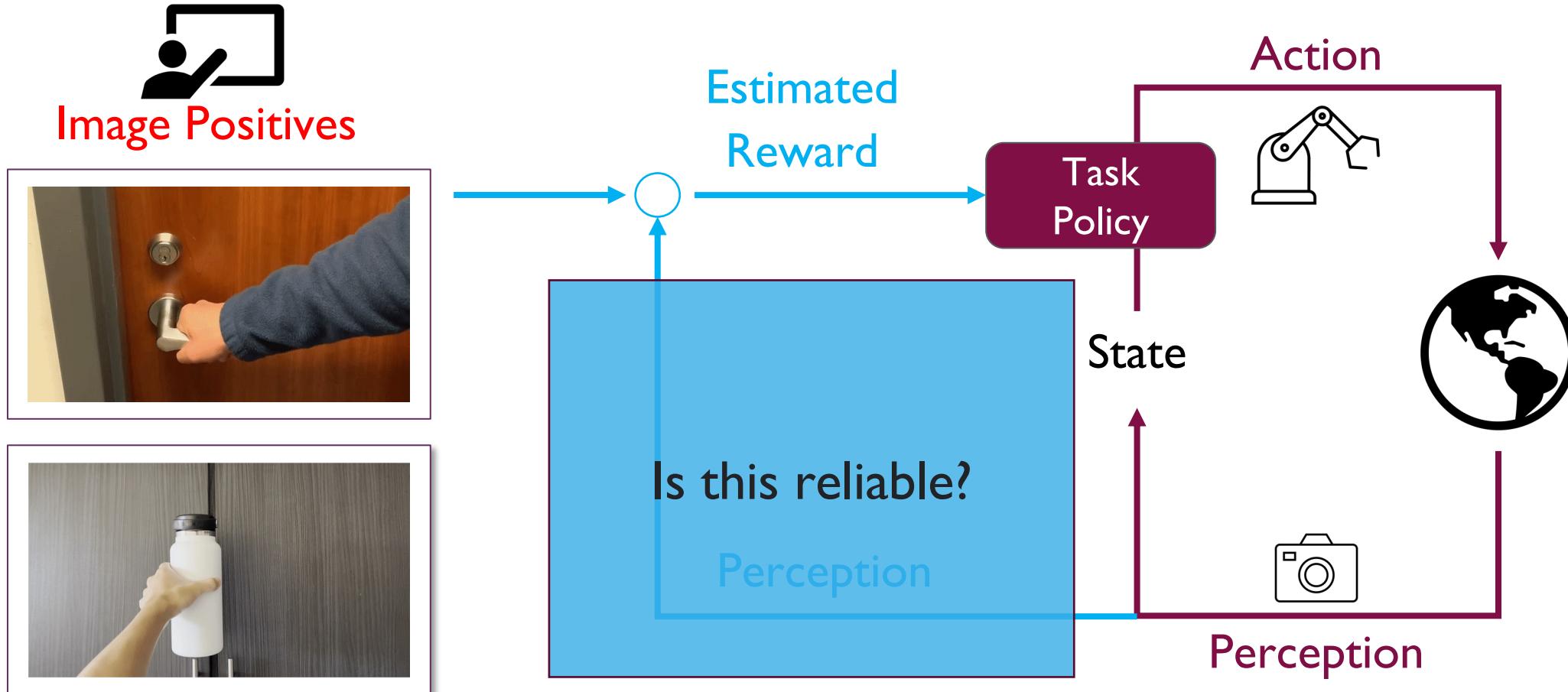
Kun Huang, Edward Hu, and Dinesh Jayaraman,

“Training Robots to Evaluate Robots: Interactive Reward Functions for Task Policy Learning”

CORL 2022 (Best Paper Award)



Goal Snapshots Might Not Fully Specify A Task

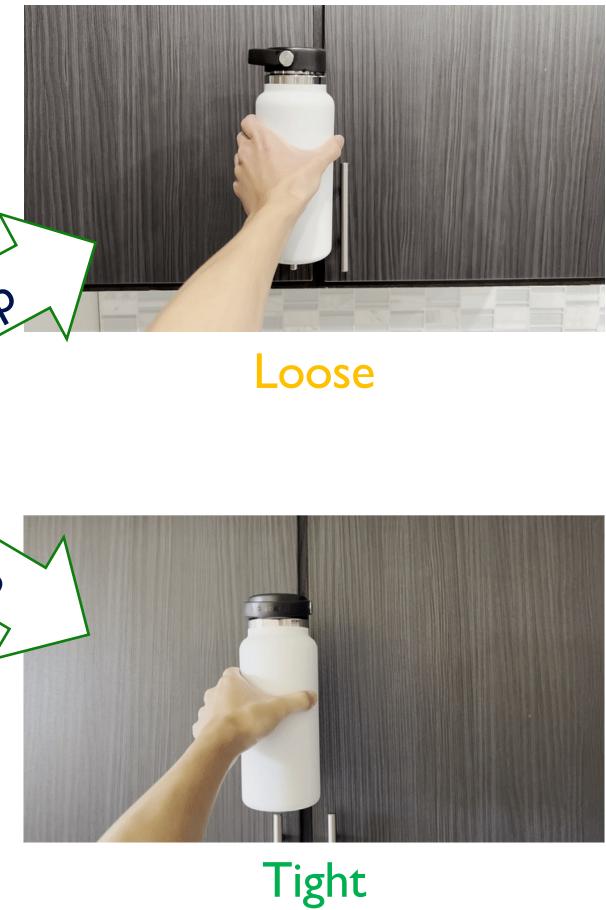
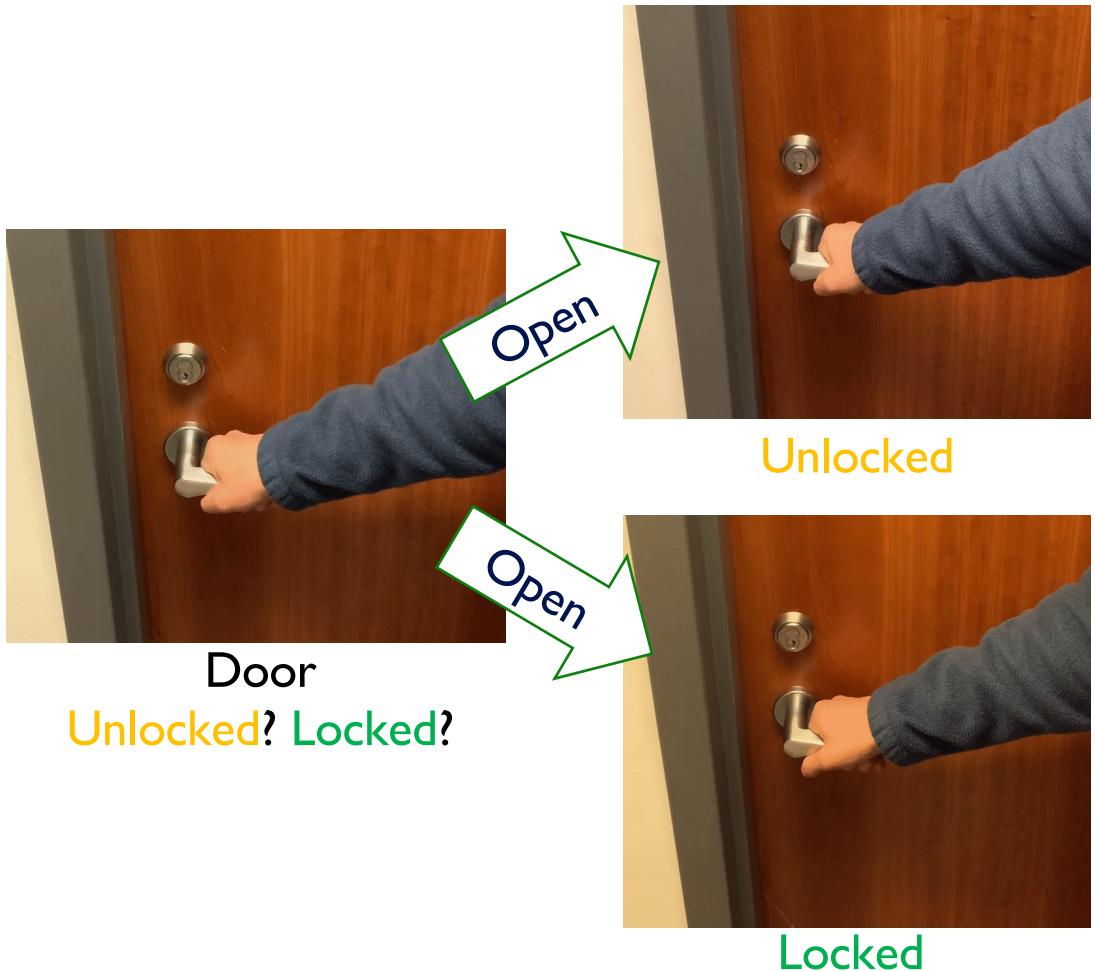


Fu, Justin, et al. "Variational inverse control with events: A general framework for data-driven reward definition." *Advances in neural information processing systems* 31 (2018).

Singh, Avi, et al. "End-to-end robotic reinforcement learning without reward engineering." *Robotics: Science and Systems* (2019).

Eysenbach, Ben, et al. "Replacing rewards with examples: Example-based policy search via recursive classification." *Advances in Neural Information Processing Systems* 34 (2021).

Perceiving Task Rewards is Often Hard!



**Specifying such verification behaviors would
constitute a more complete goal specification**

Akin to a kind of interactive unit test for software development

Verification Behaviors As Task Specifications

Task: Close & Lock Door

Reward: 

Unlocked



Reward: 

Locked

But where do verification behaviors come from?

Could we automatically acquire “interactive reward function” policies to specify the task to a skill learner?

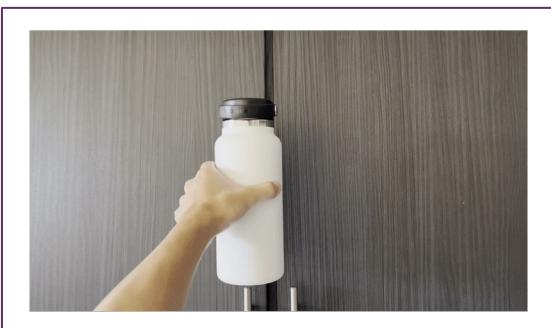
Solution: ~~Image Snapshots~~ Actionable Examples

Image Positives



Actionable Positives

Disambiguating actions

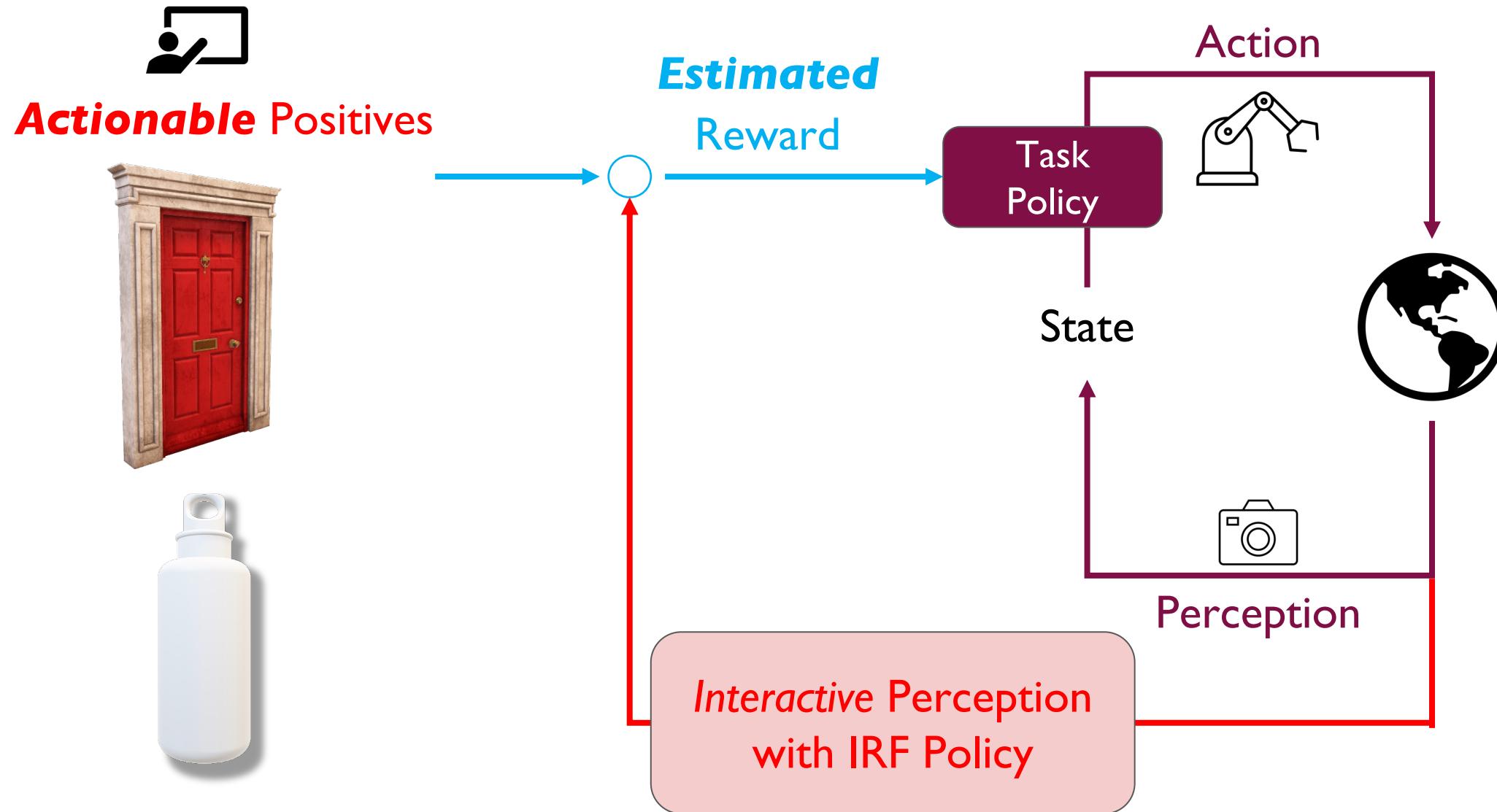


Provided by the user

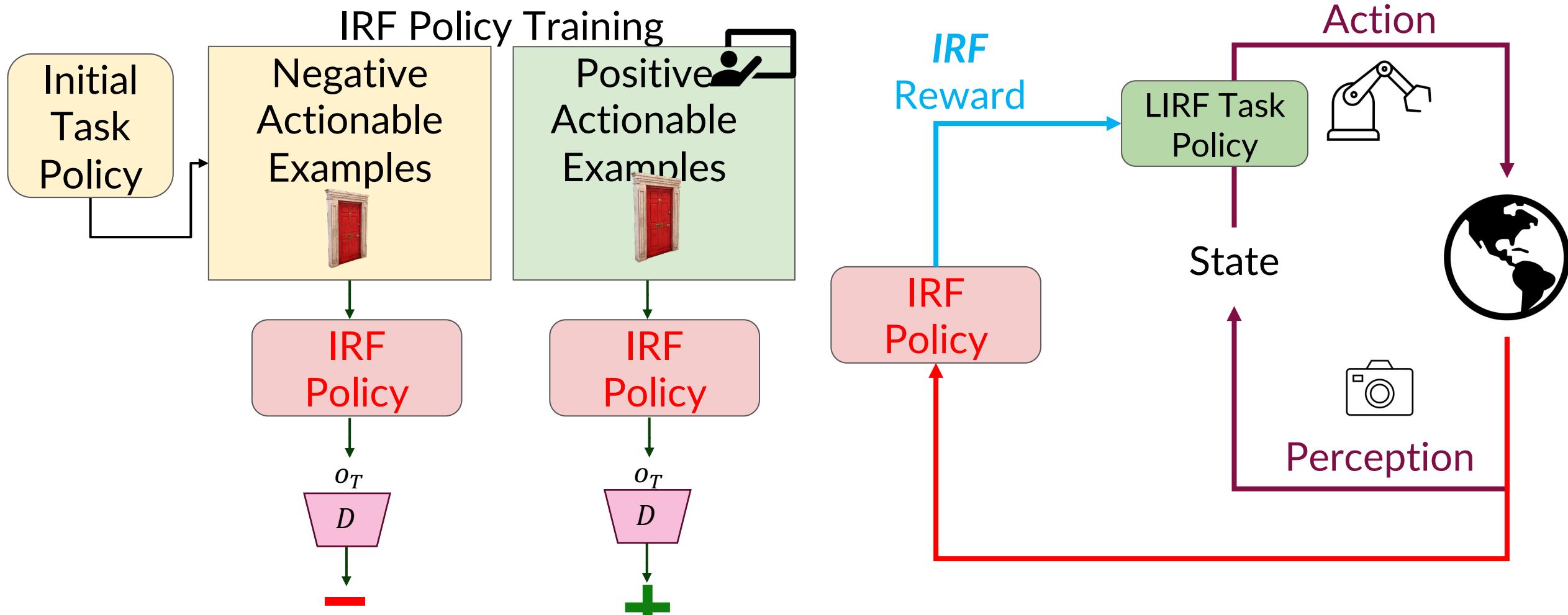


We will learn these with
RL!

RL Training Loop with Actionable Positives



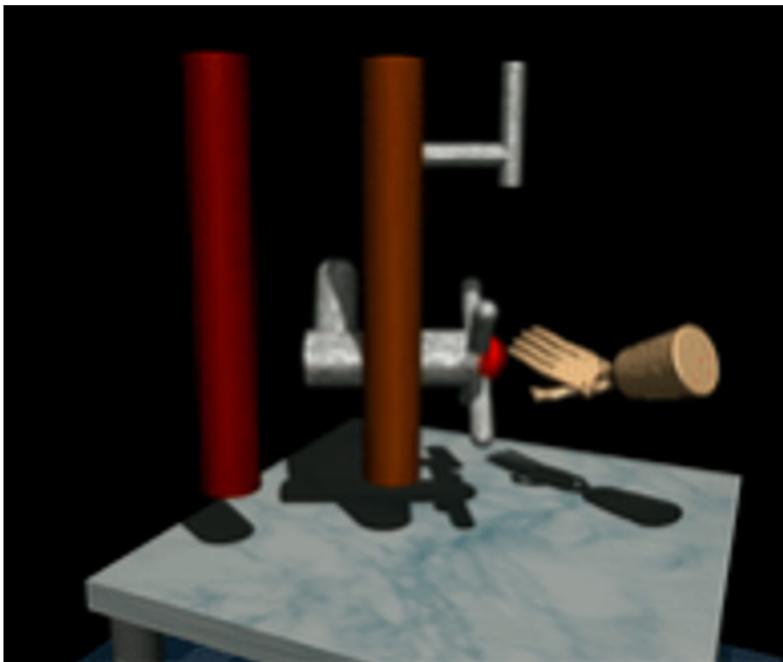
Learning from Interactive Reward Functions (LIRF) Framework



Bonus: IRF policy can even run at test time, as an in-the-loop verification behavior!
“Run the task policy until the IRF evaluation looks good.”

Experiments: Qualitative Results

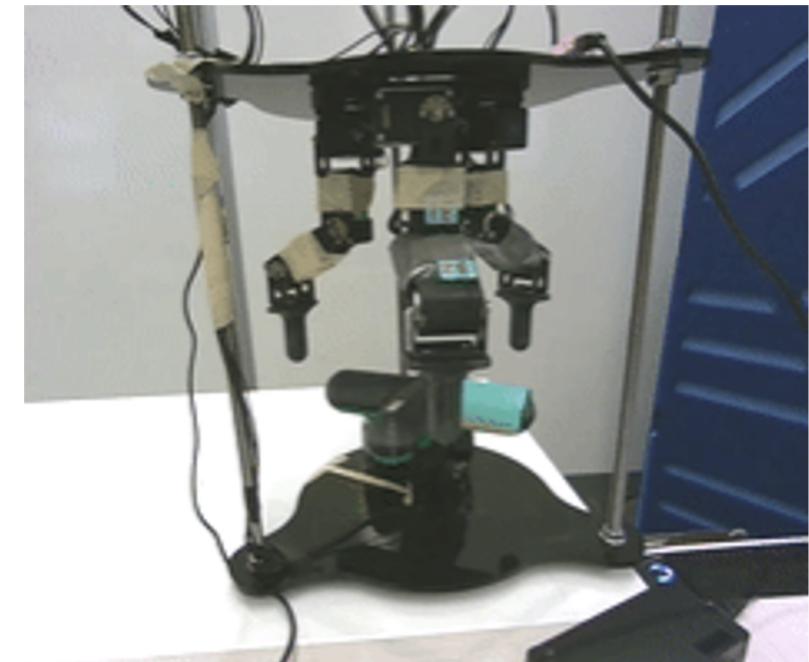
Door Locking



Block Stacking



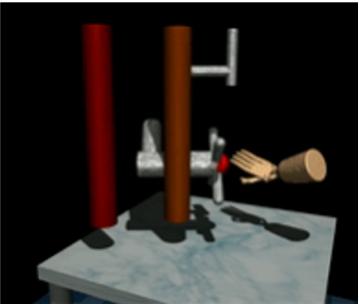
Screwing



Block colors for visualization only.
Green = heaviest block.

Task Policy Success Rates

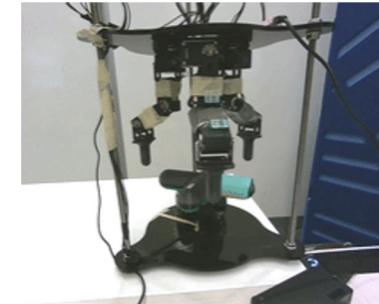
Door Locking



Block Stacking



Screwing

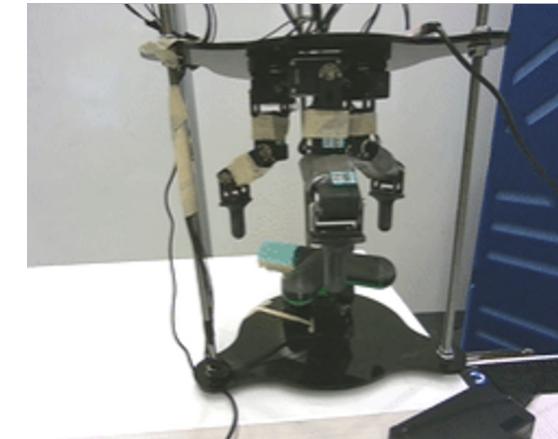


[1] Fu, Justin, et al. "Variational inverse control with events: A general framework for data-driven reward definition." *Advances in neural information processing systems* 31 (2018).

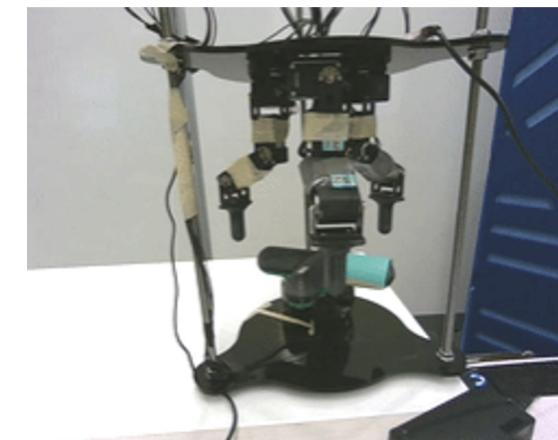
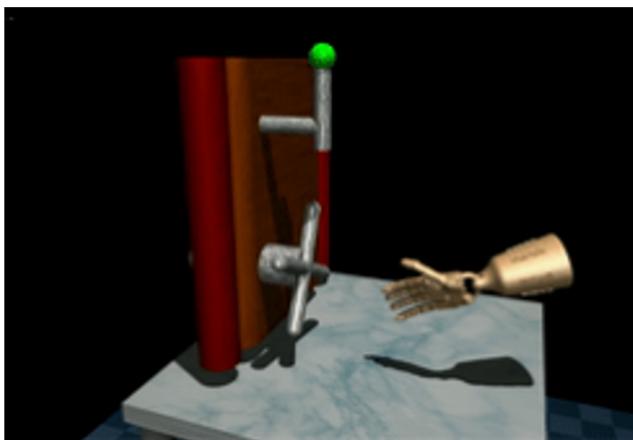
[2] Torabi, Faraz, Garrett Warnell, and Peter Stone. "Generative adversarial imitation from observation." *arXiv preprint arXiv:1807.06158* (2018).

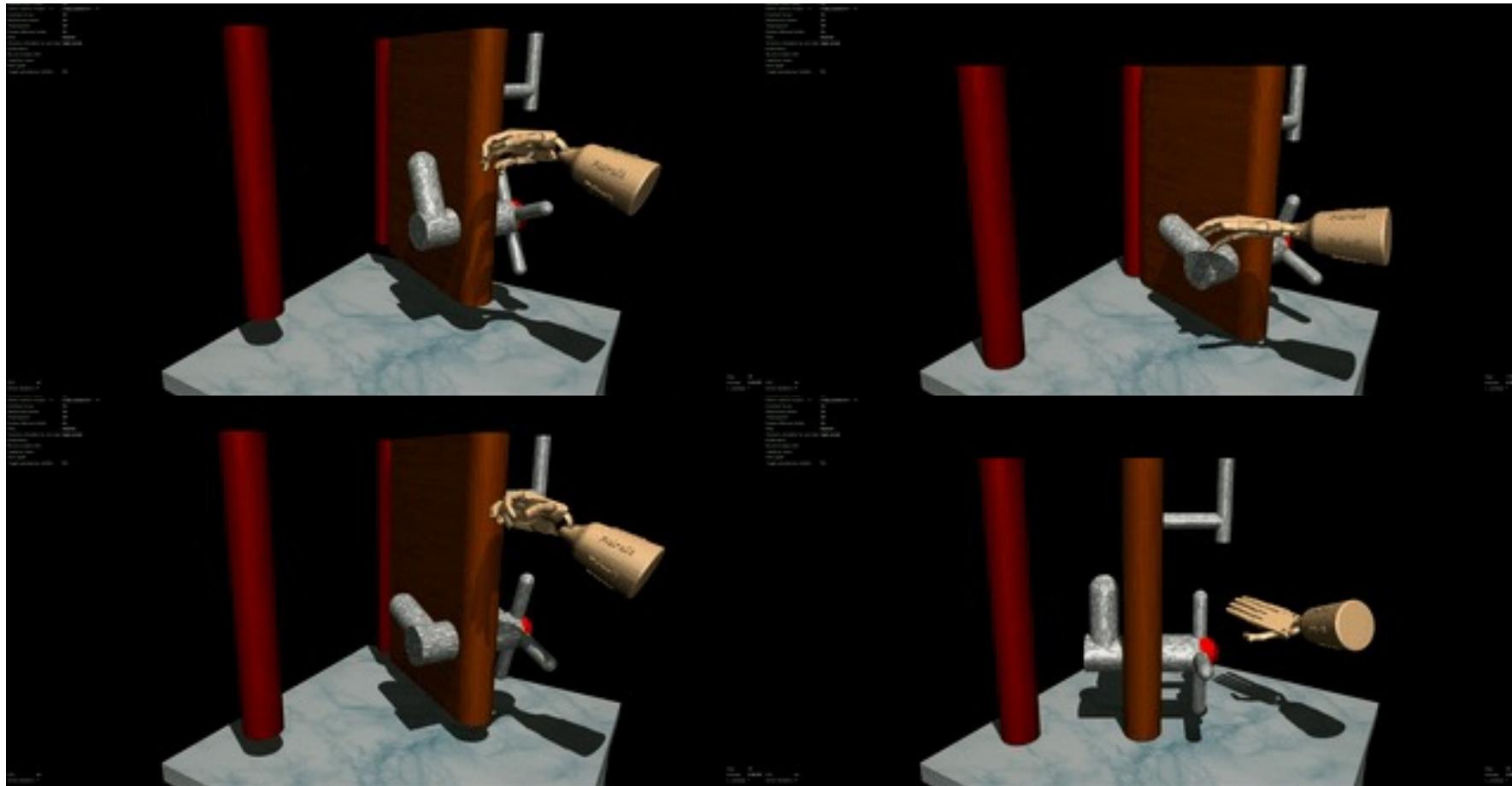
IRF Policy Rollouts

On a positive example



On a negative example





- Red ball appears: LIRF policy execution; Green ball appears: IRF policy execution

Takeaway

Physical objects can be used as “actionable examples” of desirable goal states



learning interactive “unit test” functions to specify skills



learning actual skill policies

Talk Outline

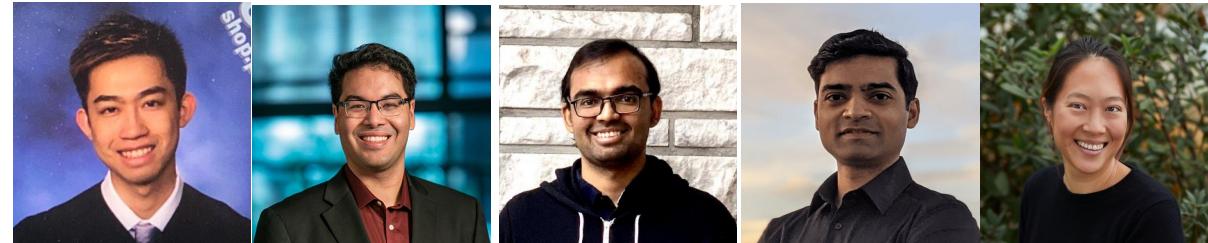
- Language and Image-Based Goal Specifications
 - Ma et al, VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training. ICLR 2023
 - Ma et al, Language-Image Representations and Rewards for Robotic Control (under review)
- Physical Objects as Goal Specifications
 - Huang et al, Training Robots to Evaluate Robots: Example-Based Interactive Reward Functions for Policy Learning. CORL 2022
- Exploration to Discover Goal-Based Skills
 - Hu et al. Planning Goals for Exploration. ICLR 2023

Vision-Language Representations For Robot Learning

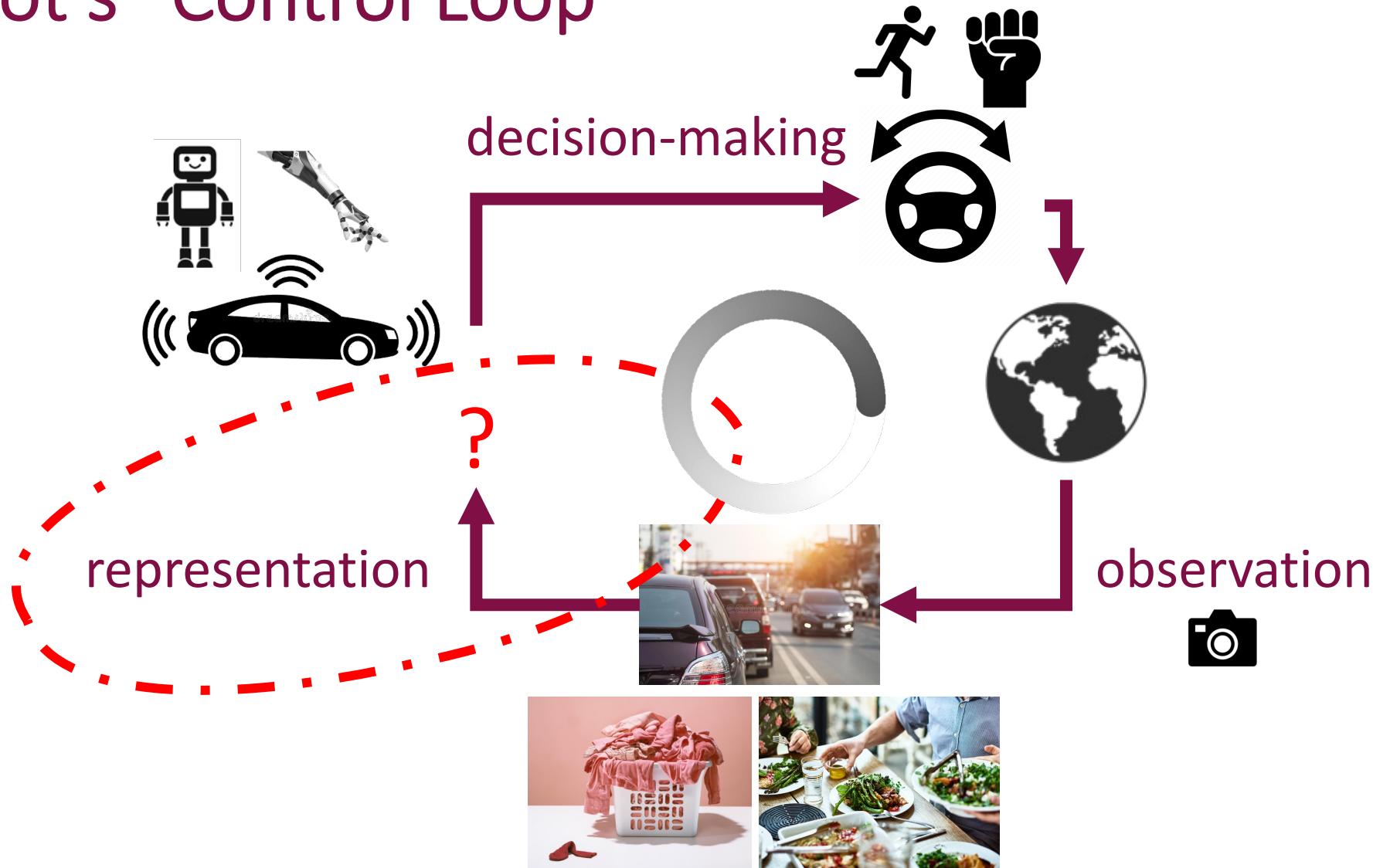
With Jason Ma, Osbert Bastani (UPenn), Shagun Sodhani, Vikash Kumar (FAIR), Amy Zhang (FAIR & UT Austin)

VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training, ICLR 2023

Language-Image Representations and Rewards for Robotic Control. (under review)



A Robot's “Control Loop”



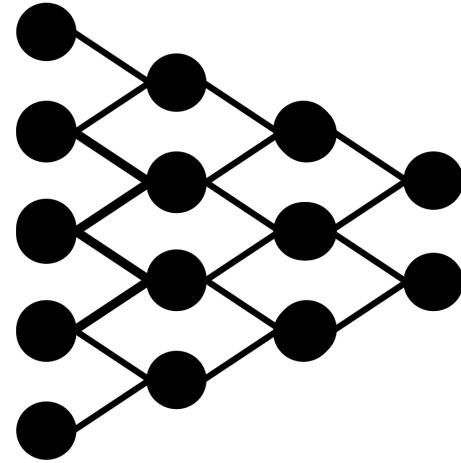
How should the robot represent the information in its visual observations?

What is a Good Visual Representation *for Recognition?*

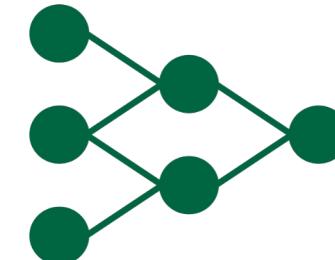
Image x



$\phi(\cdot)$



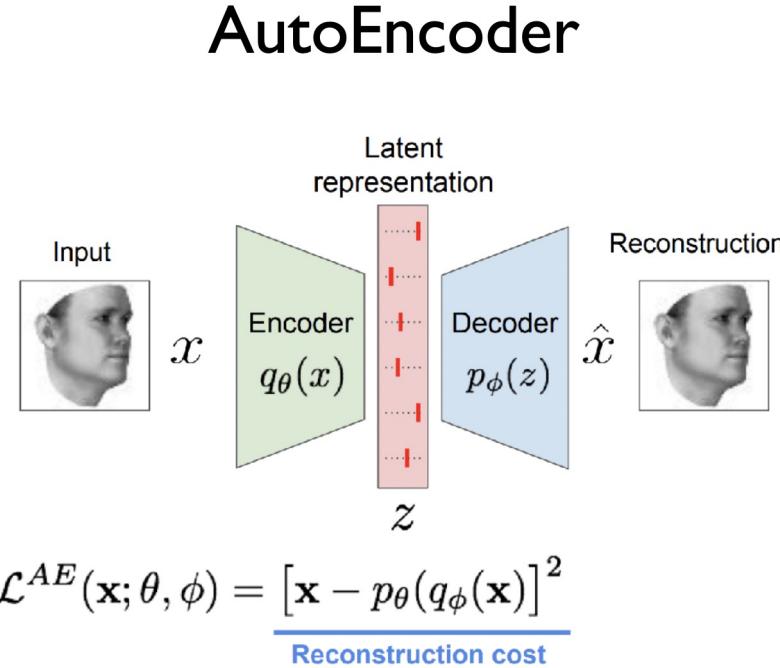
$\phi(x)$



“dog”

Good representations organize information conveniently *for the task.*

Aside: An Intro to Visual Representation Learning Autoencoders

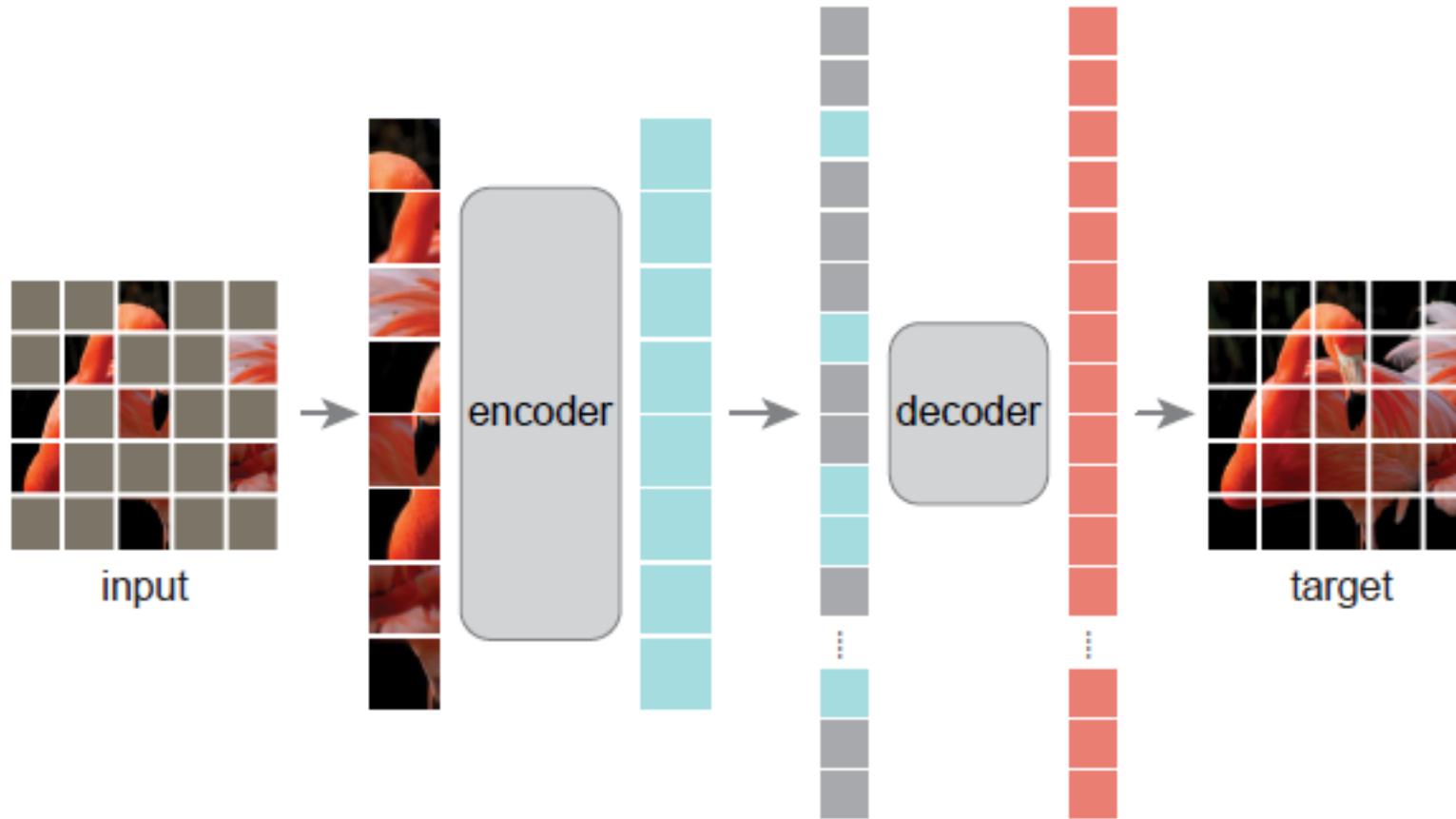


Slide from Alex Graves

Variations of this include: *Variational Autoencoders*

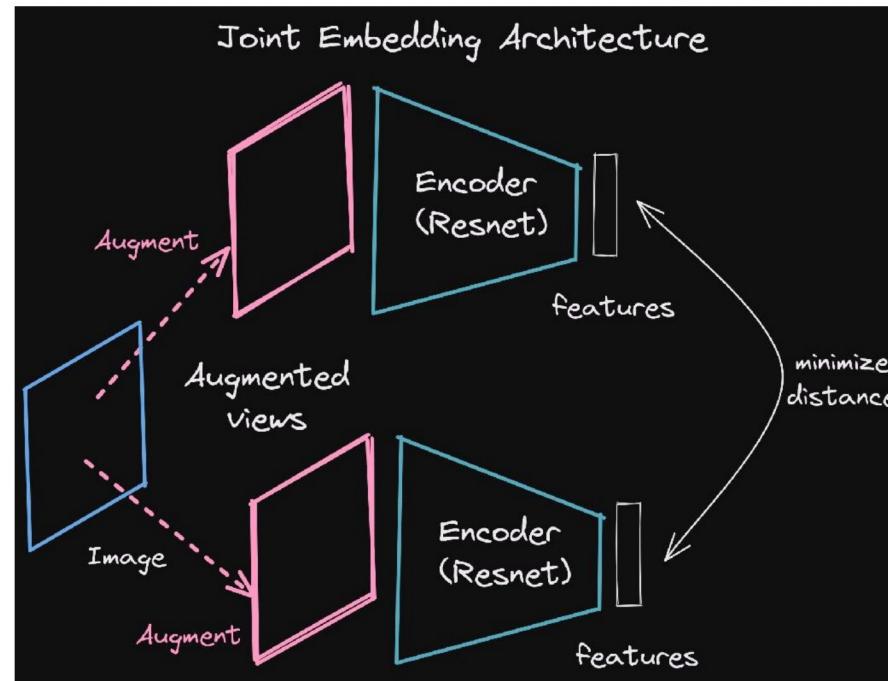
Aside: An Intro to Visual Representation Learning

Masked Autoencoders



Aside: An Intro to Visual Representation Learning

Contrastive Learning

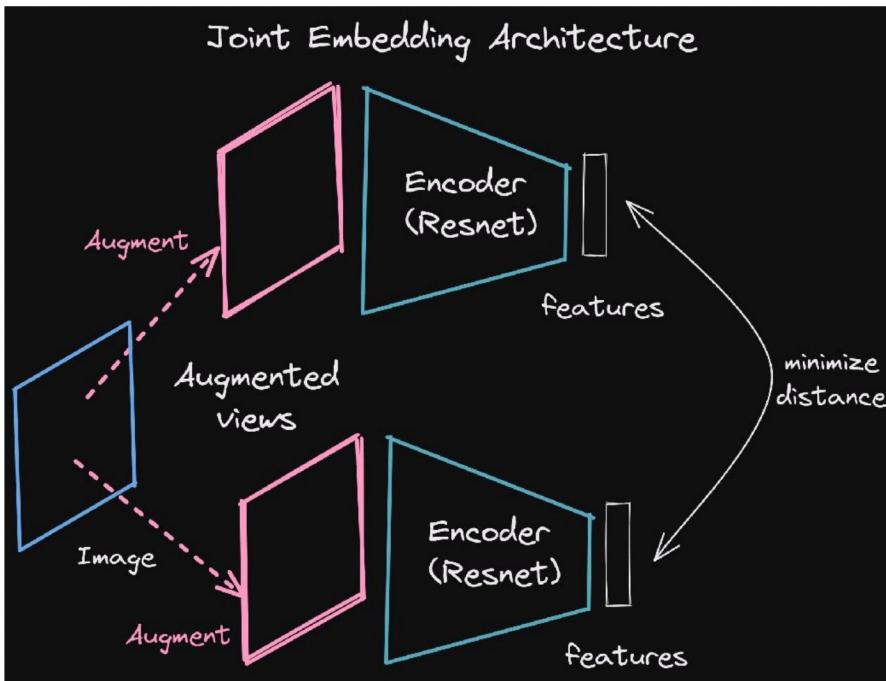


But what is to stop the representation from collapsing to $\mathbf{z}(x) = \mathbf{0} \forall x$?

Contrastive Learning

Aside: An Intro to Visual Representation Learning

Contrastive Learning



Contrastive Learning

Take a datapoint (an image), and try to fit a scoring function to make sure it aligns more with a positive relative to a negative.

$$score(x, x_{pos}) > score(x, x_{neg})$$

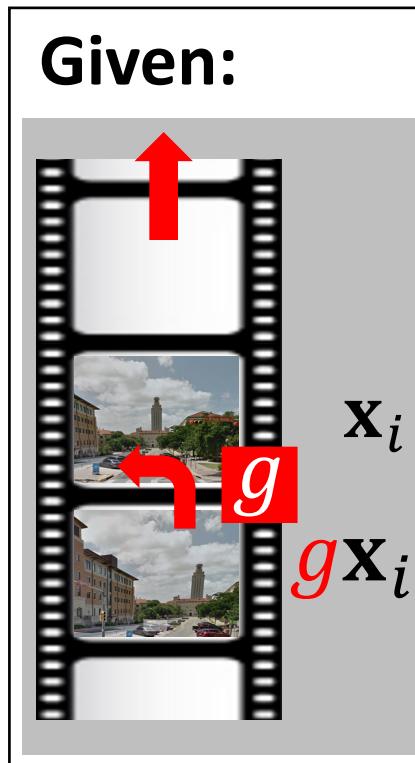
$$L_{\text{InfoNCE}} = -\mathbb{E} \left[\log \frac{s(x, x_{pos})}{s(x, x_{pos}) + \sum_{y_j \neq x_{pos}} s(x, y_j)} \right]$$

Slide adapted from Aaron van den oord

This can be shown to approximate a lower bound on MI between the two views

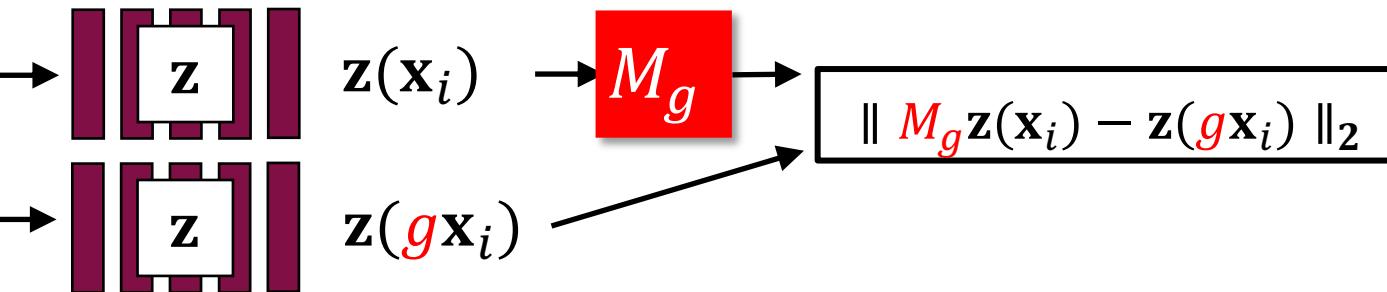
1. Representation Learning with Contrastive Predictive Coding (van den Oord et al 2018)
2. Improved Deep Metric Learning with Multi-Class N-Pairs Loss - (Sohn et al 2016)
3. Deep InfoMax, AMDIM (Hjelm, Bachman, et al 2019)

Egomotion-Equivariant Contrastive Representations

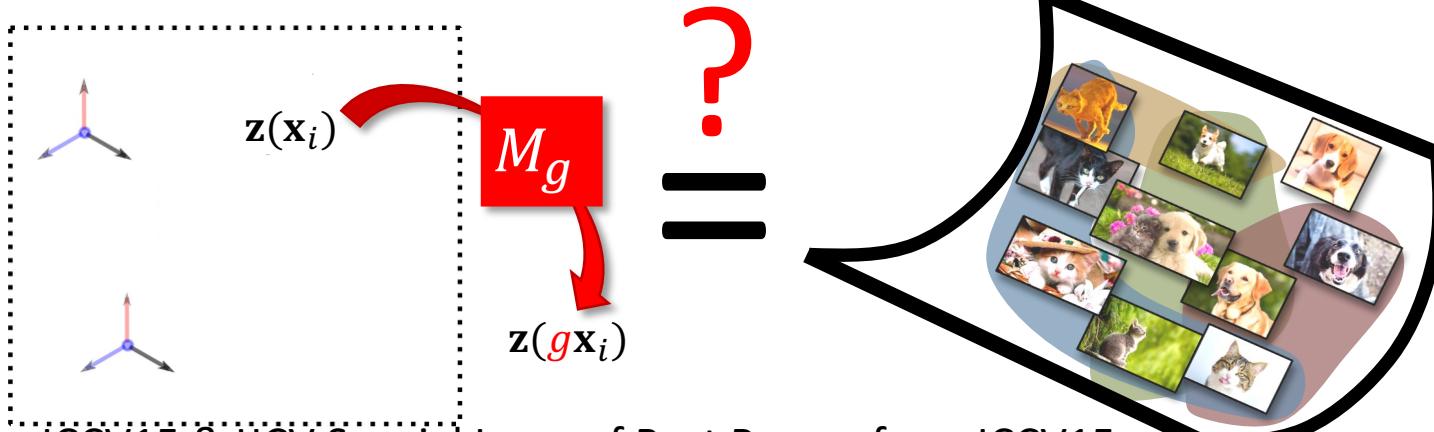


Desired: for all motions g and all images \mathbf{x} ,

$$\mathbf{z}(g\mathbf{x}) \approx M_g \mathbf{z}(\mathbf{x})$$



Feature space:

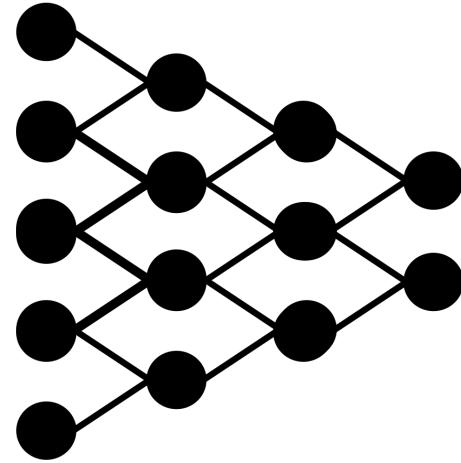


What is a Good Visual Representation *for Recognition?*

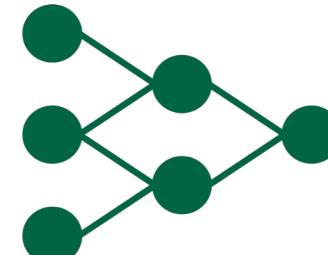
Image x



$\phi(\cdot)$



$\phi(x)$



“dog”

Good representations organize information conveniently *for the task*.

Self-supervised representations: contrastive learning, masked autoencoding etc.

Great results *for recognition*. Recently shown to also transfer to robots sometimes, but ...

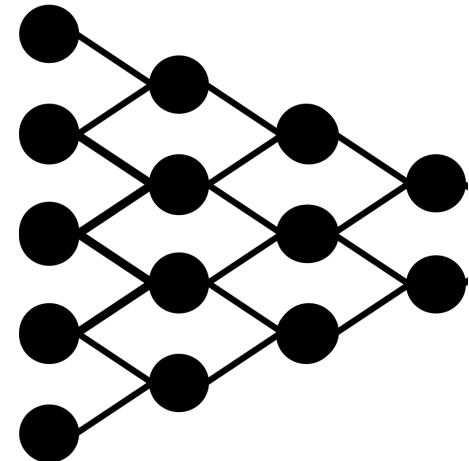
Could we construct representations specialized for control?

What is a Good Visual Representation for Robotics?

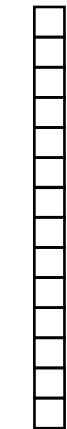
Image \mathbf{o}



$\phi(\cdot)$



$\phi(\mathbf{o})$



$\pi(\cdot)$



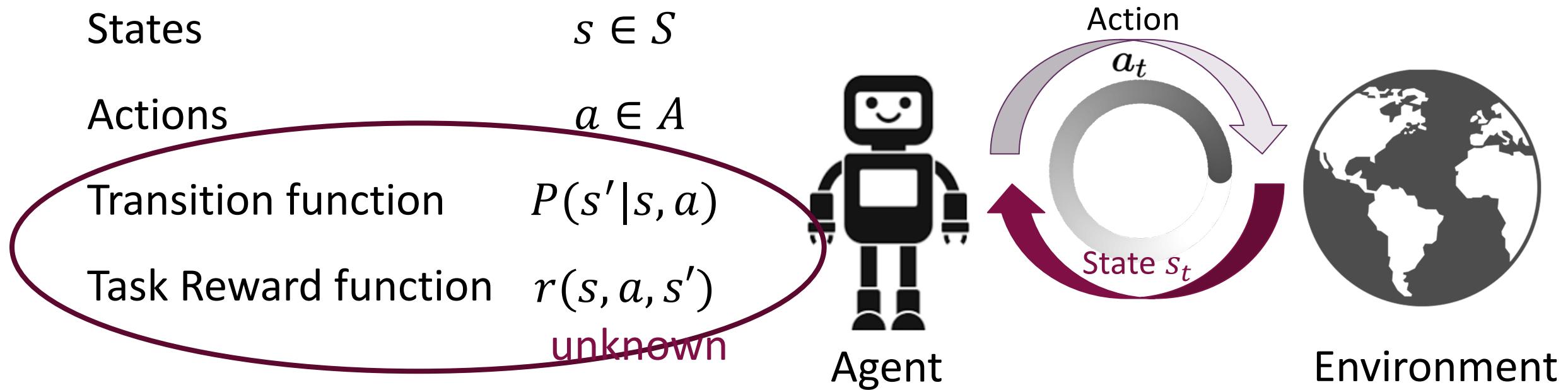
$a = \pi(\phi(\mathbf{o}))$

“rotate
gripper +4°”

Learning Objective: What does it mean to organize the visual information to present to a controller / policy?

Data: What datasets could we train on, that might be useful for robotic manipulation?

Overview: The Reinforcement Learning Formalism



Agent's objective: maximize the discounted sum of “reward” over time by executing a good action sequence a_1, a_2, \dots ,

$$\max_{\pi} R(\pi) = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, s_{t+1}) \right]$$

Universal Value Functions

- Optimal Value Function of A State conditioned on a task g [Schaul et al 2015]

$$V^*(s_0; g) = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, s_{t+1}; g) \right]$$

“How good is this state for completing the task g (if acting optimally)?”

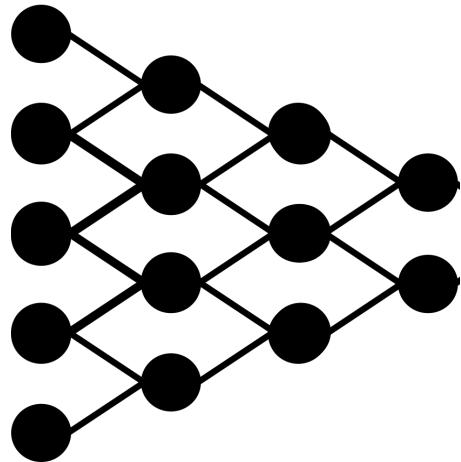
- Value functions are a useful abstraction towards policy learning:
 - Bellman equations and TD-learning
 - Don’t require known actions
 - Can guide policy improvement (trajectory opt, RL, ...)

Representations as Value Functions

Image \mathbf{o}



$\phi(\cdot)$



$\phi(\mathbf{o})$



$\pi(\cdot)$



$a = \pi(\phi(o))$

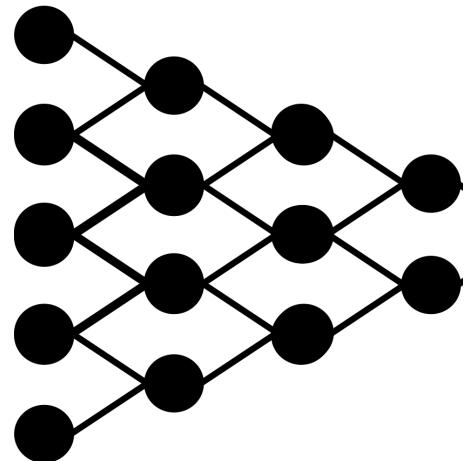
“rotate
gripper +4°”

Representations as Value Functions

Image \mathbf{o}



$\phi(\cdot)$



$\phi(\mathbf{o})$



$V^*(\cdot ; \mathbf{g})$



0.8

task specification \mathbf{g}



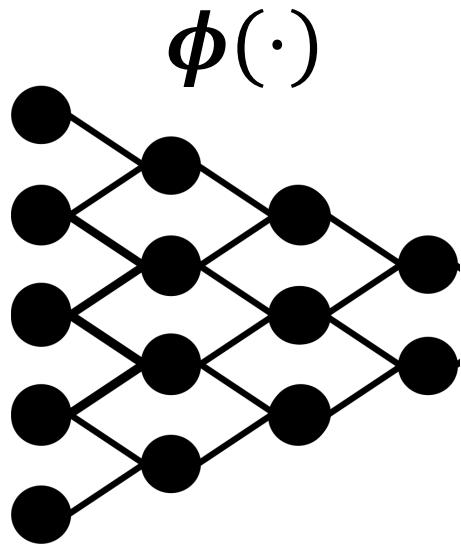
or

“squeeze
the brush
dry”

Representations as Value Functions

VIP, ICLR '23

Image \mathbf{o}



$\phi(\mathbf{o})$



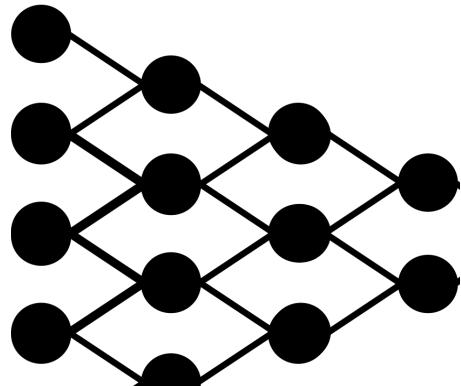
$V^*(\cdot; \mathbf{g})$

$V^*(\mathbf{o}; \mathbf{g})$

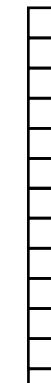
distance

0.8

Image \mathbf{g}



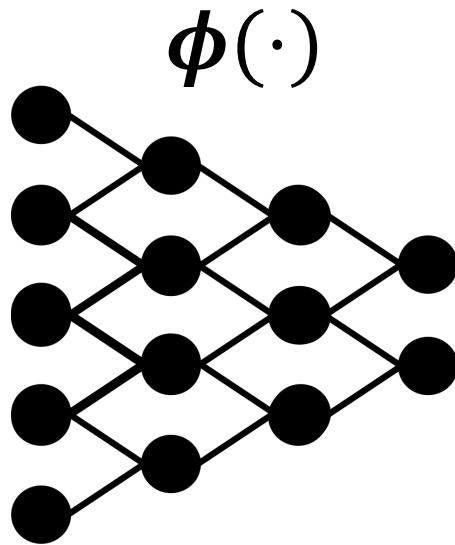
$\phi(\mathbf{g})$



Representation $\phi(\cdot)$ should be rich enough so that it easily expresses V^*

Representations as Value Functions

Image \mathbf{o}



$\phi(\mathbf{o})$



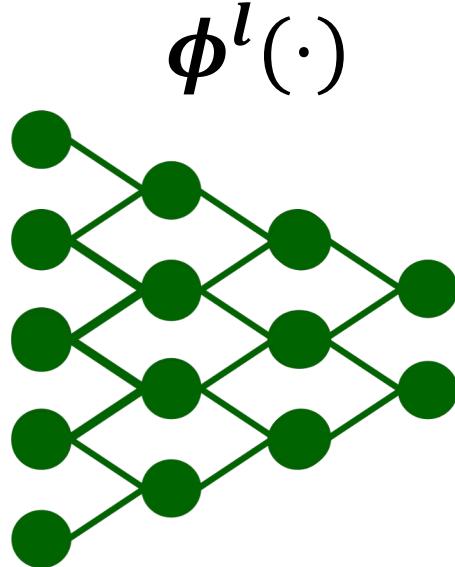
LIV (under review)

$V^*(\cdot; \mathbf{g})$

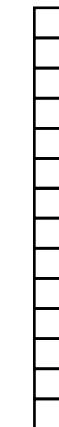
$V^*(\mathbf{o}; \mathbf{g})$

Language \mathbf{g}

“squeeze
the brush
dry”



$\phi^l(\mathbf{g})$



distance

0.8

Core Idea:
Train $\phi(\cdot), \phi^l(\cdot)$ to minimize
TD error in $V^*(\phi(\cdot), \phi^l(\cdot))^*$

Data To Train a Universal Value Function

Data: What datasets to train on?

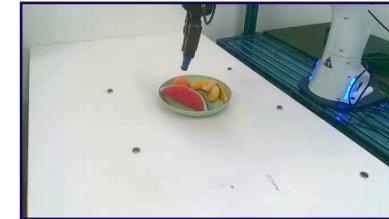
V^* (

Observation



,

Goal

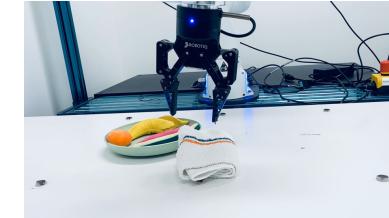


)

V^* (



,



)



In-domain, task-specific robot demonstration data are inherently scarce and expensive to collect.

Not enough robot data for pre-training and generalization

Pre-Train on In-the-Wild Human Videos

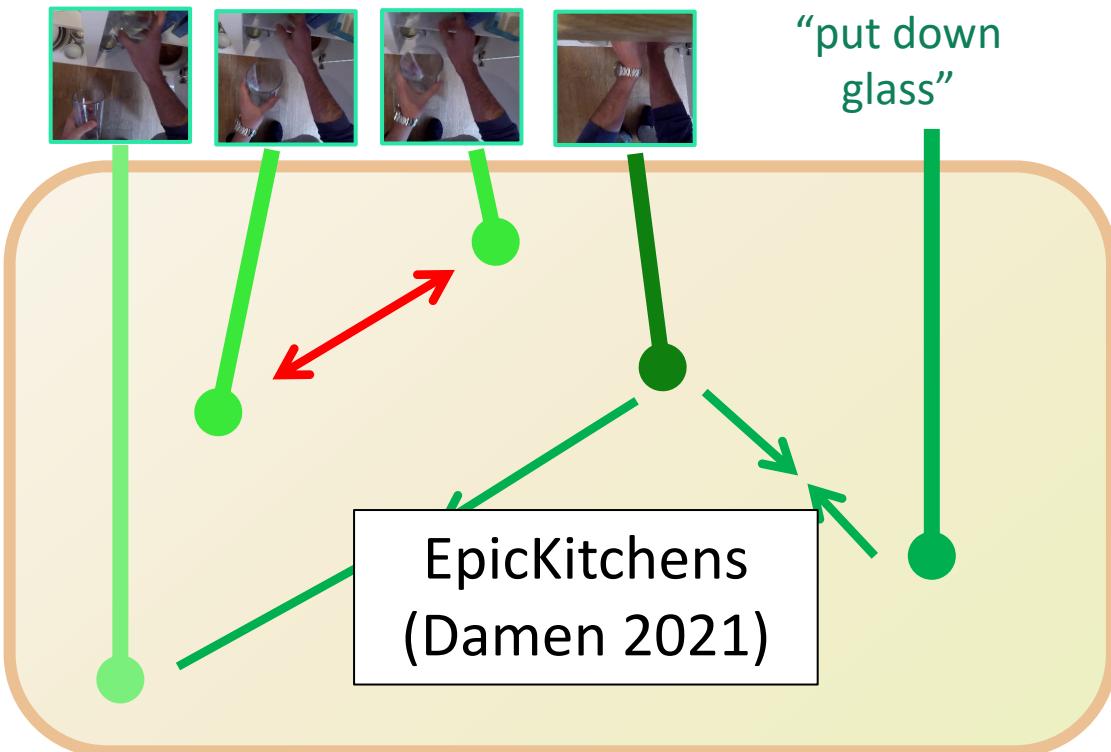
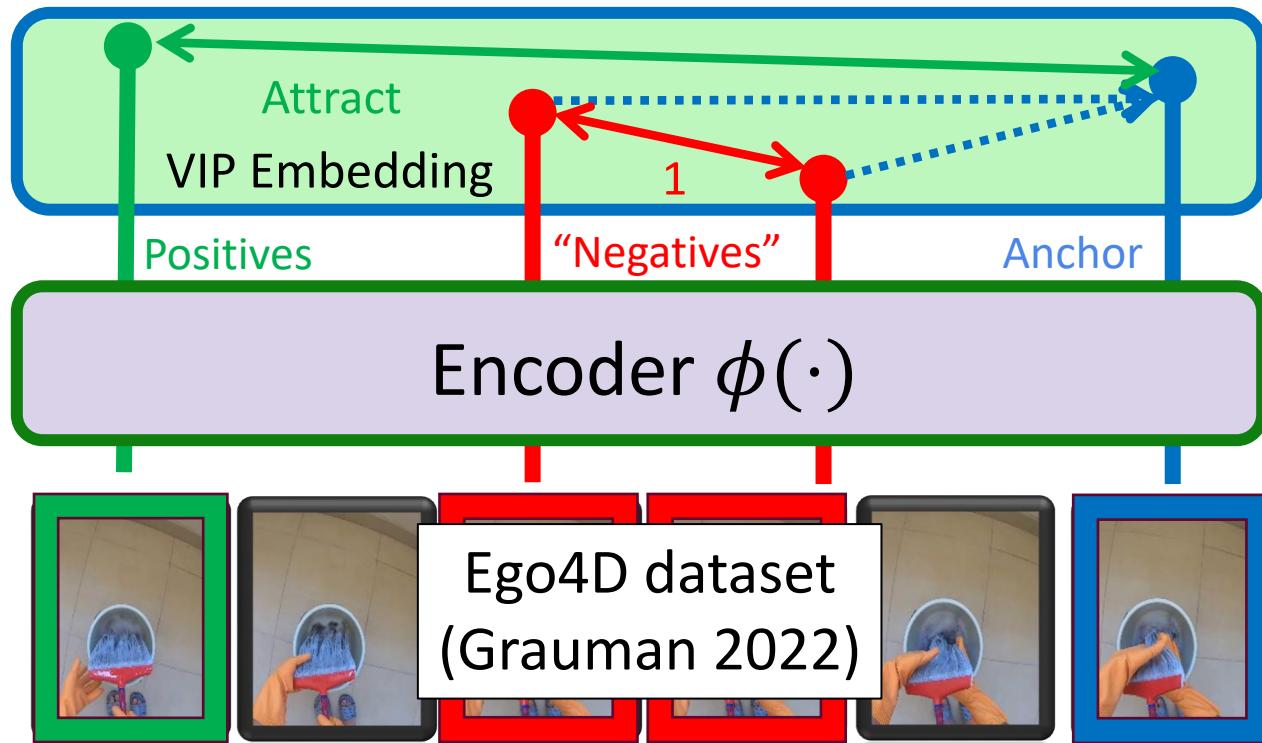
Human videos are abundant, and cover many diverse tasks!



- **Advantage of goal-reaching rewards:** every video reaches *some* goal! Just treat the final frame* of any video as the goal
- Reward function? $r = 1$ for last step of video, $\epsilon < 1$ elsewhere.
- Actions not available, but no problem: we only care for $V^*(s)$

Training Objective

$$L_{\text{InfoNCE}} = -\mathbb{E} \left[\log \frac{s(x, x_{\text{pos}})}{s(x, x_{\text{pos}}) + \sum_{y_j \neq x_{\text{pos}}} s(x, y_j)} \right]$$



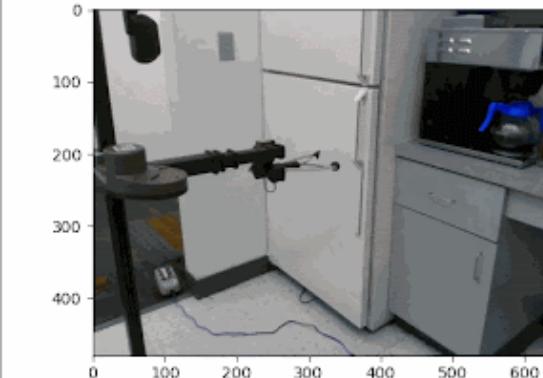
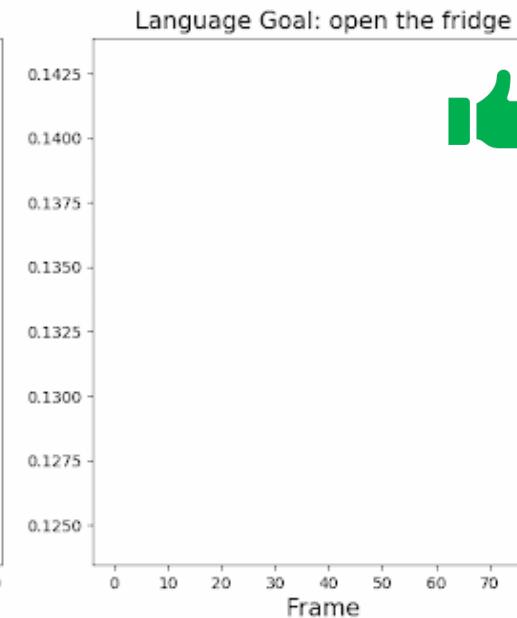
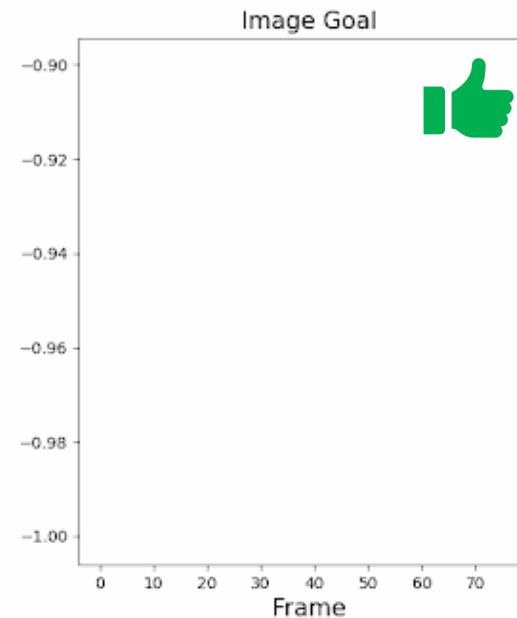
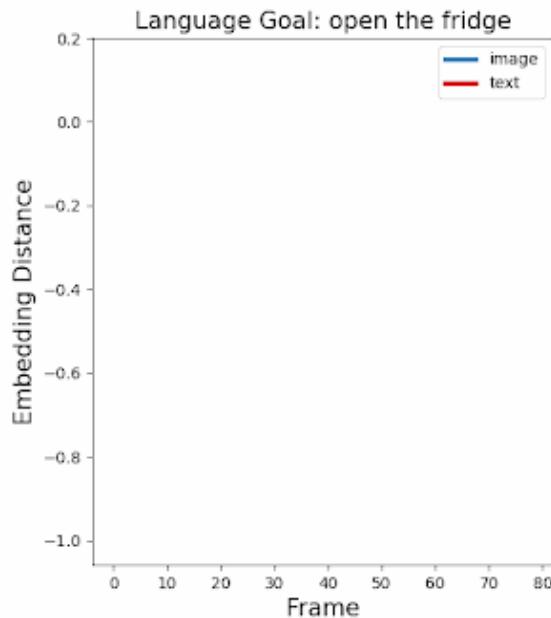
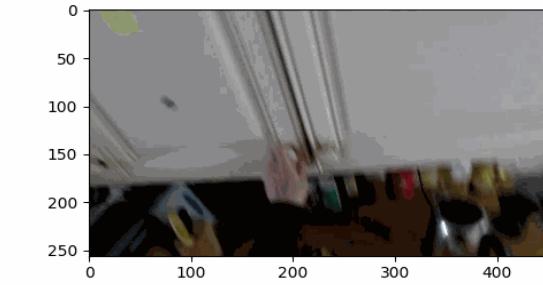
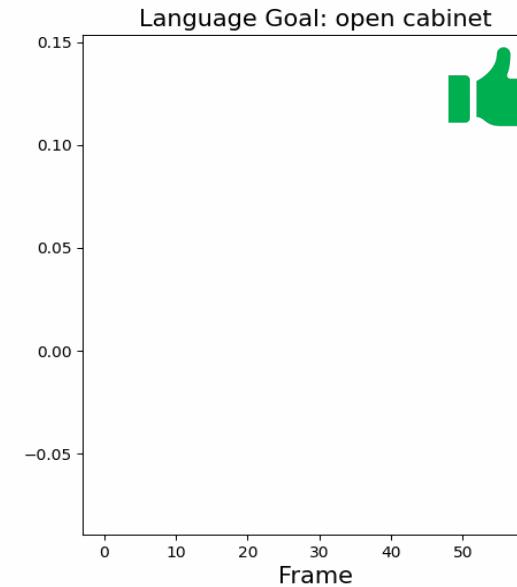
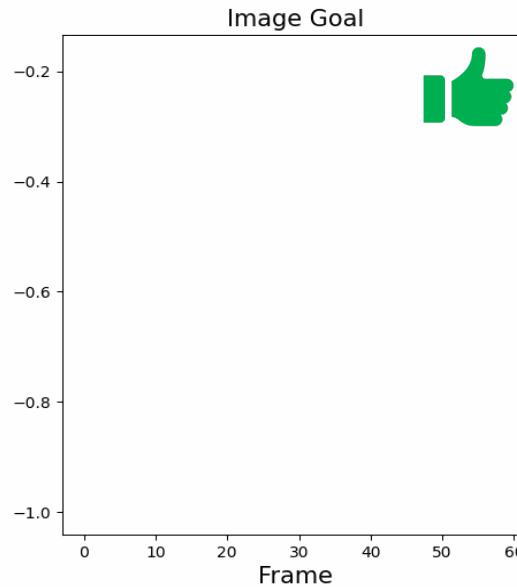
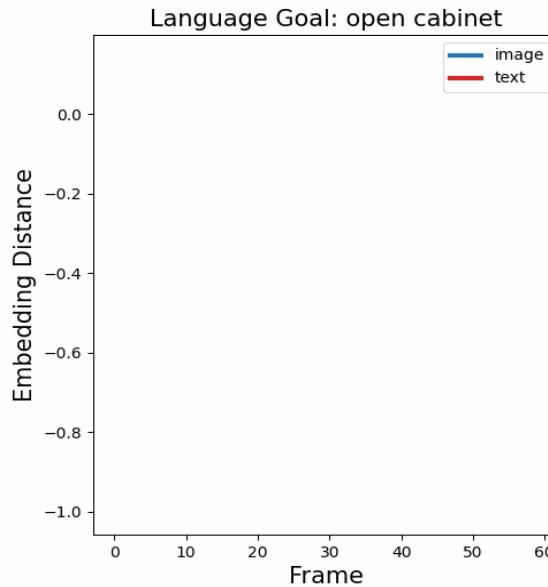
$$\mathbb{E}_{p(g)} \left[(1 - \gamma) \mathbb{E}_{\mu_0(o;g)} [\|\phi(o) - \phi(g)\|_2] + \log \mathbb{E}_{(o,o';g) \sim D} \left[\exp \left(\|\phi(o) - \phi(g)\|_2 - \tilde{\delta}_g(o) - \gamma \|\phi(o') - \phi(g)\|_2 \right) \right] \right]$$

Pulls $\phi(o)$ towards $\phi(g)$

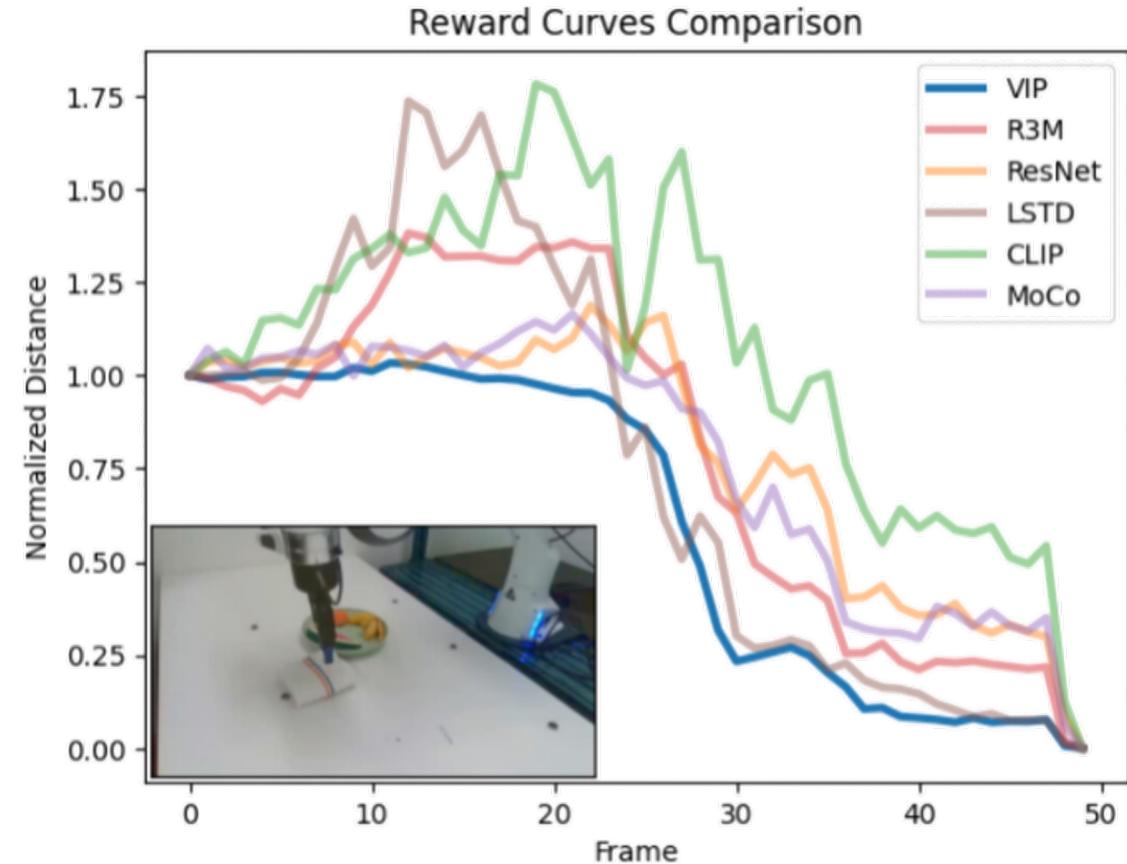
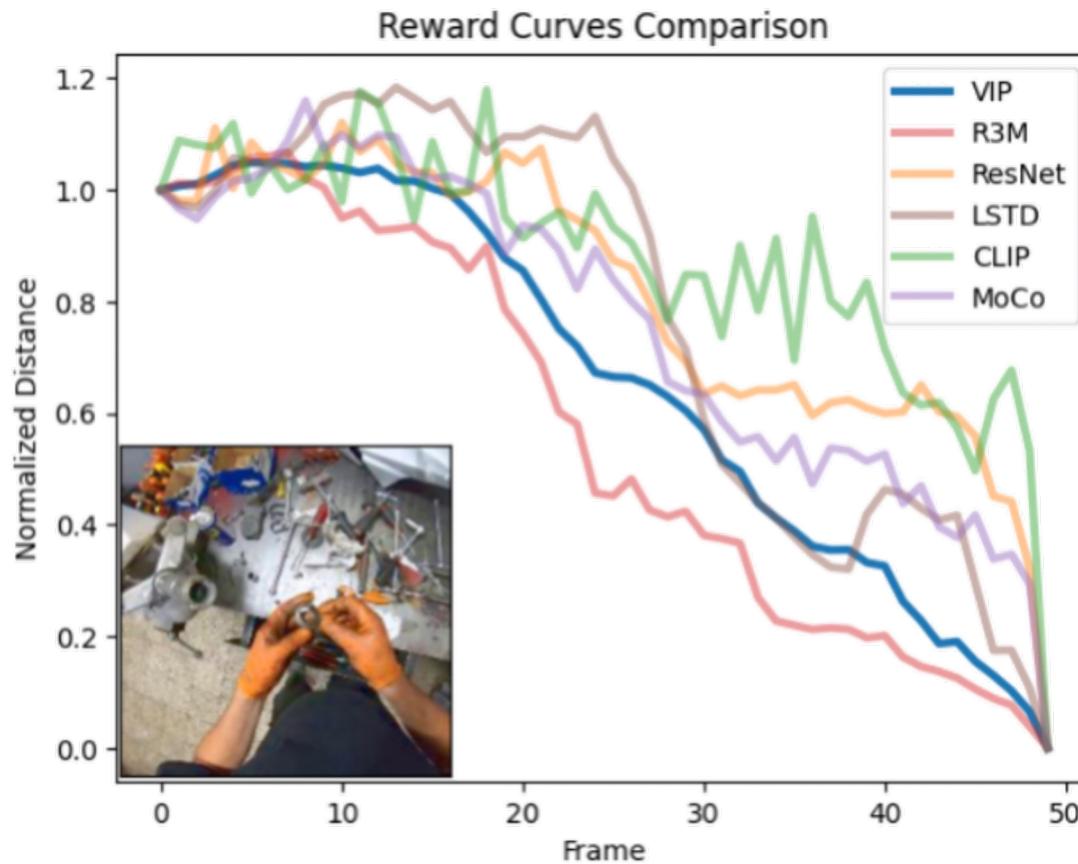
preceding $V^*(o ; g)$

$$\min_{\phi} (1 - \gamma) \mathbb{E}_{p(g), \mu_0(o;g)} \left[-\log \frac{e^{V^*(\phi(o); \phi(g))}}{\mathbb{E}_{D(o, o'; g)} [\exp(\tilde{\delta}_g(o) + \gamma V^*(\phi(o'); \phi(g)) - V^*(\phi(o), \phi(g)))]^{\frac{-1}{(1-\gamma)}}} \right] \|\phi(\cdot) - \phi(g)\|_2$$

Results: Language-Goal Value Function $d(\phi(o), \phi^l(g))$



Results: Image-Goal Value Function $d(\phi(o), \phi(g))$



On demo data, our representations predict smooth goal-conditioned V^* on human and robot videos.

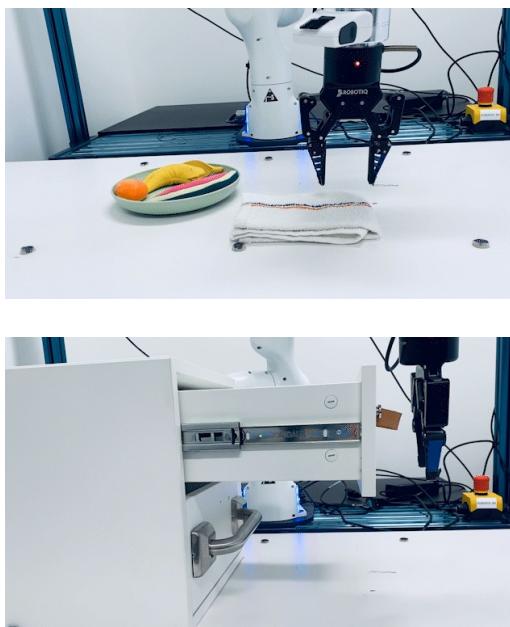
What Can We Do With $\phi(\cdot)$ and $\phi^l(\cdot)$?

- **Use as representations for robot learning:**
 - Training robot policies on image representation with:
 - behavior cloning
 - language-conditioned behavior cloning [Lynch '20]
- **Use as dense reward functions to guide reinforcement policy learning:**
 - $R(o, a, o'; g) = V^*(o', g) - V^*(o, g) = \|\phi(o') - \phi(g)\|_2 - \|\phi(o) - \phi(g)\|_2$
 - offline RL (reward-weighted regression [Peters '07]) for policy learning from noisy demos
 - online policy improvement with trajectory optimization and RL (natural policy gradient [Kakade '01])

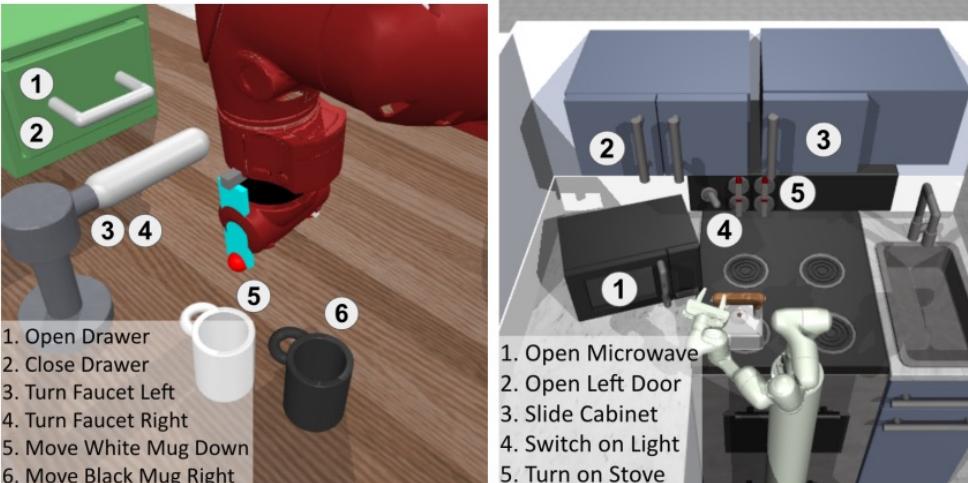
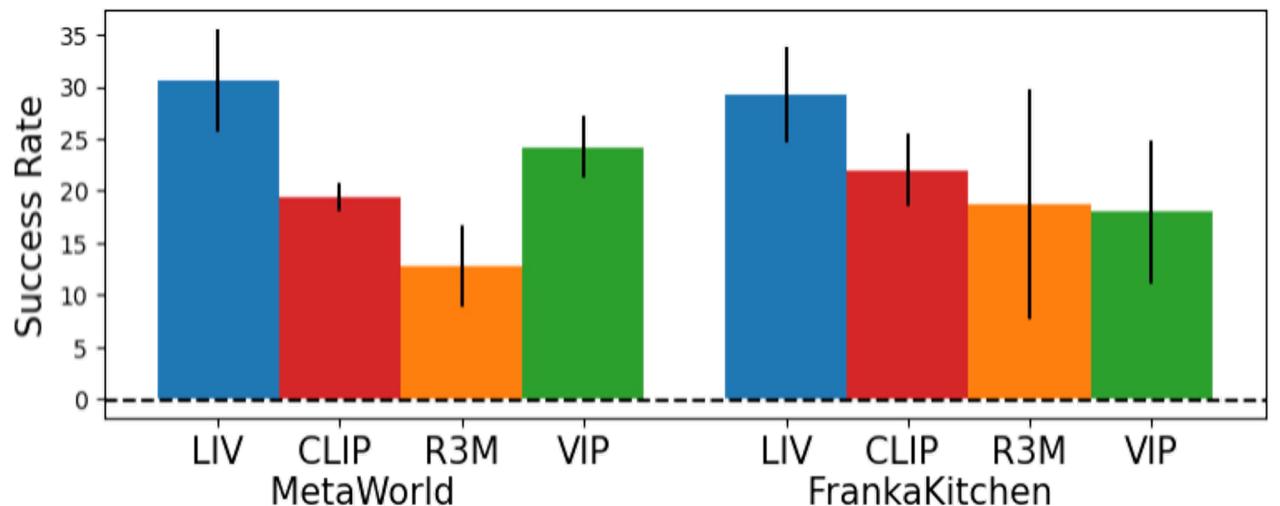
Quantitative Results Summary

Results: Real-World BC / Offline RL From 20 Demos

Environment	Pre-Trained				In-Domain		
	VIP-RWR	VIP-BC	R3M-RWR	R3M-BC	Scratch-BC	VIP-RWR	VIP-BC
CloseDrawer	100 ± 0	50 ± 50	80 ± 40	10 ± 30	30 ± 46	0 ± 0	0* ± 0
PushBottle	90 ± 30	50 ± 50	70 ± 46	50 ± 50	40 ± 48	0* ± 0	0* ± 0
PlaceMelon	60 ± 48	10 ± 30	0 ± 0	0 ± 0	0 ± 0	0* ± 0	0* ± 0
FoldTowel	90 ± 30	20 ± 40	0 ± 0	0 ± 0	0 ± 0	0* ± 0	0* ± 0

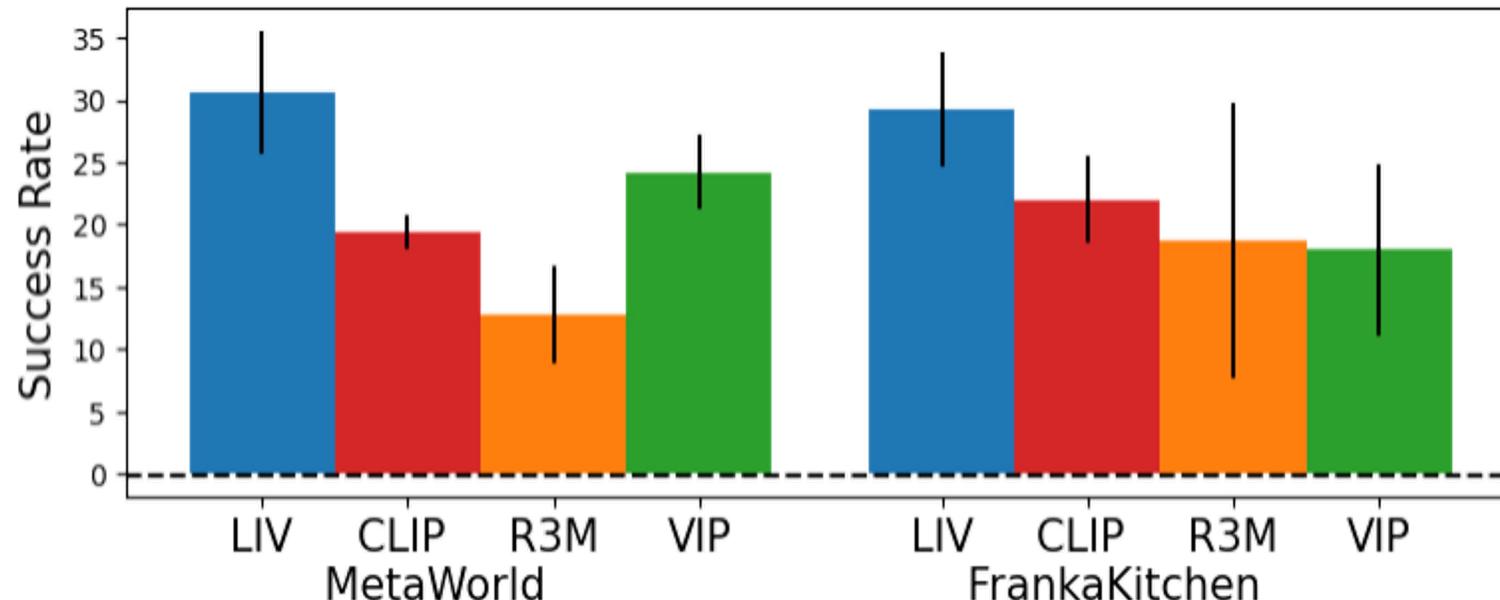


Results: Language-Conditioned Behavior Cloning

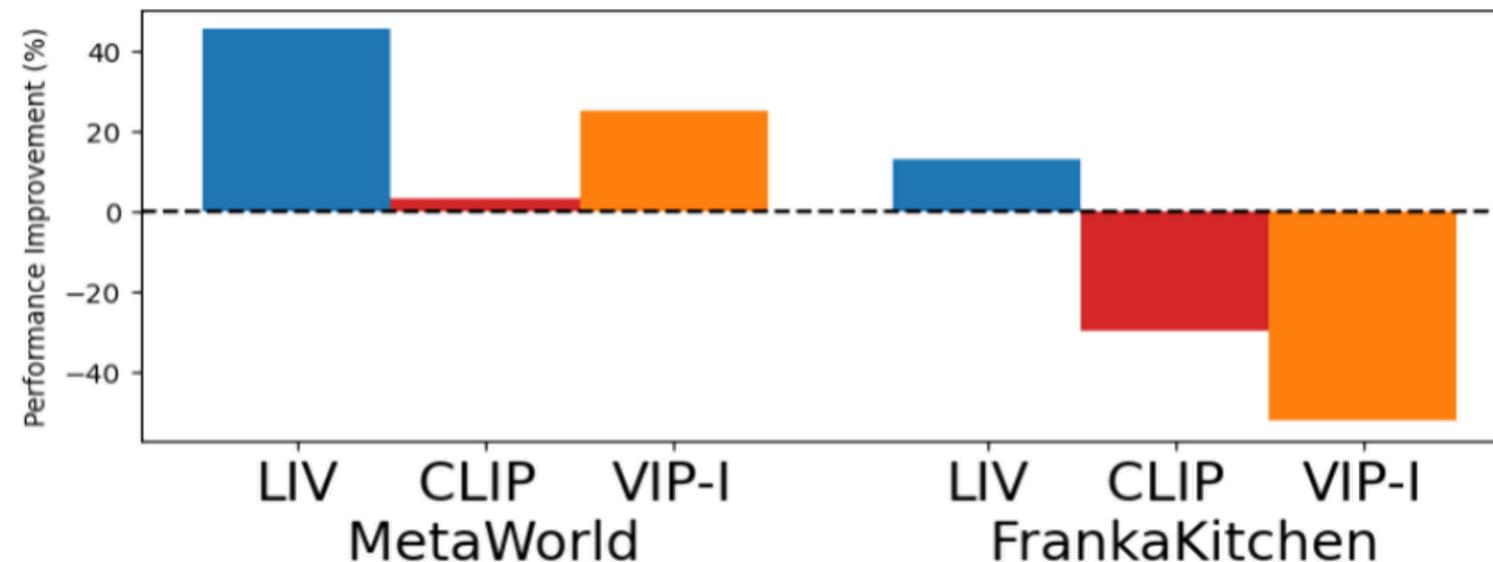


Results: Language-Conditioned Behavior Cloning

Pretraining-only performance

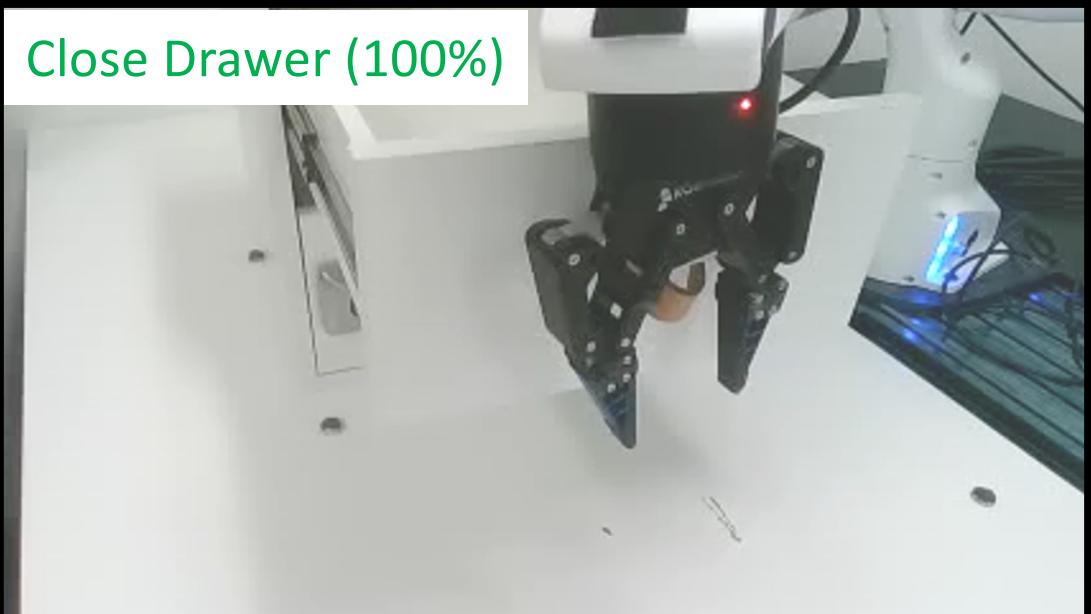


Performance improvement from in-domain finetuning

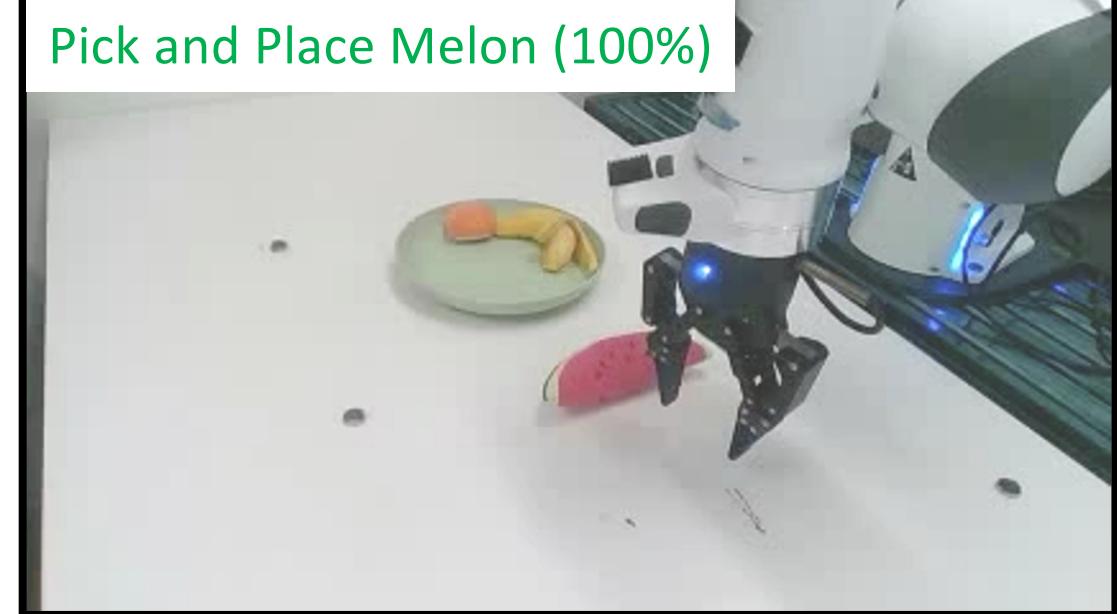


Results: Offline RL Examples

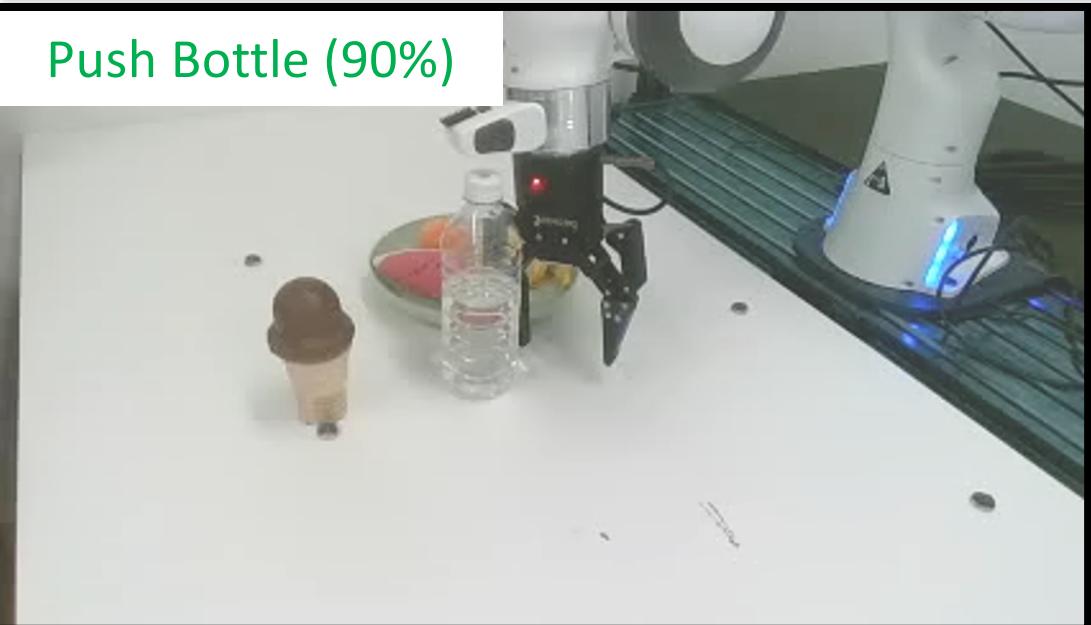
Close Drawer (100%)



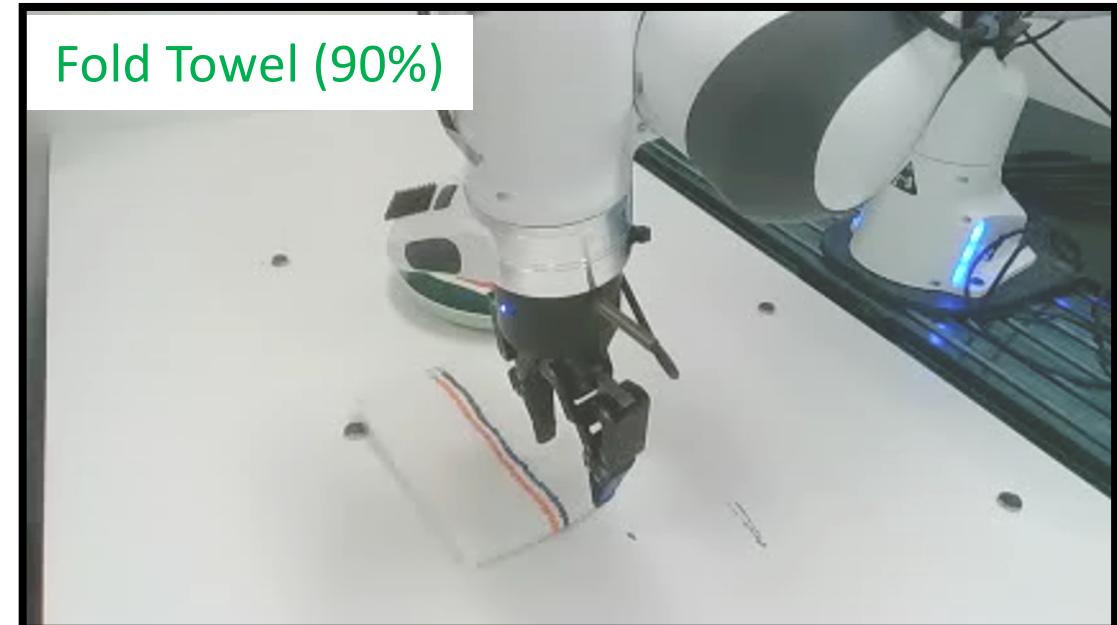
Pick and Place Melon (100%)



Push Bottle (90%)



Fold Towel (90%)



Results: Image Goal-Conditioned Trajectory Opt. & Online RL

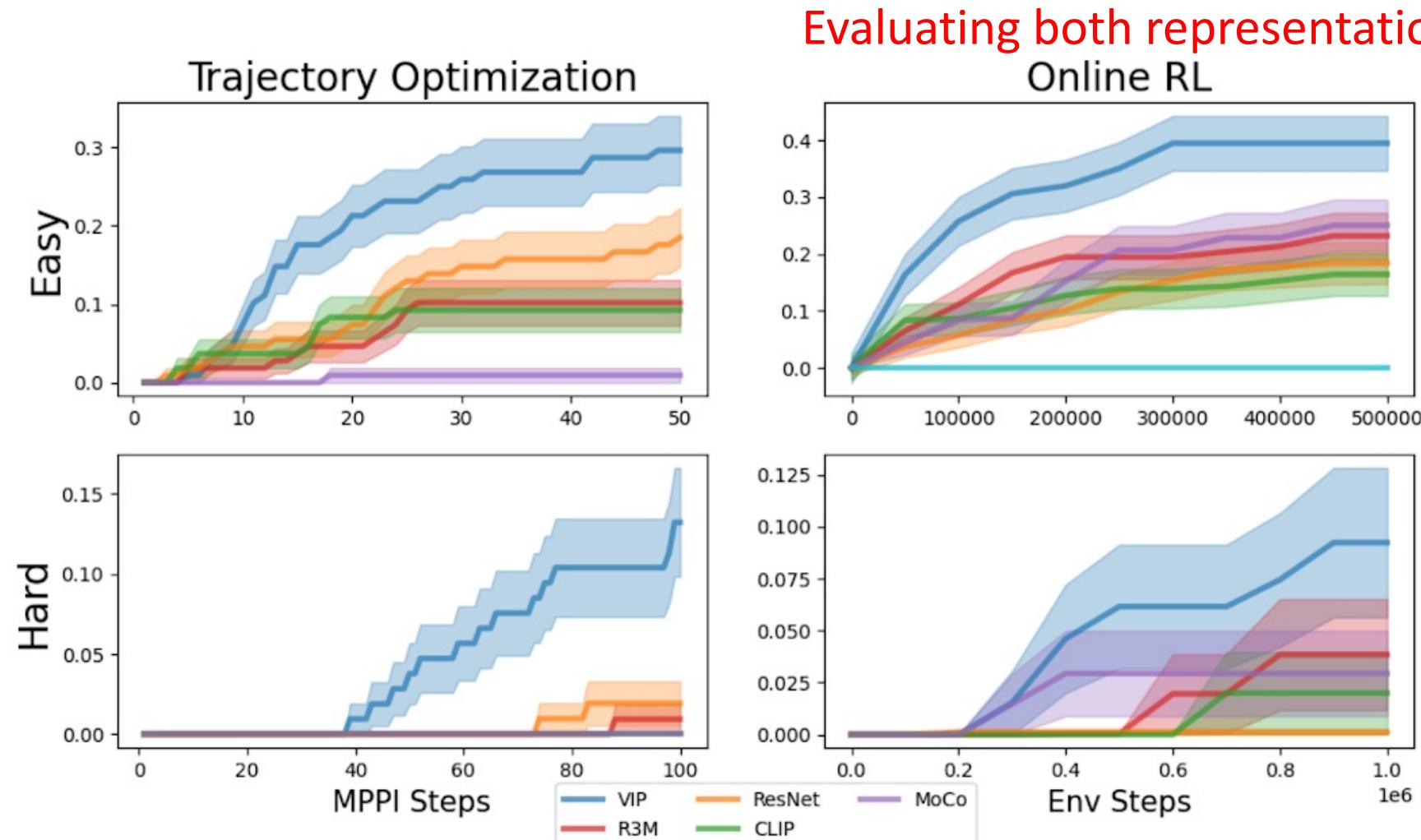
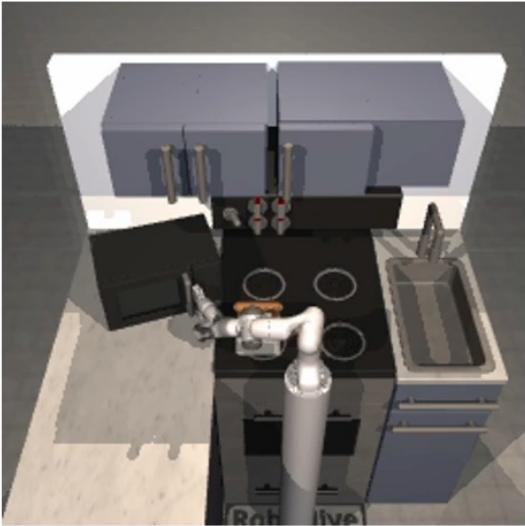
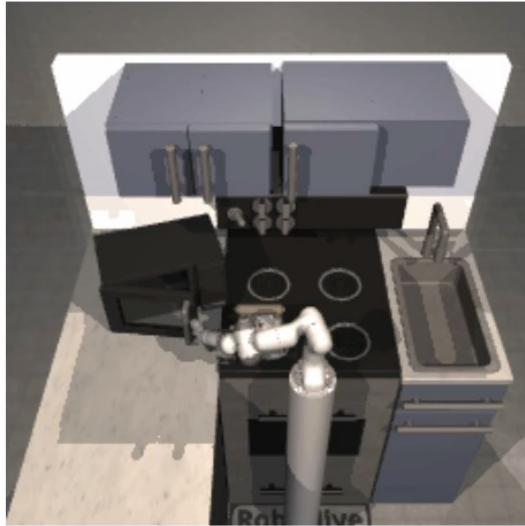


Figure 4: Visual trajectory optimization and online RL aggregate results (cumulative success rate %).

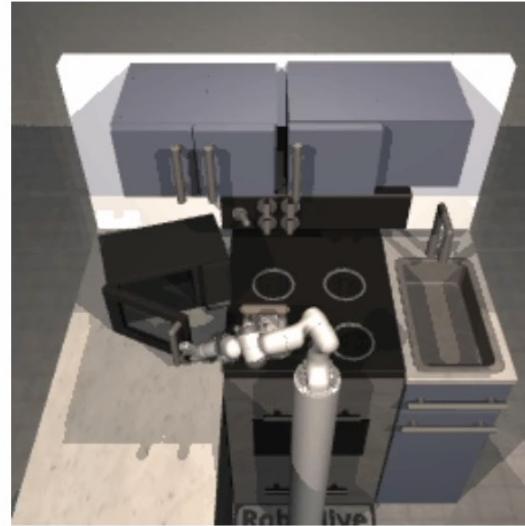
Results: Image Goal Trajectory Opt Examples



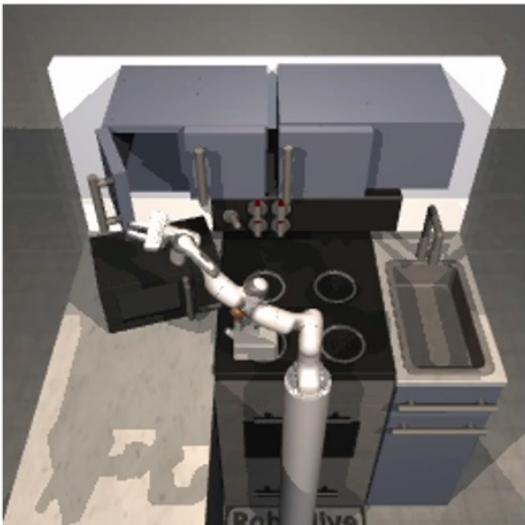
Microwave-Close Goal Image (center view)



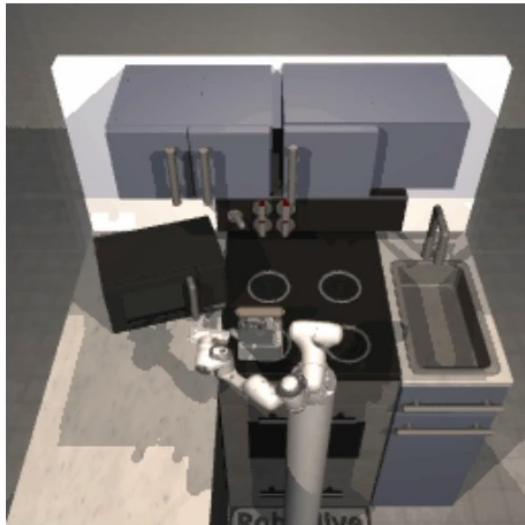
VIP



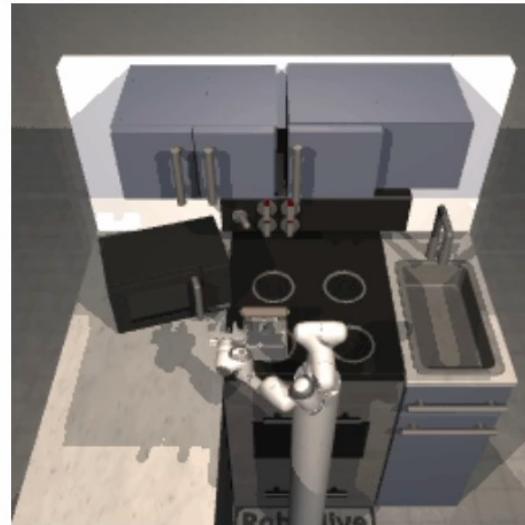
R3M



Leftdoor-open Goal Image (center view)



VIP



R3M

Takeaway

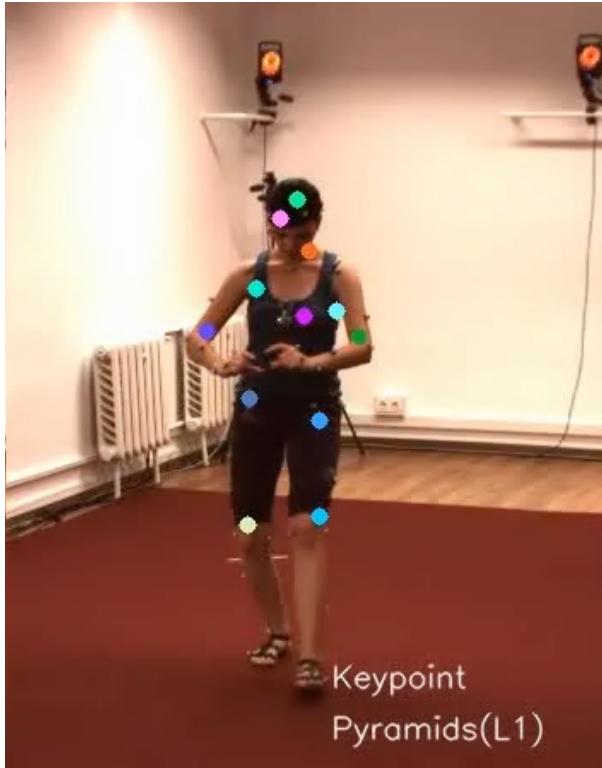
Representations as goal-conditioned “universal value functions” offer a powerful new way to learn *control-aware* vision, language, (and other?) representations.

Object-Structured Visual Representations

E.g. Unsupervised, hierarchically structured entity-centric representations.

“Keypoint Pyramids” (ECCV 2022)

H3.6M

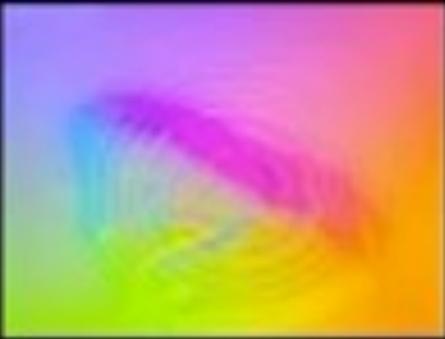


UPenn B&O

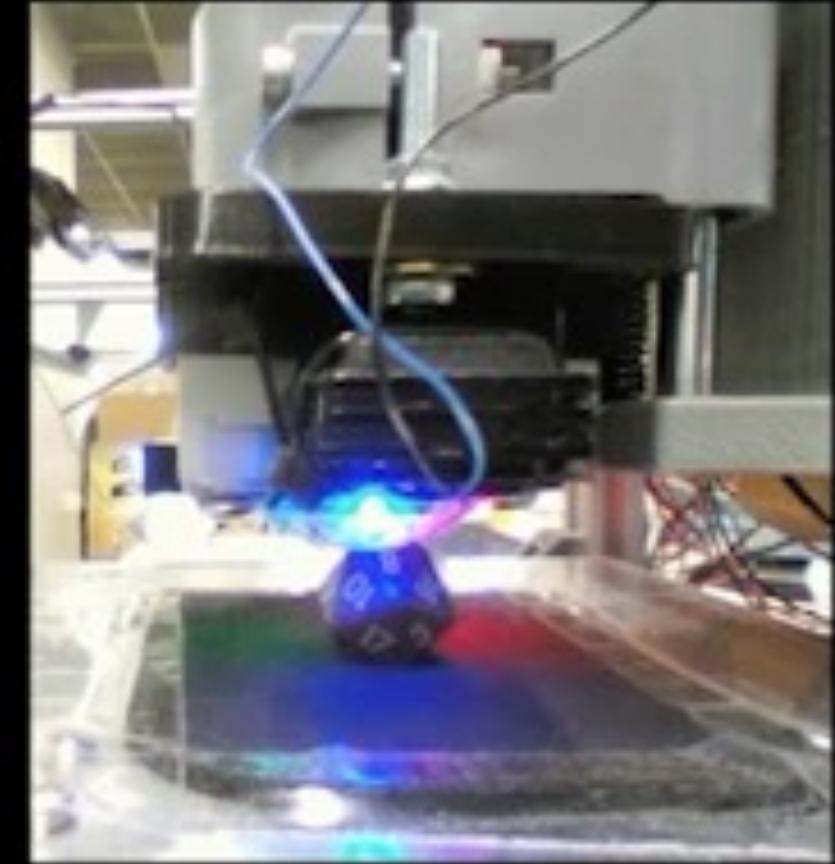


Goals In Other Modalities: Tactile Servoing

Goal Image



Actual rollout



Takeaways

- Control-specific multimodal representations can be trained *as value functions* from large-scale offline data
 - Physical objects can be used to specify task goals by training interactive reward function policies.
 - Goal-directed exploration through learned models can discover skills.
-
- **Future work:**
 - Shared representations, encoding objects etc., to improve the task specification interface.
 - Logical task specifications, safety constraints ...
 - Learners that can flexibly recognize and exploit many different types of learning signals on-the-fly.

Acknowledgements



National
Science
Foundation



NEC Laboratories America
Relentless passion for innovation



GE Research