Generalization through Generation:

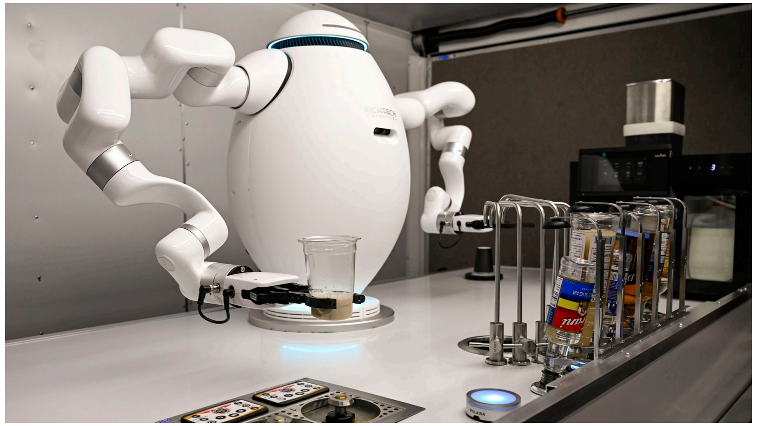
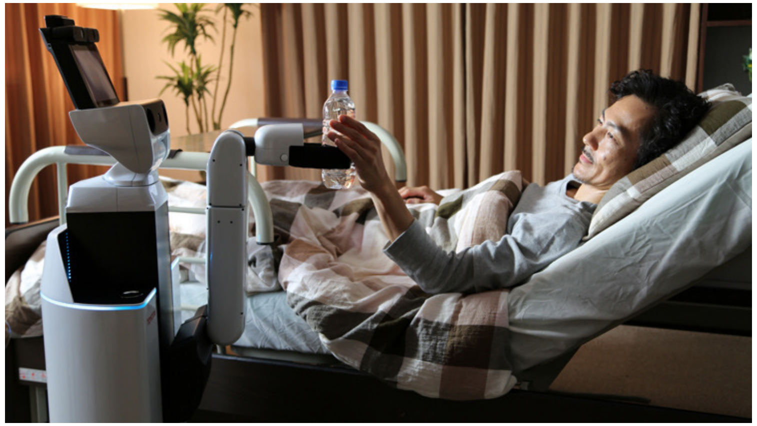
Learning Long-Horizon Tasks with Limited Supervision

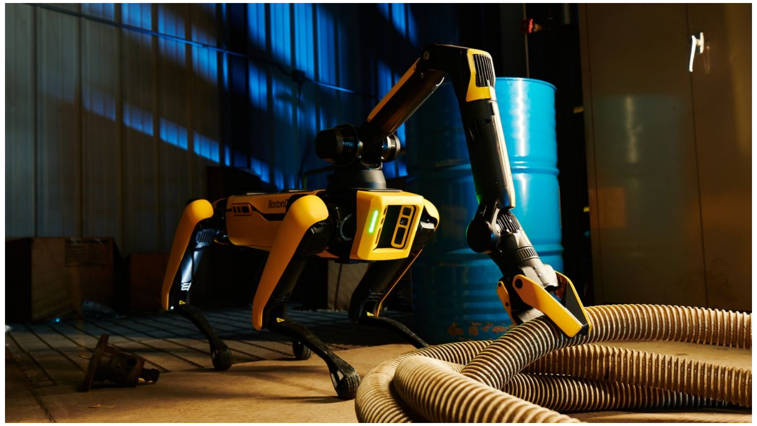
Kuan Fang

Berkeley UNIVERSITY OF CALIFORNIA

Cornell University

Long-horizon tasks in unstructured environments




Solving long-horizon tasks with general-purpose skills

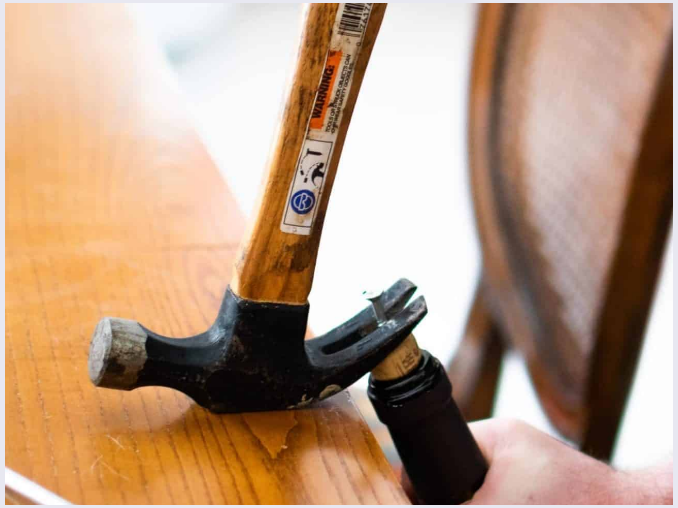
robust

Robustly handle wide variations of the task

adaptable

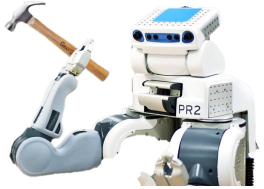
Efficiently adapt to unseen and more challenging tasks

extensible



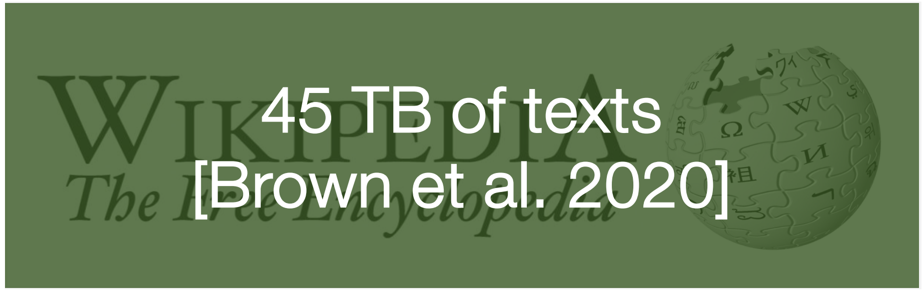
Discover novel behaviors for ever-increasing demand

Can robots acquire general-purpose skills through learning?

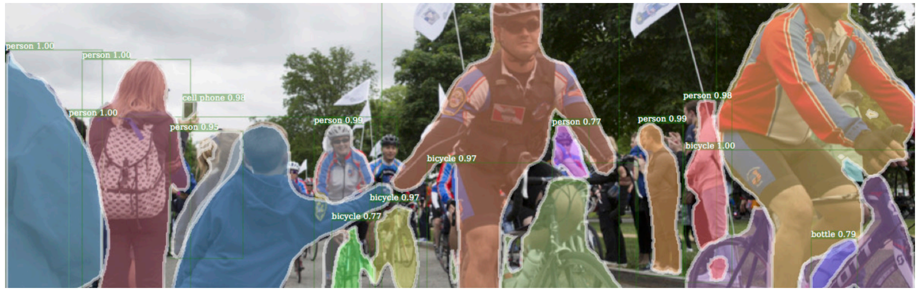


The power of scaling up

massive datasets ↑



generalizable models ↑



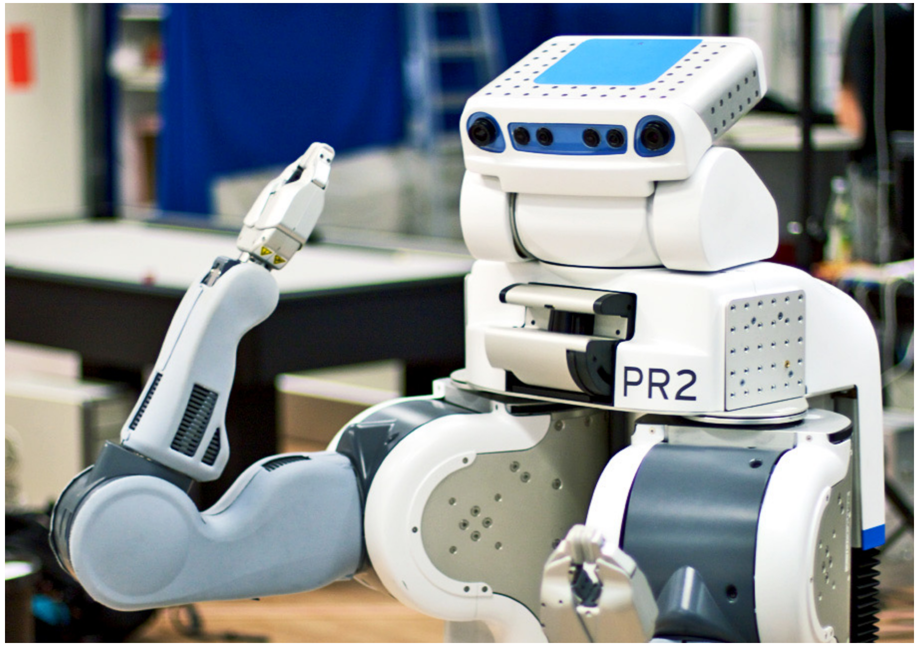
ChatGPT

[Lin et al. 2014; Liu et al. 2015; Yu et al. 2015; Chang et al. 2015; Heilbron et al. 2015; Abu-El-Haija et al. 2016; Mo et al. 2018; He et al. 2017]

Towards scaling up robot learning

wide-ranging tasks ?

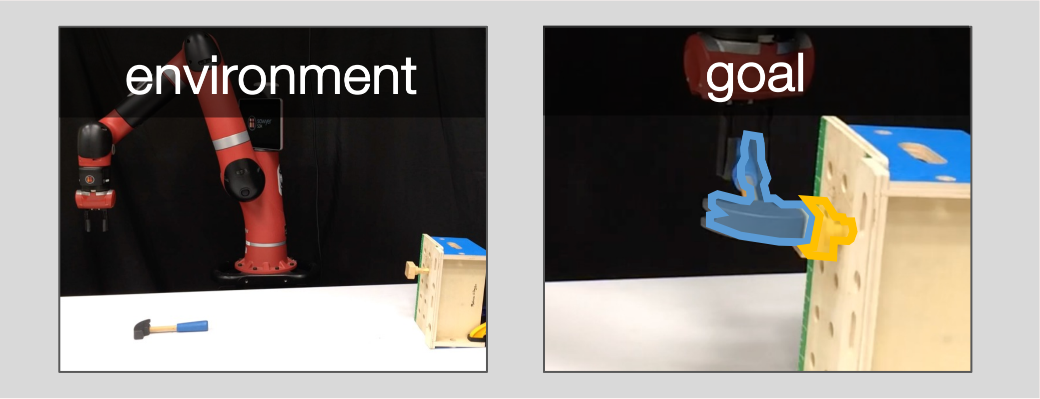
general-purpose skills ?



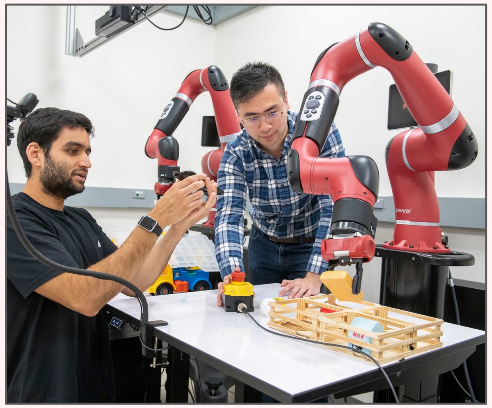
Human supervision in robot learning

understand tasks

specify tasks



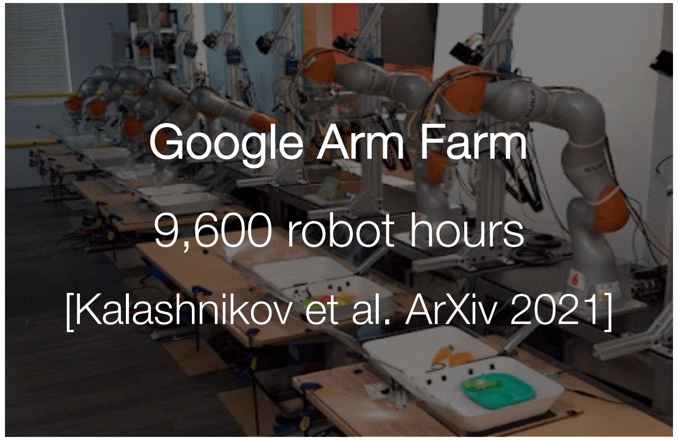
learn tasks

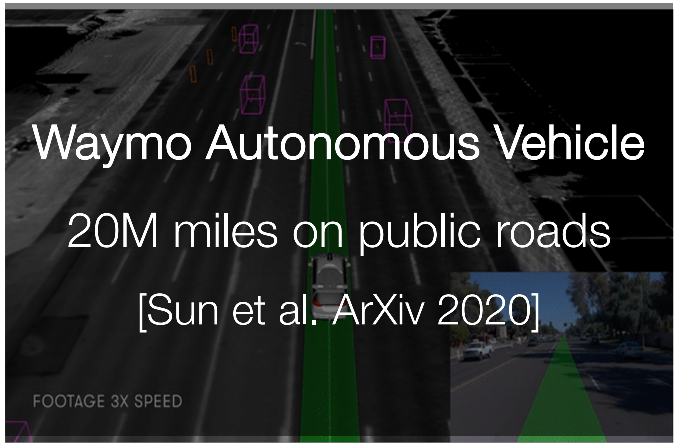


Require non-trivial **manual labor** and **domain knowledge**

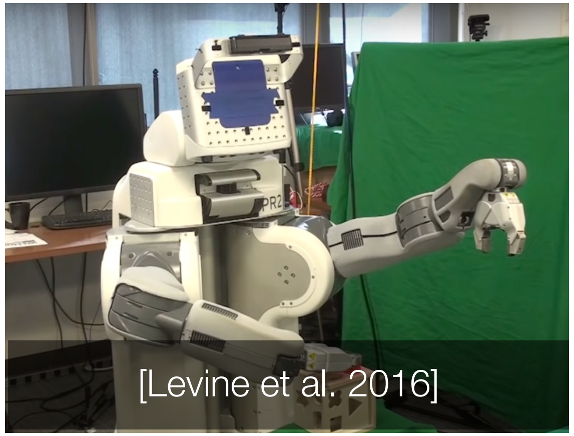
Specific for each new environment, goal, and task.

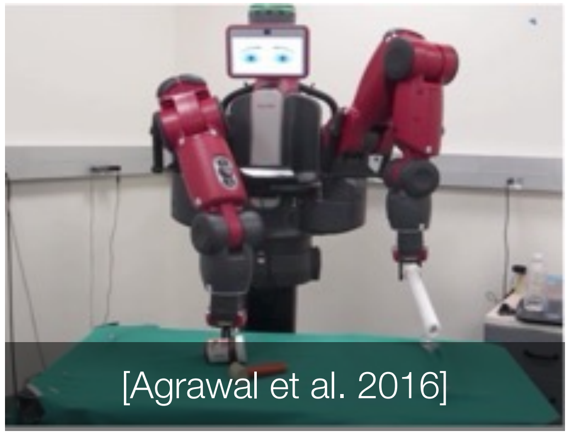
Towards scaling up human supervision





Prior efforts in industry





Prior efforts in research labs

[Byravan et al. 2017; Levine et al. 2017; Gordon et al. 2018; Xiang et al. 2018; Manuelli et al. 2019; Pinto and Gupta 2016; Punjani and Abbeel 2015; Kalashnikov et al. 2018; Srinivas et al. 2018; Peng et al. 2020; Zeng et al. 2017] ……

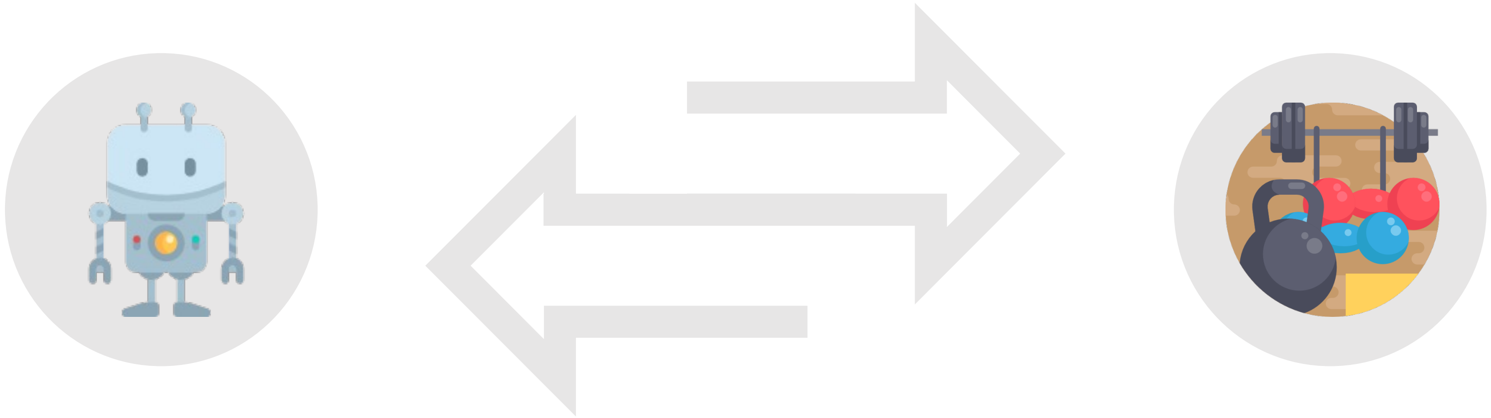

How can robots effectively learn to solve long-horizon tasks with limited human supervision?

Generalization through Generation

Core idea: train robots to autonomously acquire **general-purpose skills** through the **generation** of environments and goals.

Generalization through Generation

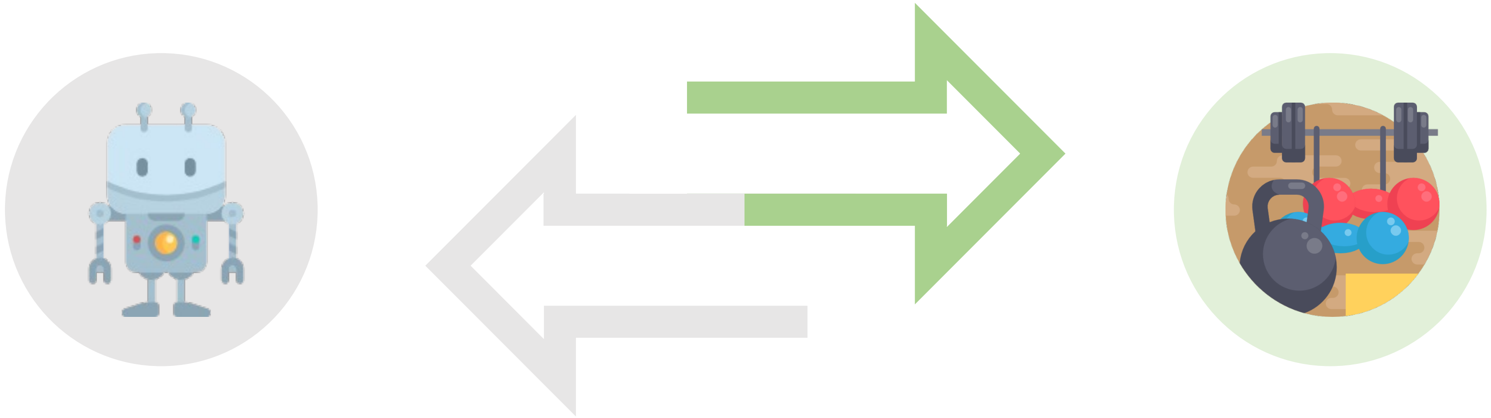
Core idea: train robots to autonomously acquire **general-purpose skills** through the **generation** of environments and goals.



Generalization through Generation

Core idea: train robots to autonomously acquire **general-purpose skills** through the **generation** of environments and goals.

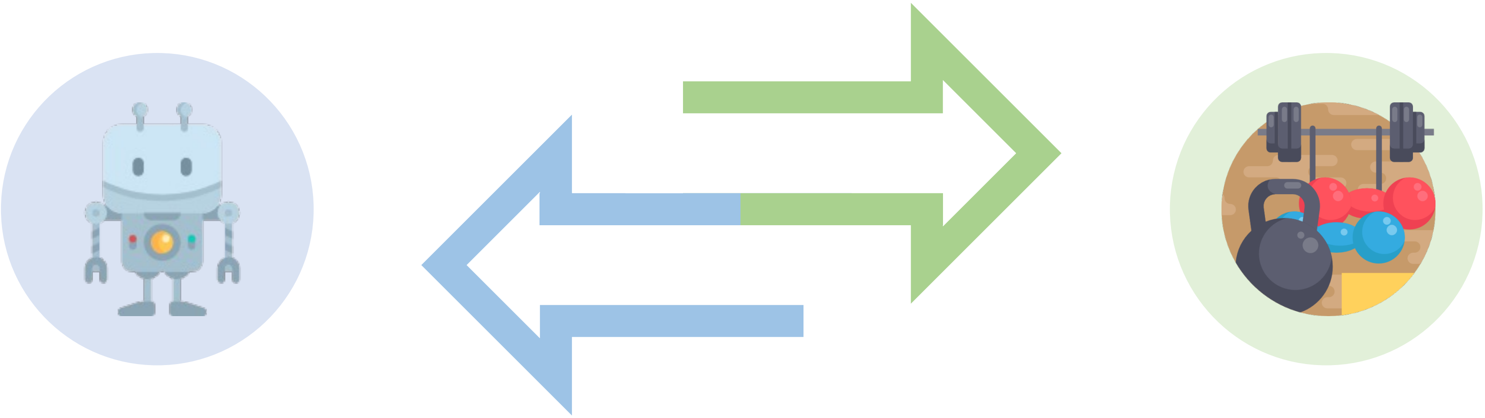
Generate feasible, diverse, and useful tasks



Generalization through Generation

Core idea: train robots to autonomously acquire **general-purpose skills** through the **generation** of environments and goals.

Generate feasible, diverse, and useful tasks

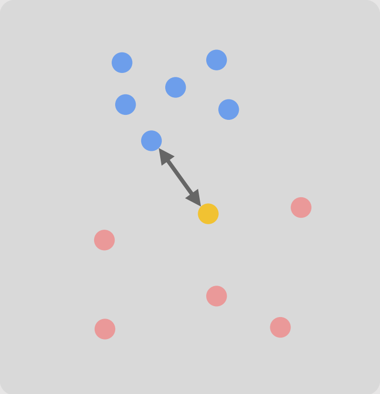


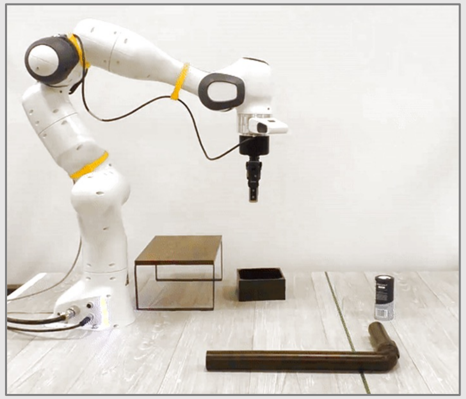
Acquire robust, adaptable, and extensible skills

Generalization through Generation:

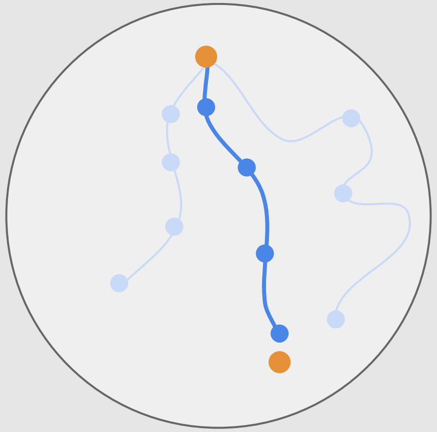
Learning Long-Horizon Tasks with Limited Supervision

Learning Robust Skill via Environment Generation





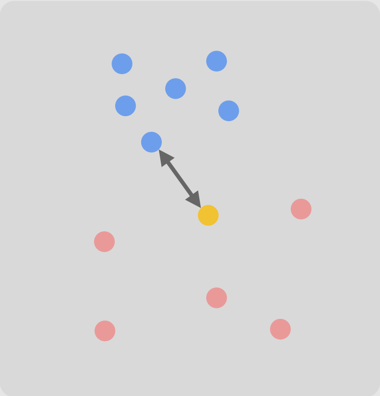
Adapting Prior Skills via Goal Generation

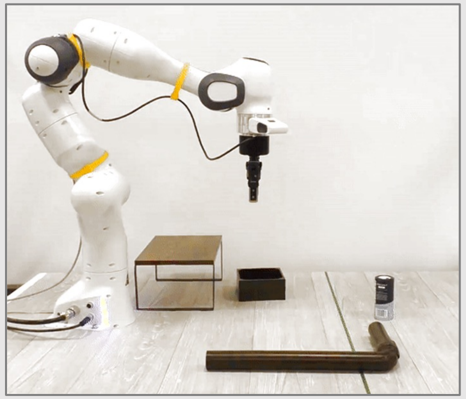


Generalization through Generation:

Learning Long-Horizon Tasks with Limited Supervision

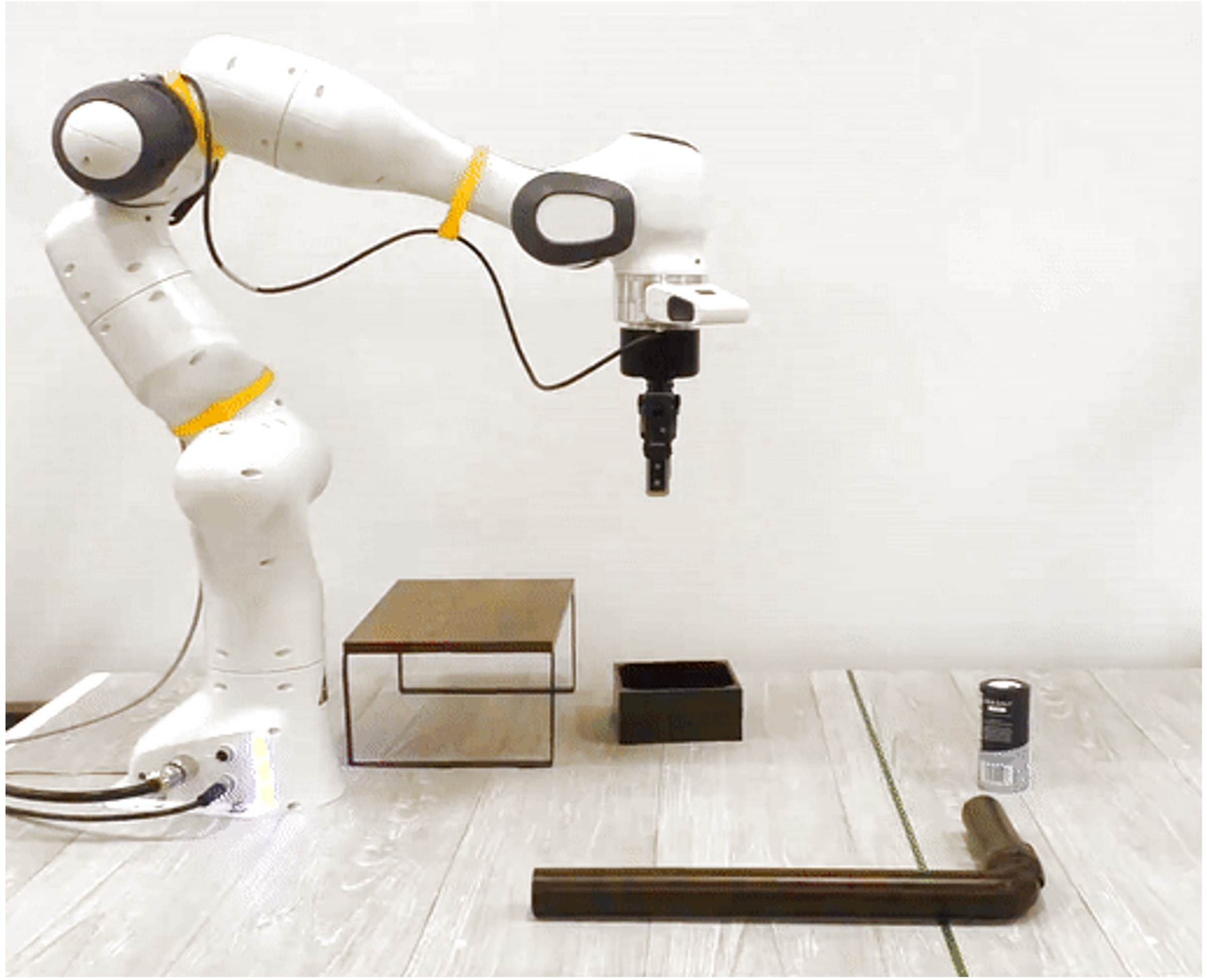
Learning Robust Skill via Environment Generation



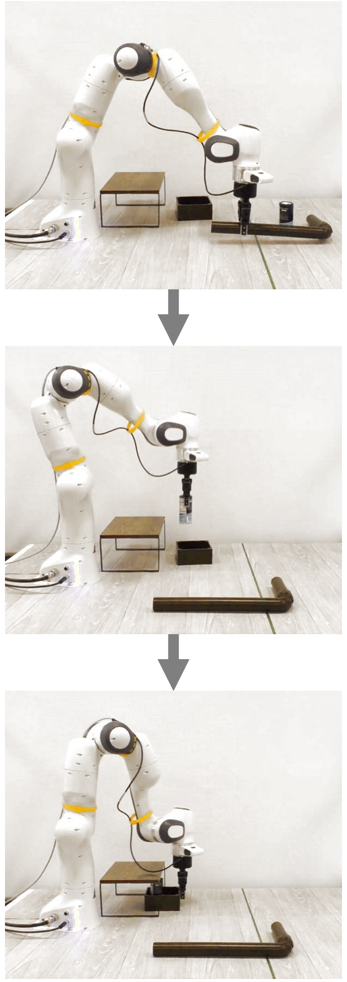


Adapting Prior Skills via Goal Generation

Robust skills for robotic manipulation

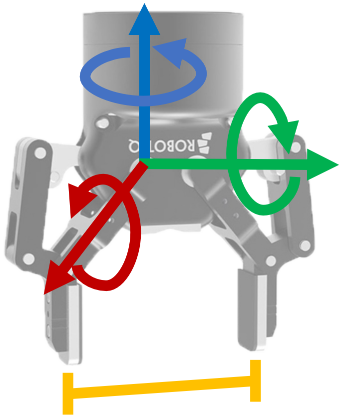


skills

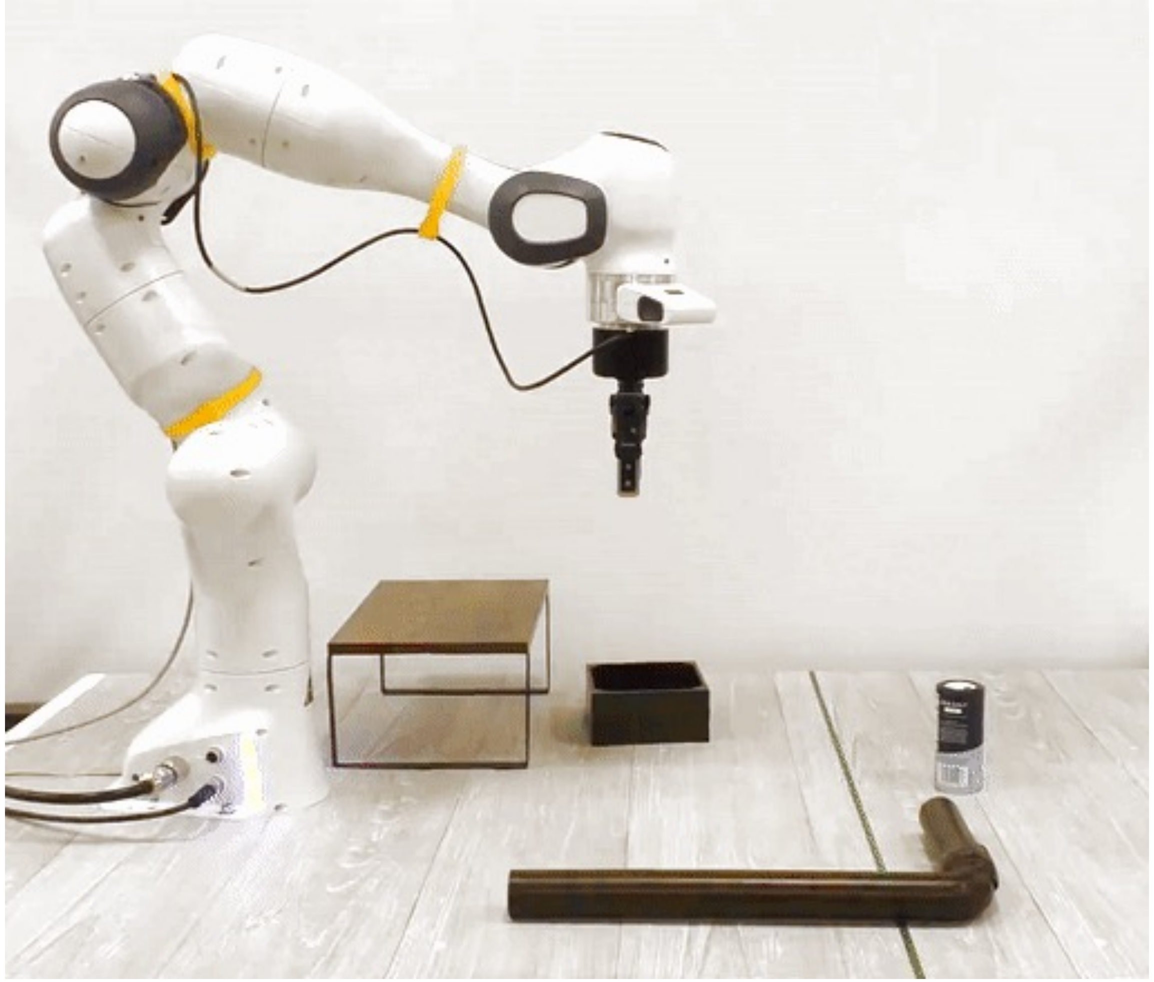


state

action

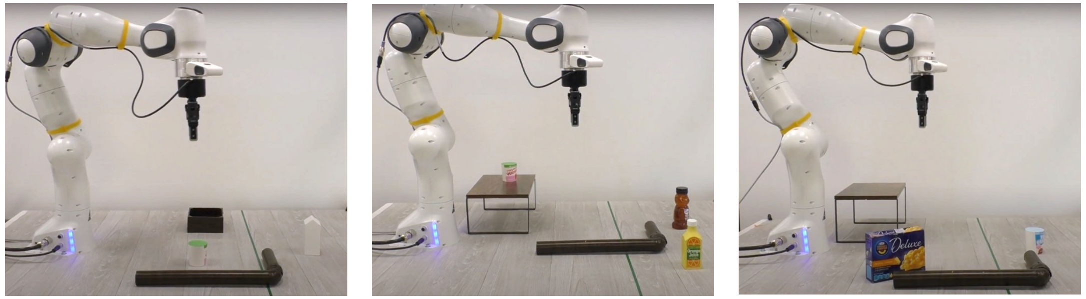


Robust skills for robotic manipulation

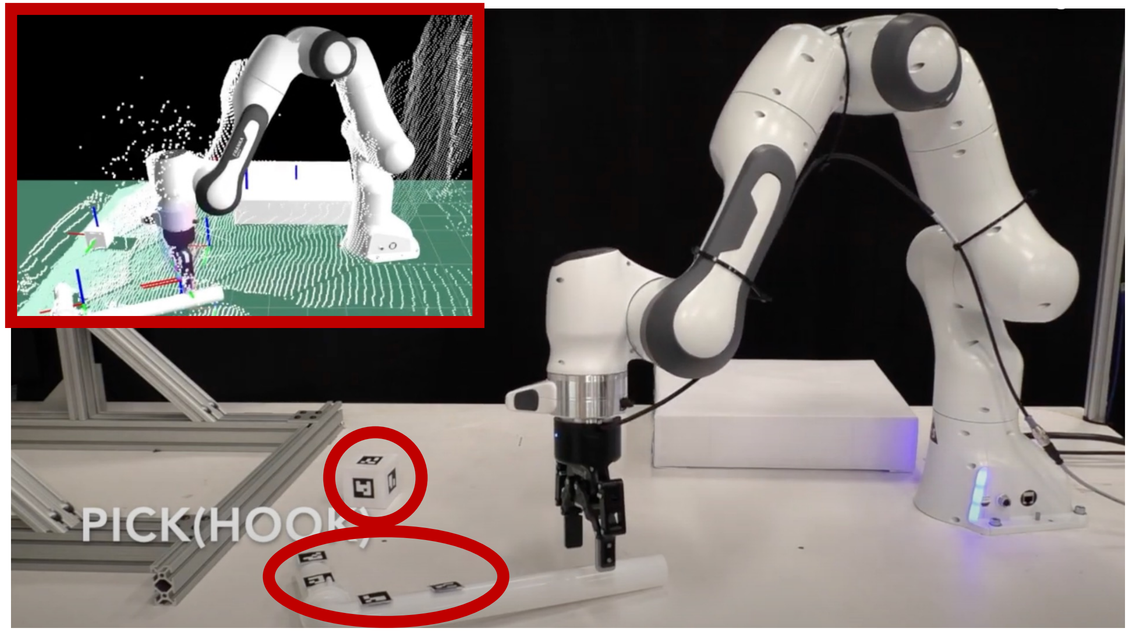


various objects

various arrangements



It is challenging to acquire robust skills that can solve complex tasks



Hand-designed skills
[Migimatsu and Bohg. RA-L 2020]

Cannot handle unknown objects

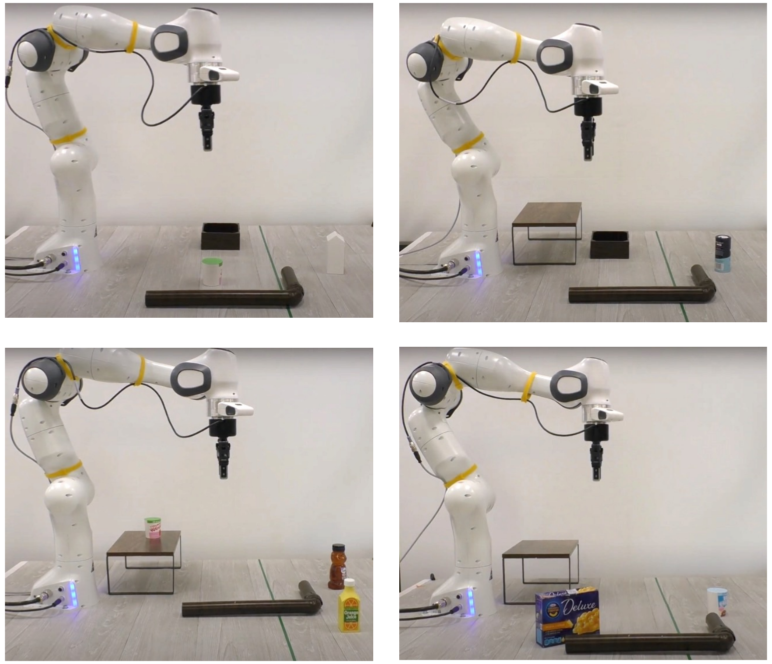
Learned skills
[Levine et al. ISER 2016]

Require large-scale data

Alternative data source: procedural content generation

Procedural content generation (PCG)

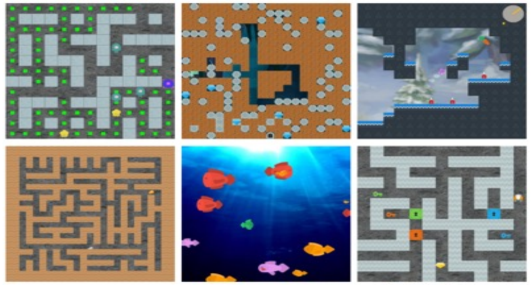
Enrich training data



Real-world robotics problems

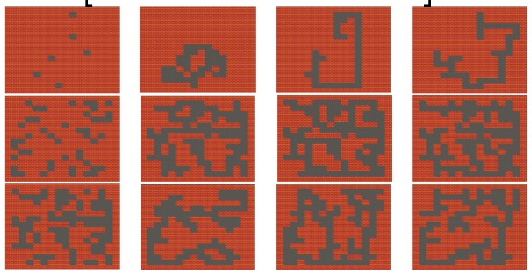
Procedural content generation for scalable learning

Challenge: how to generate **diverse** and **feasible** environments for complex skills.

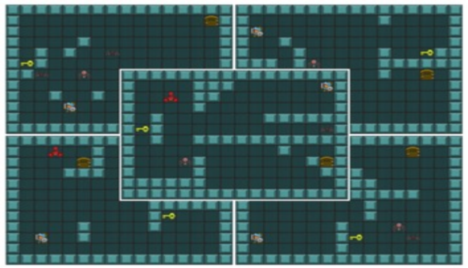


[Cobbe et al. 2019]

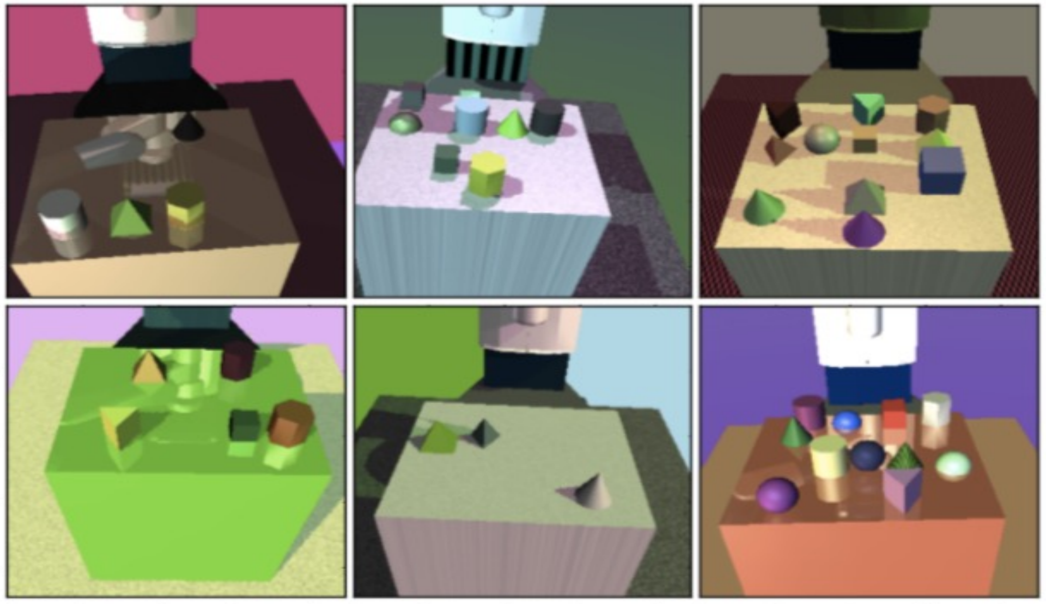
[Daniele et al. 2019]



[Khalifa et al. 2020]



[Bontrager et al. 2020]



[Tobin et al. 2017]

Grid-world domain

discrete spaces, simple dynamics

Robotic manipulation

limited variations, simple environments

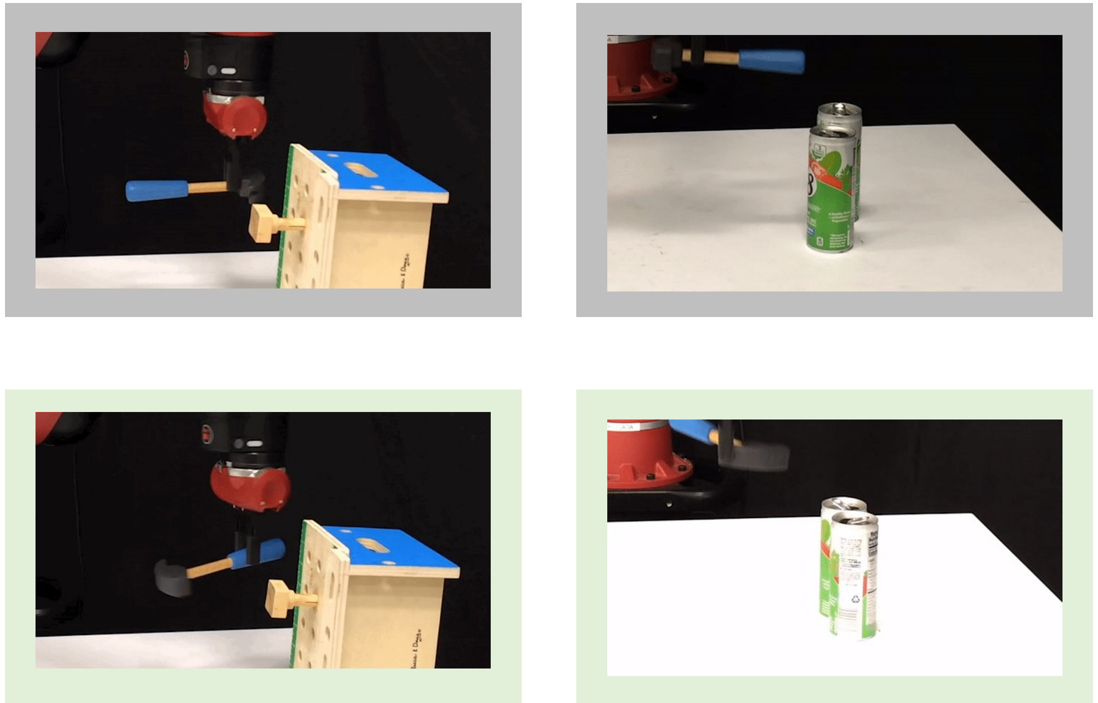
[Wang et al. 2018; Izatt and Tedrake 2020; Fisher et al. 2012; Izadinia et al. 2017; Majerowicz et al. 2013; Schwarz and Behnke 2020; Wang et al. 2019; Raileanu & Rocktaschel, 2020; Kolve et al., 2017; Xia et al., 2018; Savva et al., 2019; Yu et al., 2019]

Tool-use skills

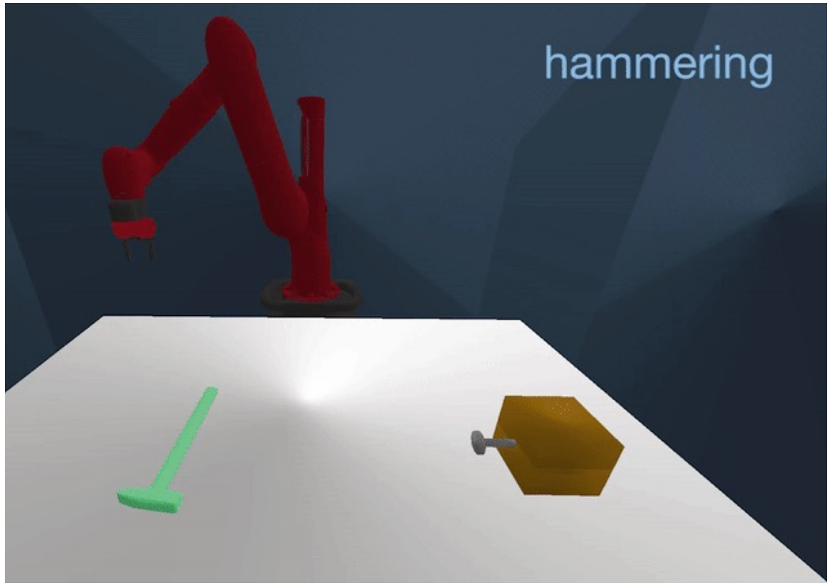
hammering sweeping

failure

success

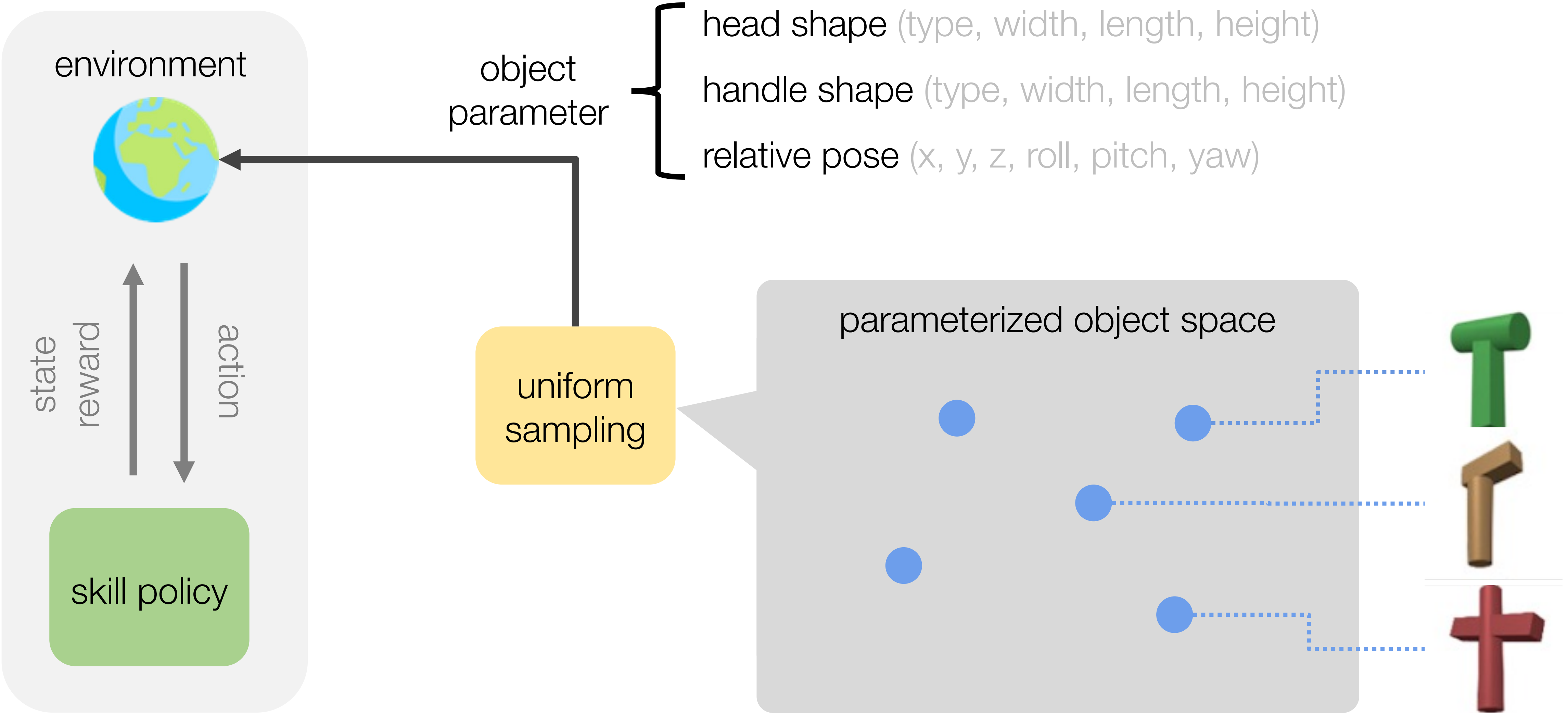


simulation



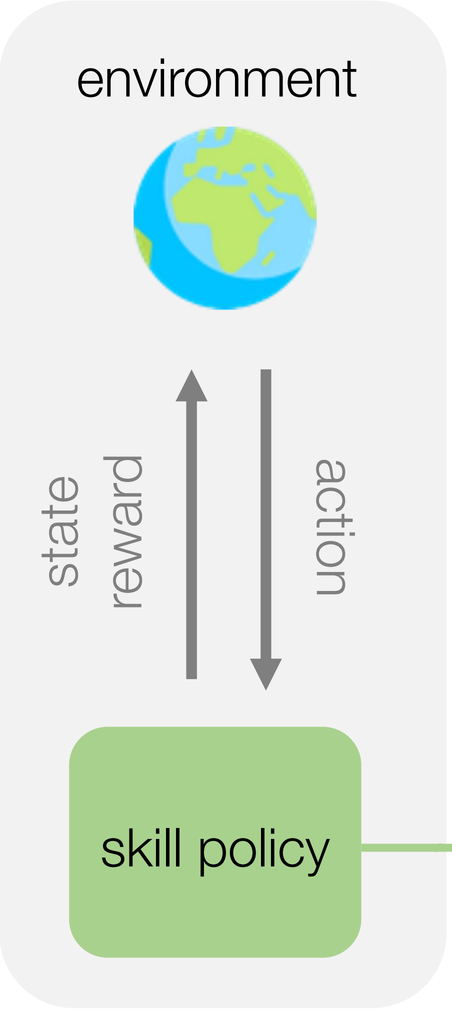
Fang, Zhu, Garg, Kurenkov, Mehta, Fei-Fei, Savarese. *IJRR* 2019

Self-supervised skill learning with procedurally generated objects

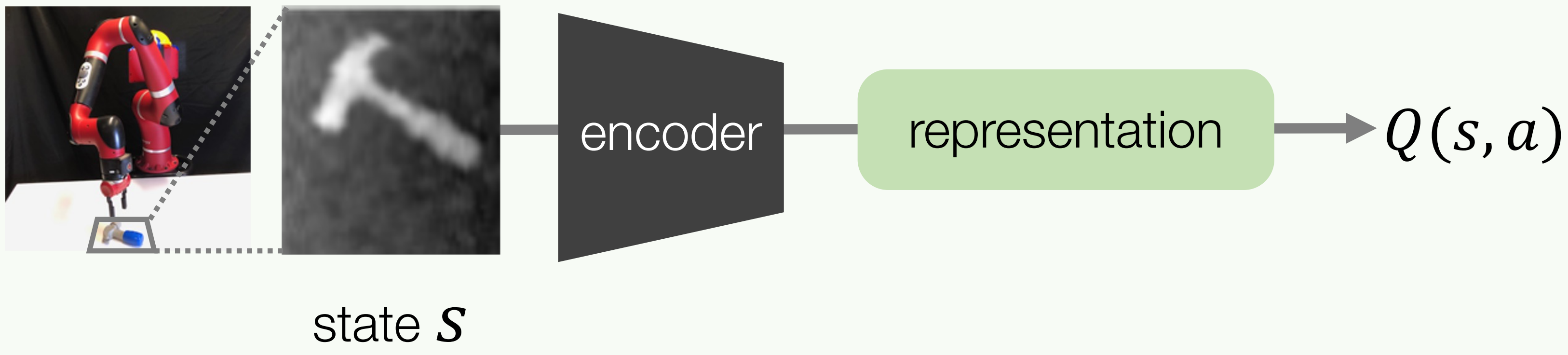


Fang, Zhu, Garg, Kurenkov, Mehta, Fei-Fei, Savarese. *IJRR 2019*

Self-supervised skill learning with procedurally generated objects



$Q(s,a)$ estimates the probability of success



The action is selected as $a^* = \arg\max_a Q(s,a)$

Fang, Zhu, Garg, Kurenkov, Mehta, Fei-Fei, Savarese. *IJRR 2019*

Training in simulation:

20,000 synthetic objects

100,000 training trajectories

(equivalent to 800 robot hours)

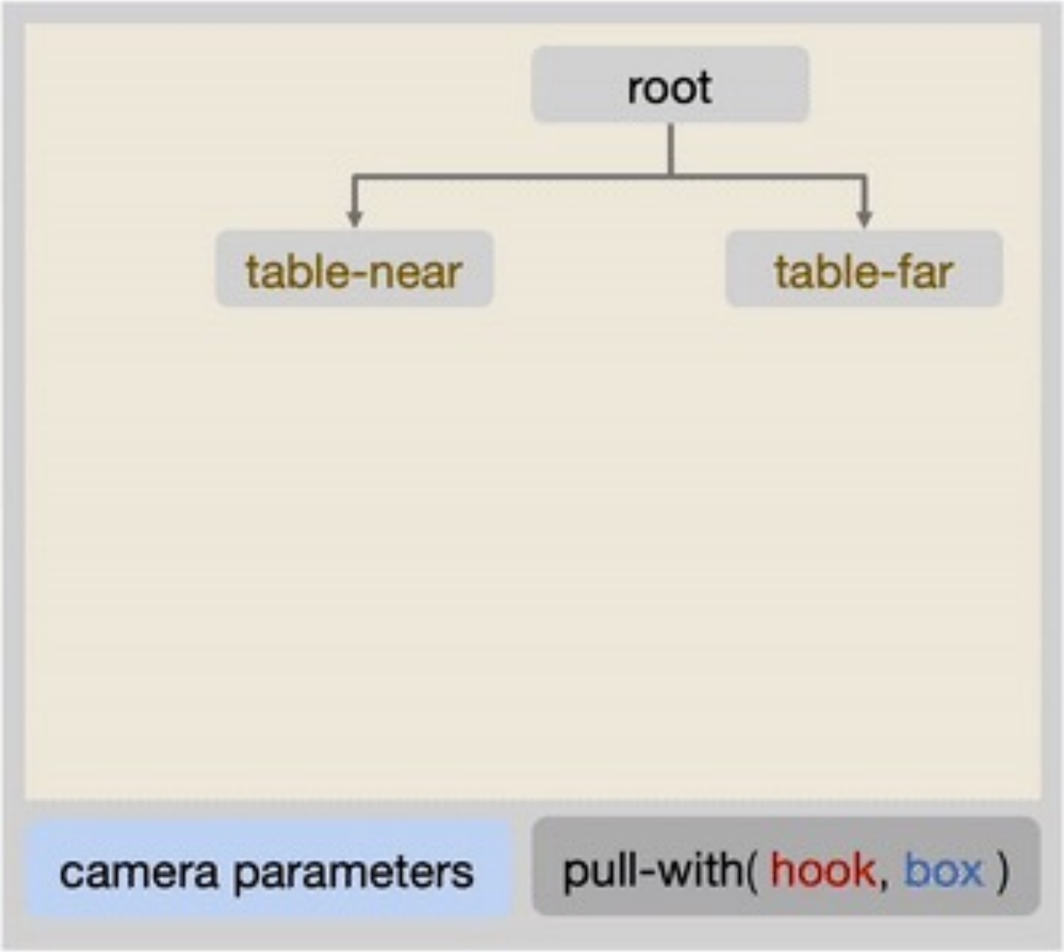
Evaluation in the real world:

Hammering: 71.1%

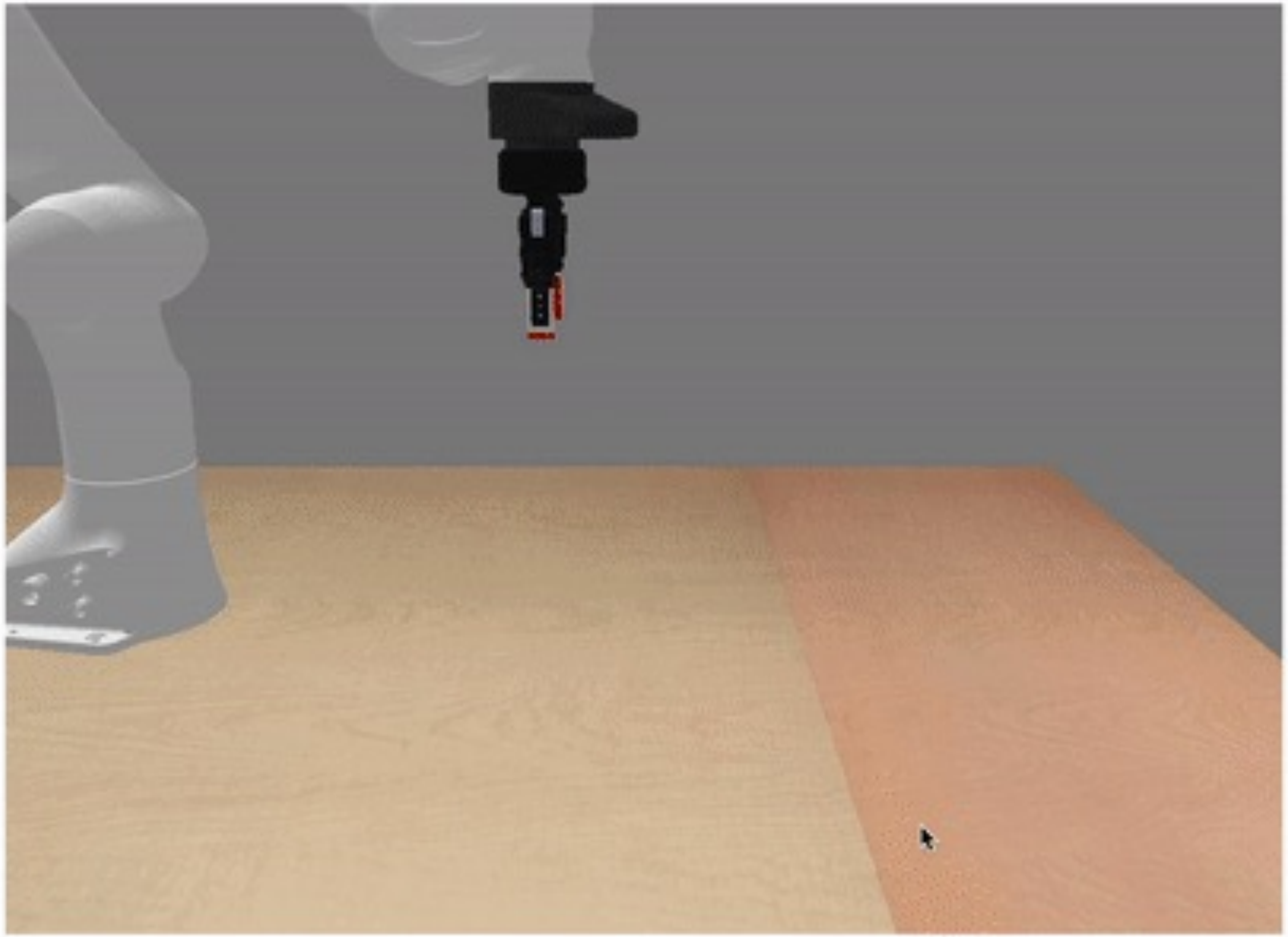
Sweeping: 80.0%

Fang, Zhu, Garg, Kurenkov, Mehta, Fei-Fei, Savarese. *IJRR 2019*

Procedural generation of complex environments



Environment parameter



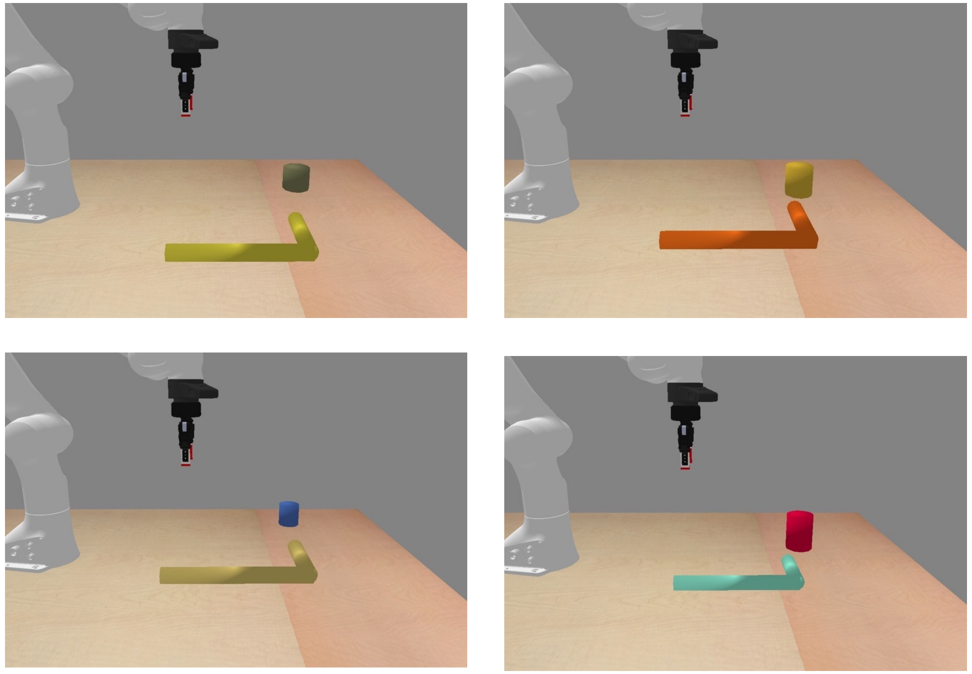
Procedurally generated environment

Fang, Zhu, Garg, Kurenkov, Mehta, Fei-Fei, Savarese. *IJRR 2019*

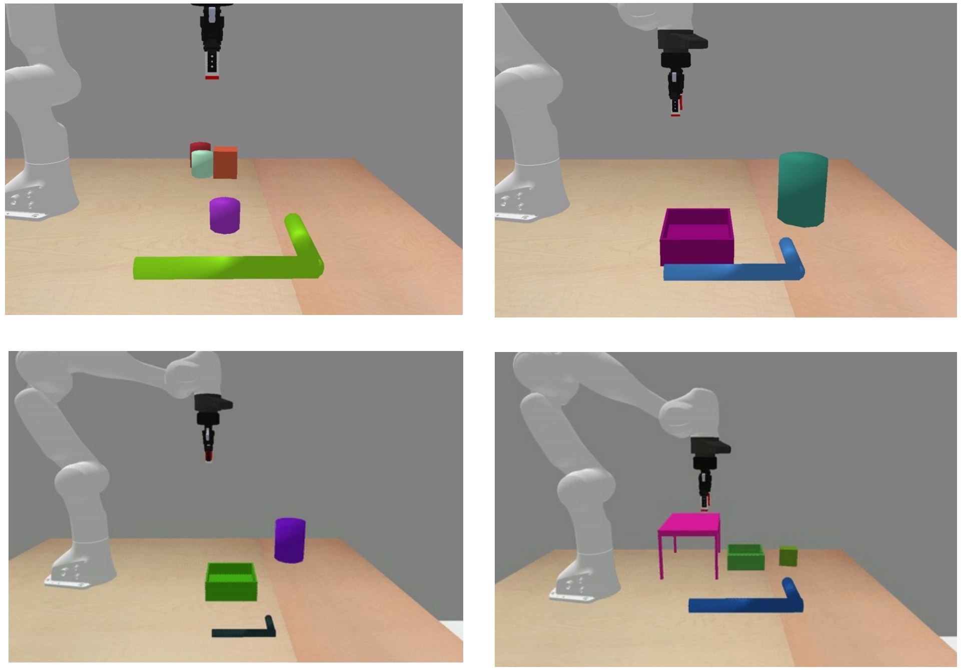
Procedural generation of complex environments

skill: $\mathrm{pull\_with}(x, y)$

reward: $r = \begin{cases} 1 & \text{if } \mathrm{on}(x, \mathrm{table\_near}) \\ 0 & \text{otherwise} \end{cases}$



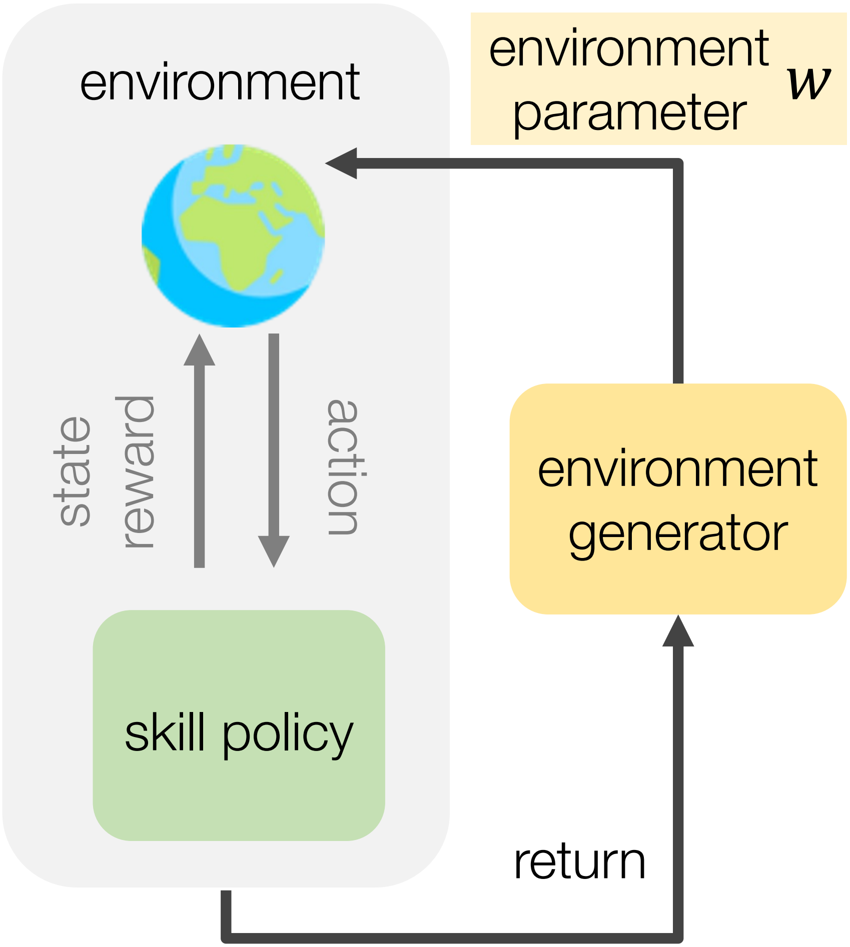
Repetitive environments



Infeasible environments

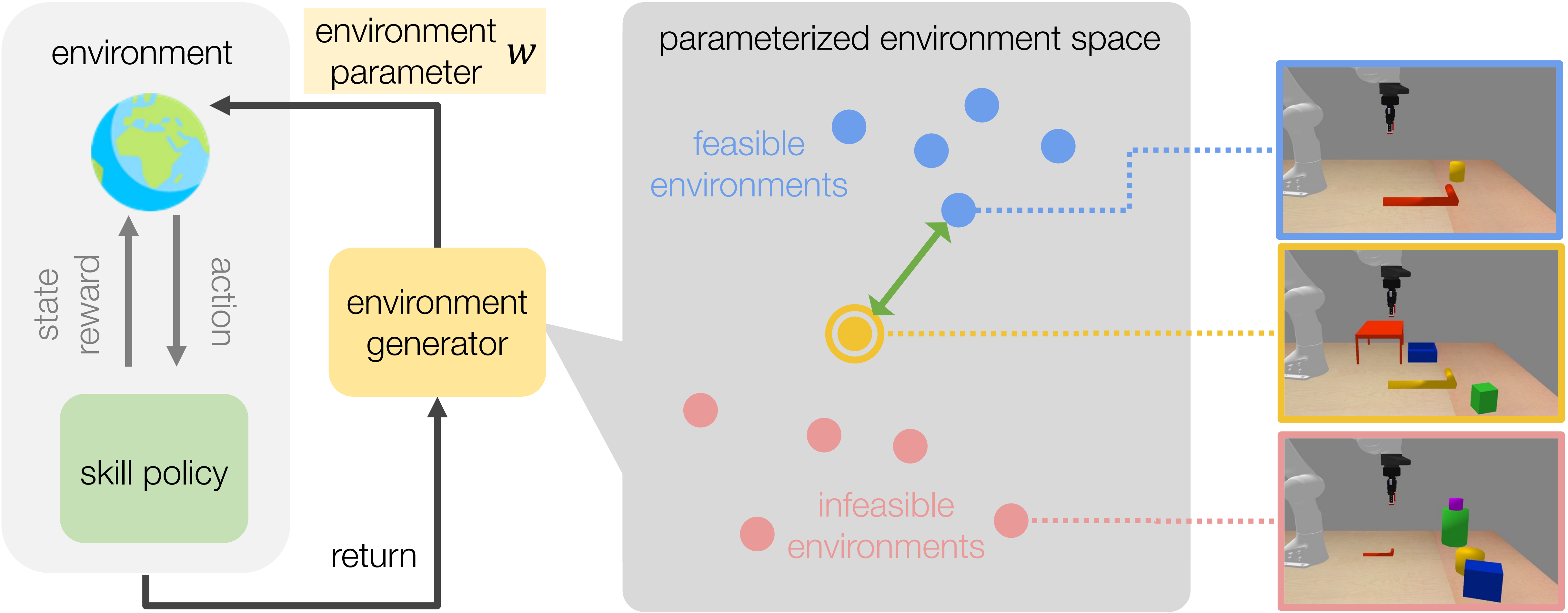
Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

Active Task Randomization (ATR)



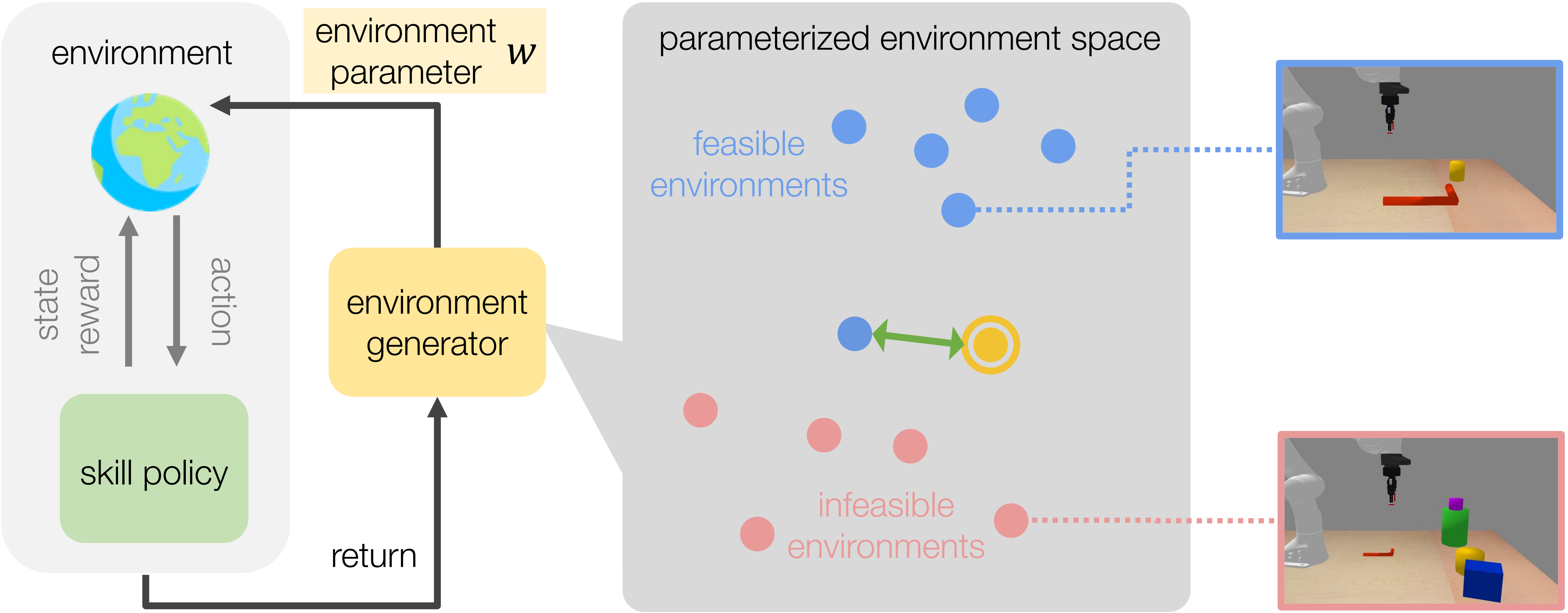
Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

Active Task Randomization (ATR)



Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

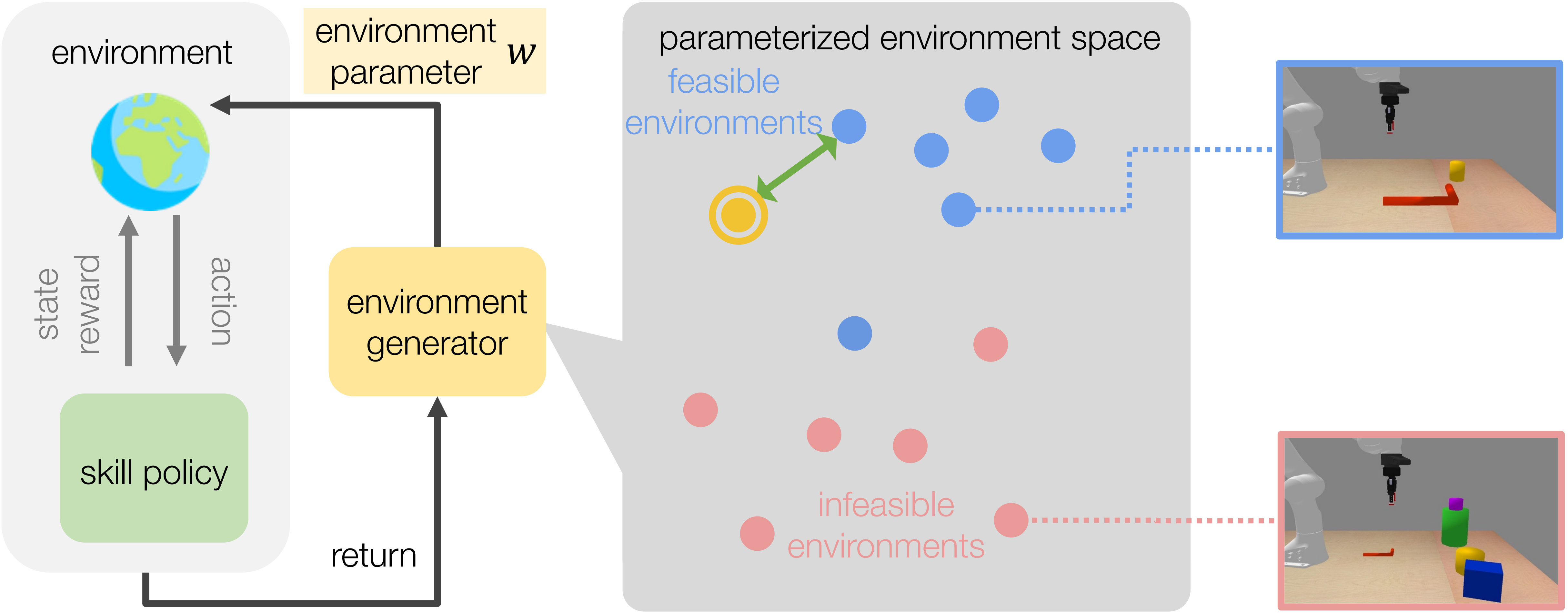
Active Task Randomization (ATR)



Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

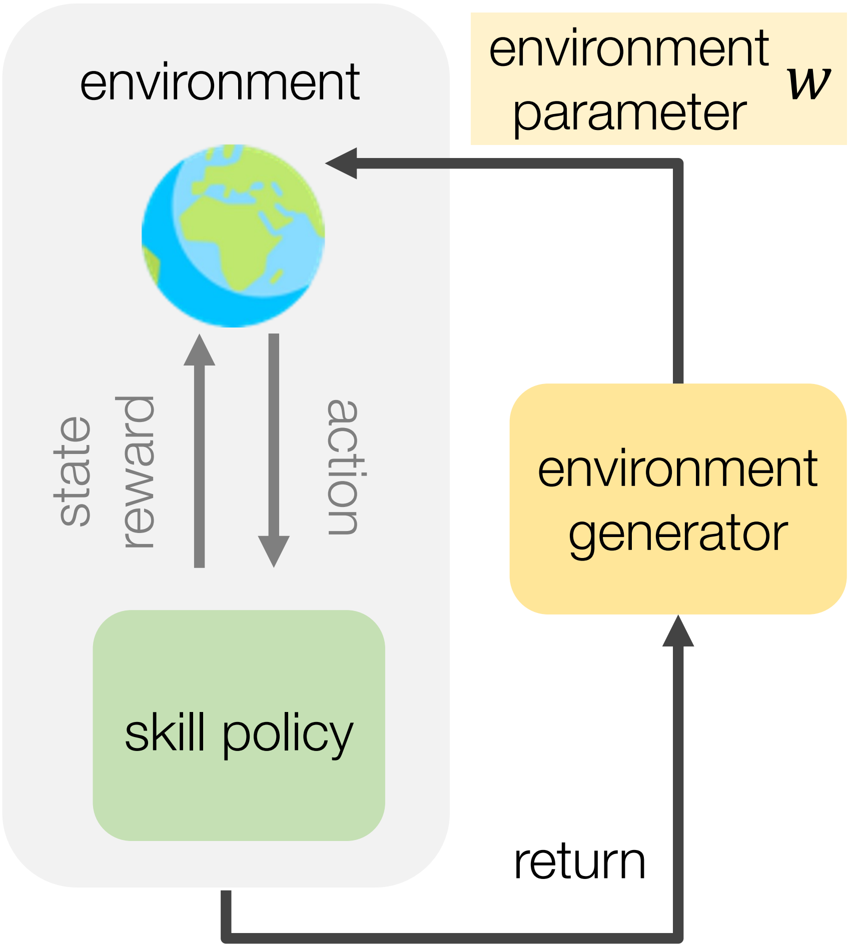
Active Task Randomization (ATR)

Key idea: adaptively estimate **feasibility** and **diversity** of the sampled environments.



Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2022

Adaptive estimation of feasibility and diversity

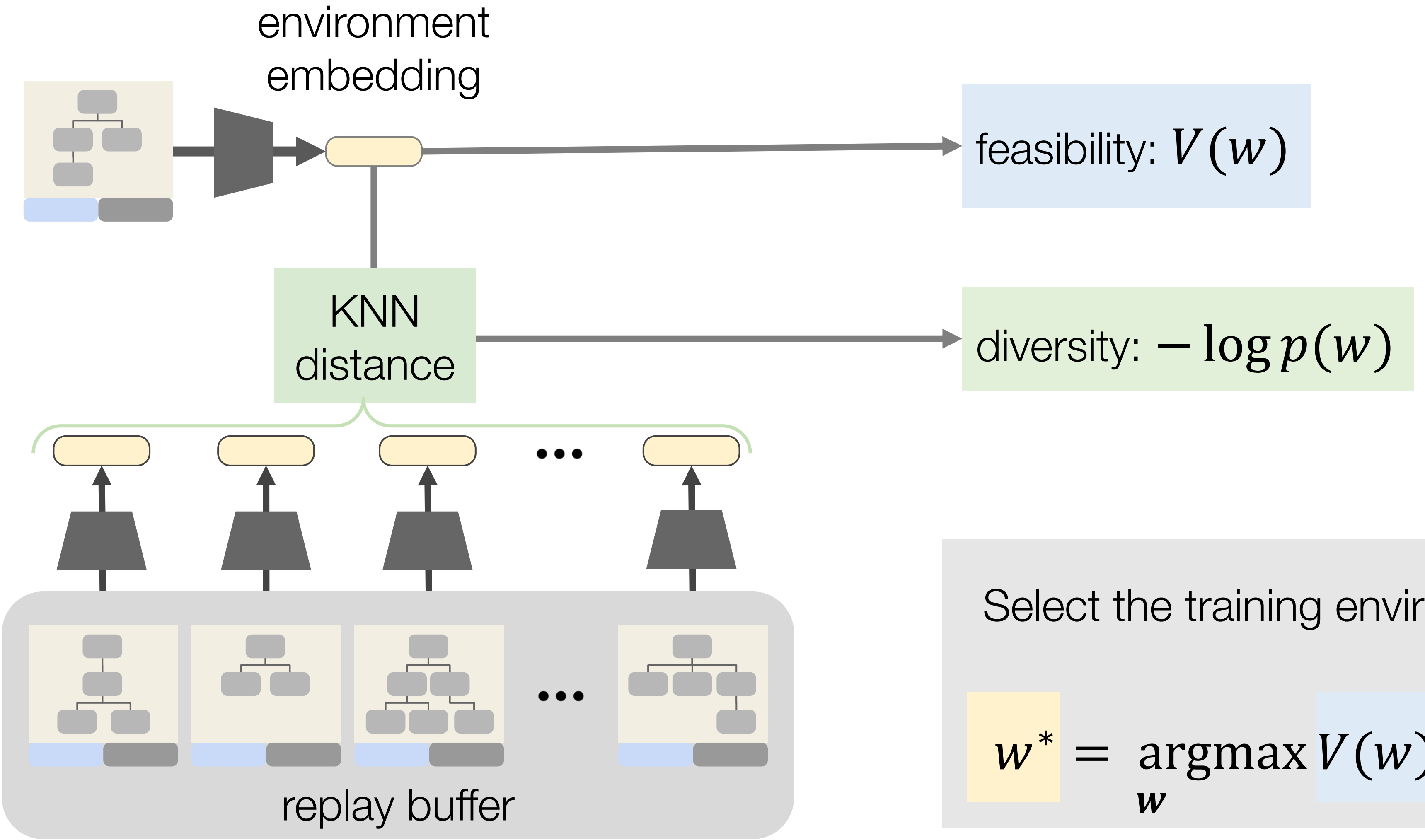


feasibility: $V(w)$ — Expected rewards achieved by the current skill policy

diversity: $-\log p(w)$ — Density of the environment parameter

Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

Adaptive estimation of feasibility and diversity



Value network

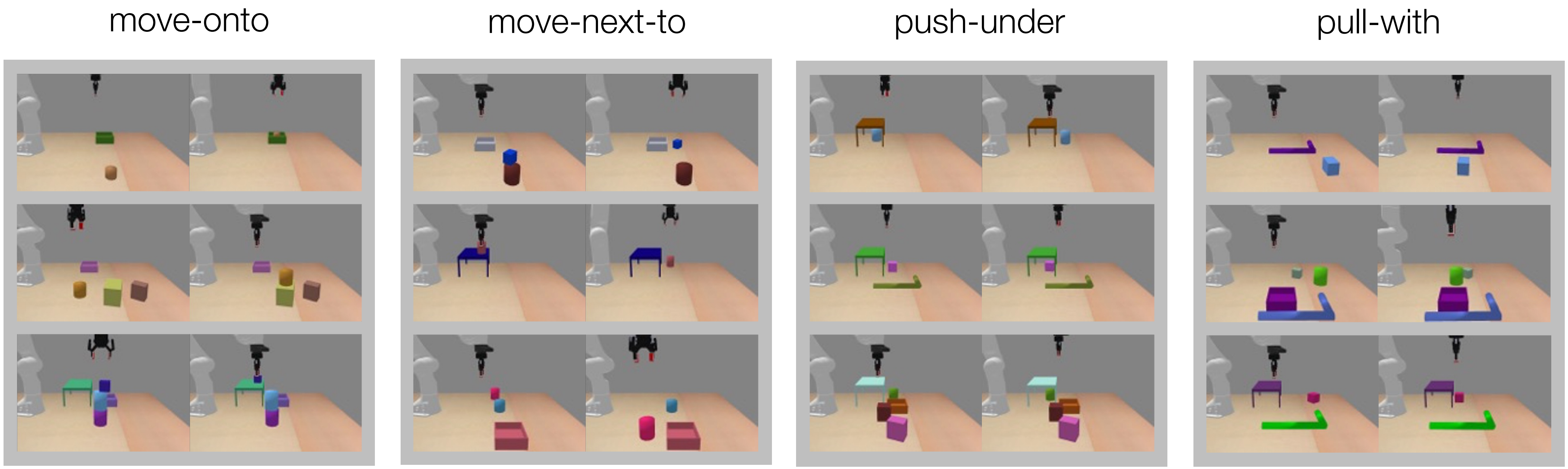
Non-parametric estimation of density

Select the training environment parameter by

$$w^* = \underset{w}{\operatorname{argmax}}\, V(w) - \beta \log p(w)$$

Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

Procedurally generated training environments

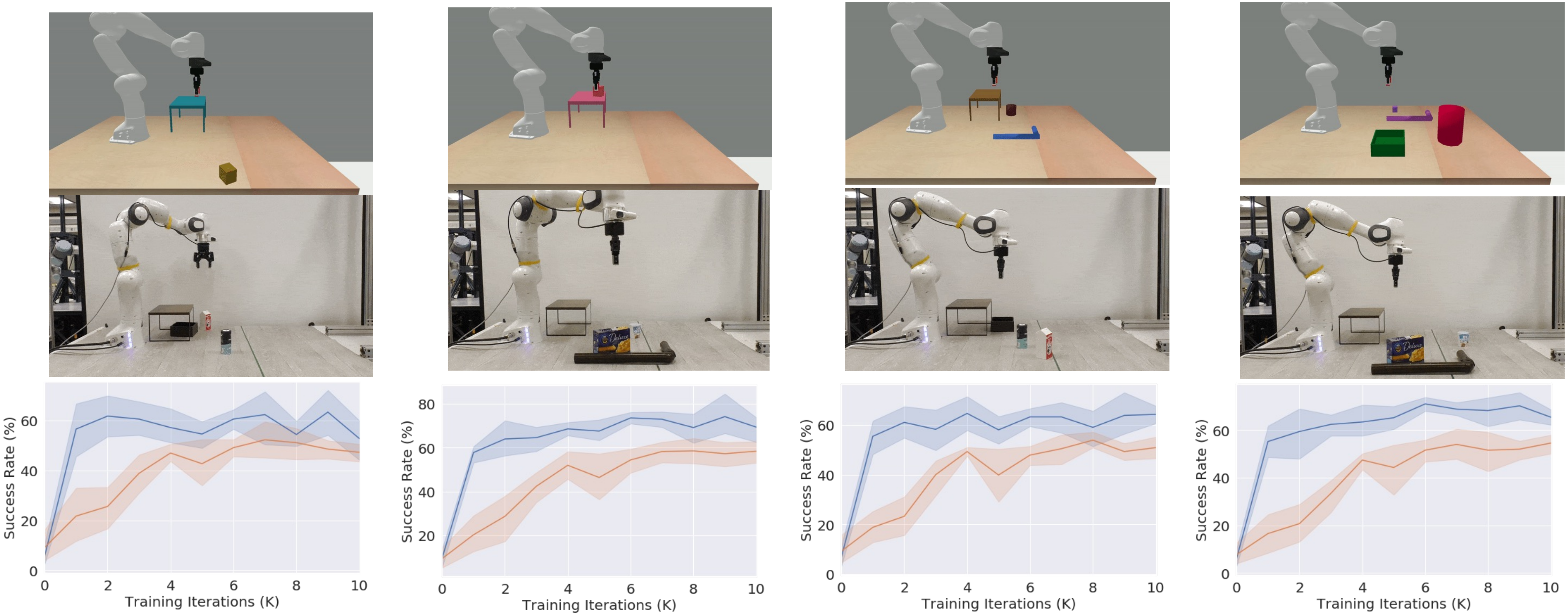


Training with 10,000 generated environments in simulation.

Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

Learned skills

move-onto | move-next-to | push-under | pull-with



ATR (Ours) Uniform

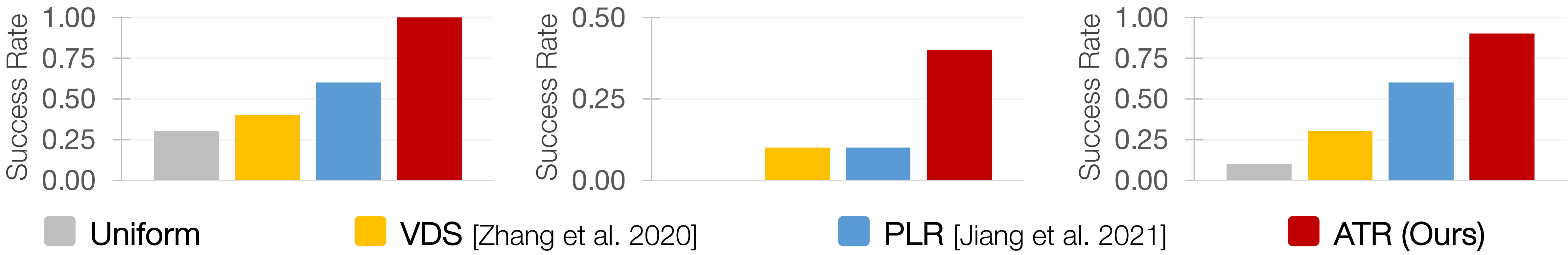
Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

Composing learned skills to solve sequential manipulation tasks

[Jiang et al. 2021; Zhang et al. 2020]

Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

Composing learned skills to solve sequential manipulation tasks

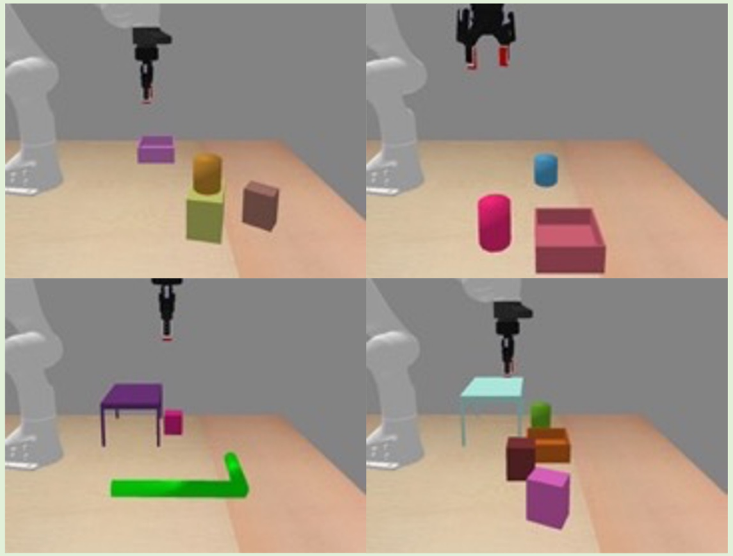


Fang*, Migimatsu*, Mandlekar, Fei-Fei, Bohg. 2023

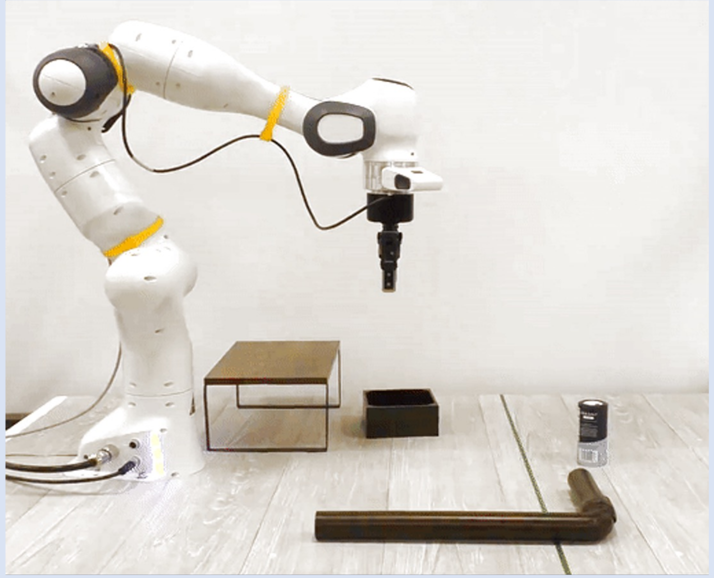
Summary

- Procedurally generate environments in simulation to enrich training data.
- Select suitable environments by adaptively estimating the diversity and feasibility.

Training



Testing

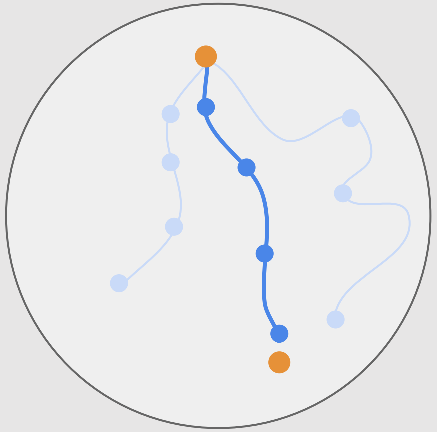


Generalization through Generation:

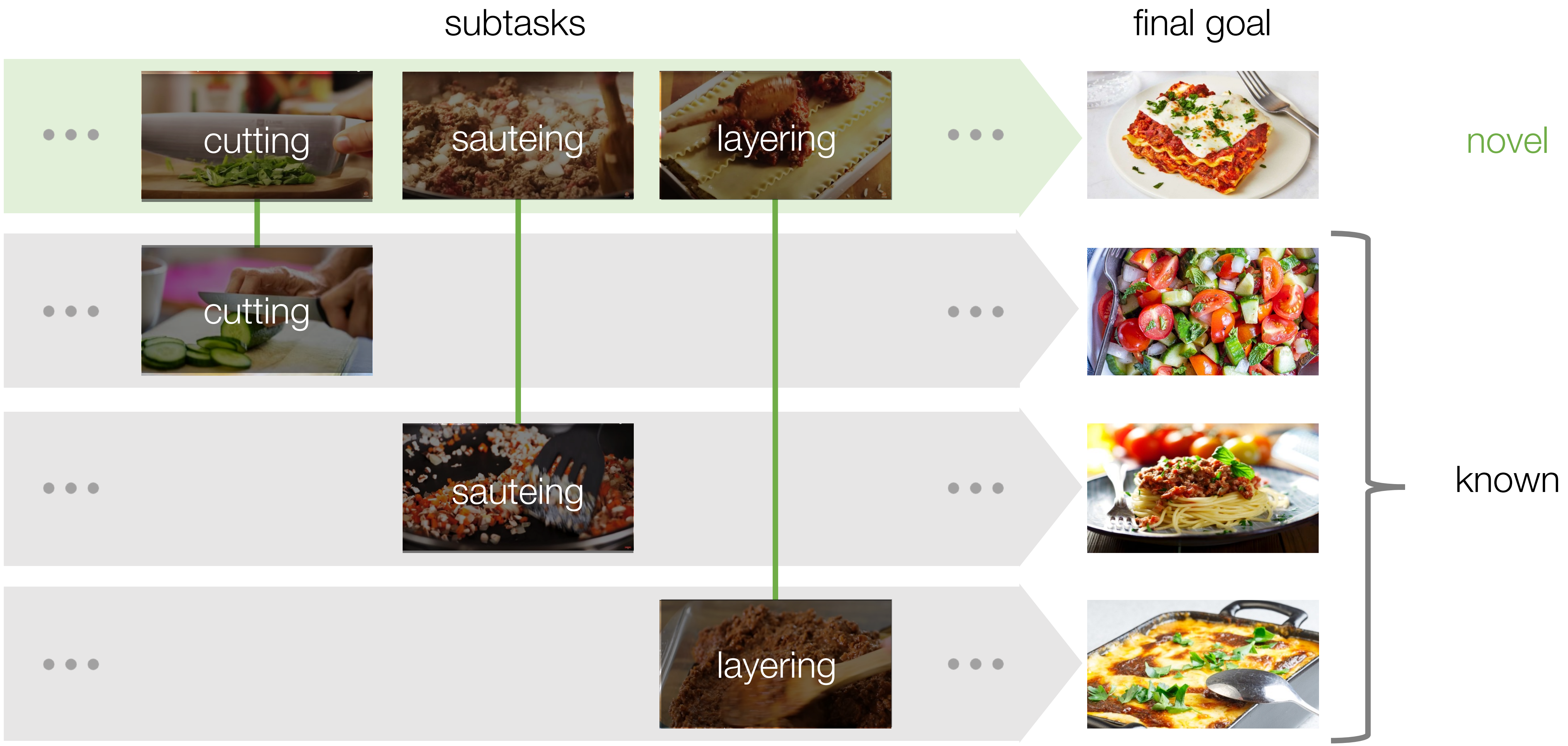
Learning Long-Horizon Tasks with Limited Supervision

Learning Robust Skill via Environment Generation

Adapting Prior Skills via Goal Generation

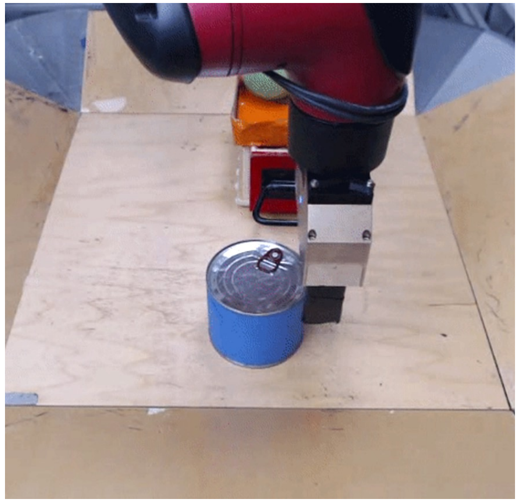


Reusing and adapting prior skills



Solving sequential tasks specified by goals

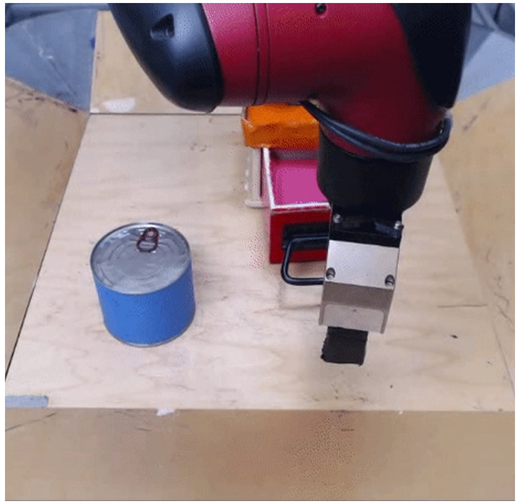
initial state



drawer → open

can → bottom left

final goal



[Kaelbling 1993; Andrychowicz et al. 2017; Pong et al. 2020; Ding et al. 2019; Khazatsky et al. 2021]

Solving sequential tasks specified by goals

solution 1: planning + motion primitives

push(can) → reach(drawer) → open(drawer)

solution 2: reward shaping

reward = reward_can + reward_drawer + reward_gripper

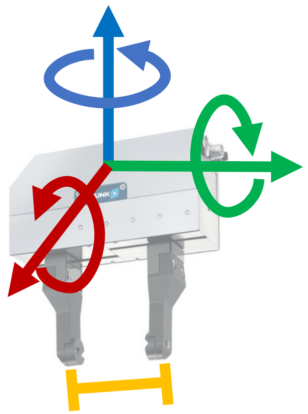
Requires non-trivial domain knowledge about the task

[Nilsson 1984; Malcolm and Smithers 1990 Kaelbling 2011; Kaelbling 2013; Srivastava 2014; Toussaint 2015; Dantam 2016; Toussaint 2018; Garret 2020]

Solving sequential tasks specified by goals

Exploring over long horizons without immediate feedback leads to poor performance.

action:



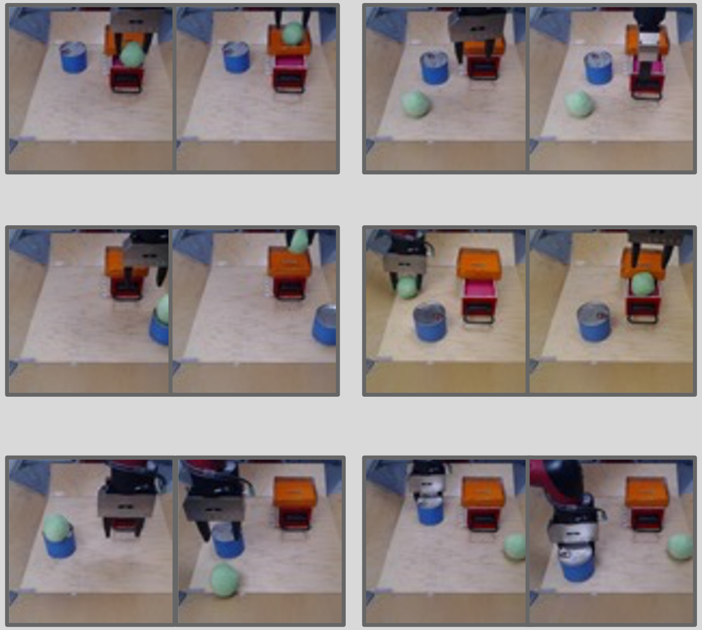
5 Hz

reward: $r = \begin{cases} 0 & \|s - g\| \leq \epsilon \\ -1 & \|s - g\| > \epsilon \end{cases}$

Learning sequential tasks by leveraging prior skills across tasks

2.3k short-horizon trajectories collected through teleoperation

prior experiences

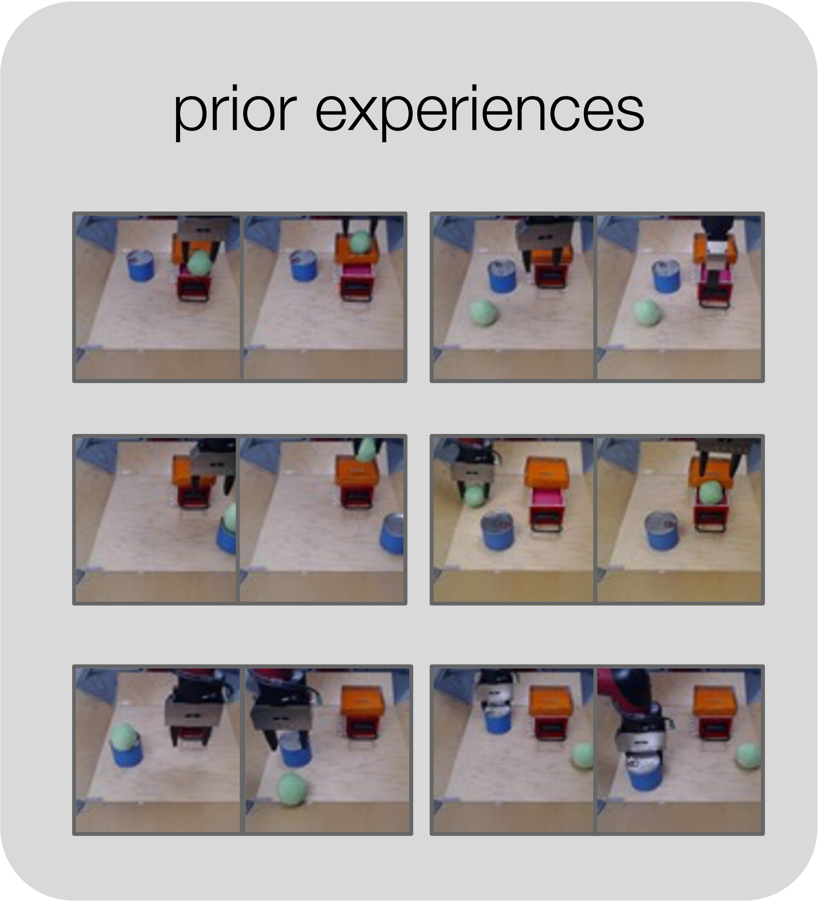


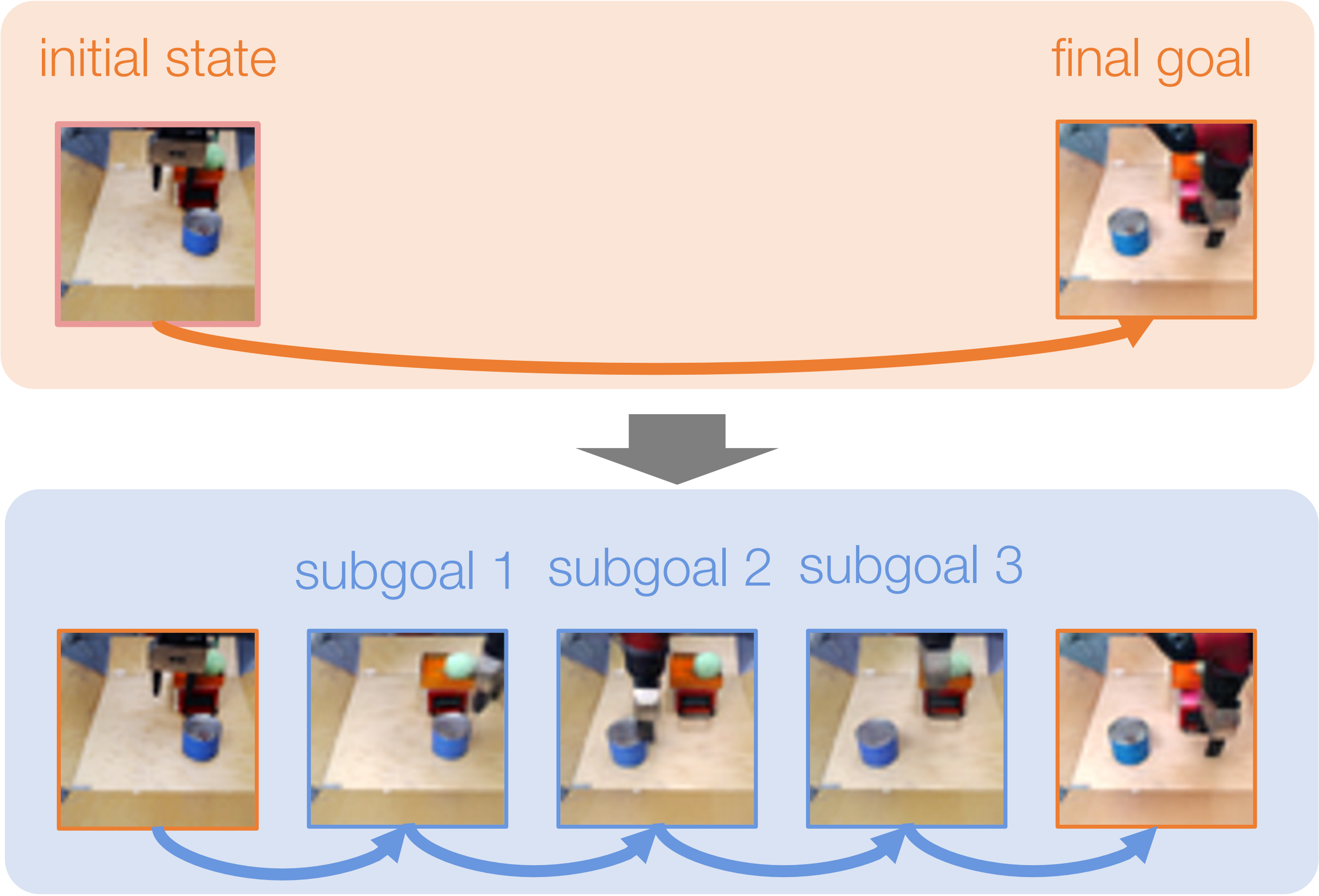
adaptation

Fang, Zhu, Garg, Savarese, Fei-Fei. CoRL 2019

Decompose the novel goal into familiar subgoals

Challenge: how to propose feasible and useful subgoals in high-dimensional space?

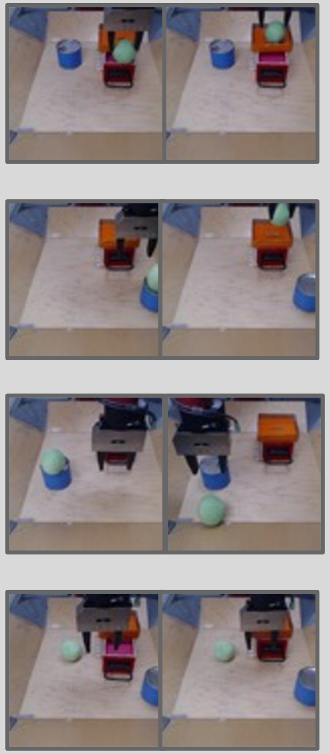




Fang, Zhu, Garg, Savarese, Fei-Fei. CoRL 2019

Planning-to-Practice (PTP)

prior experience



affordance : the possibility of an action on an object

$$m(s' \mid s, u)$$

Gibson (1979). *The Ecological Approach to Visual Perception.*

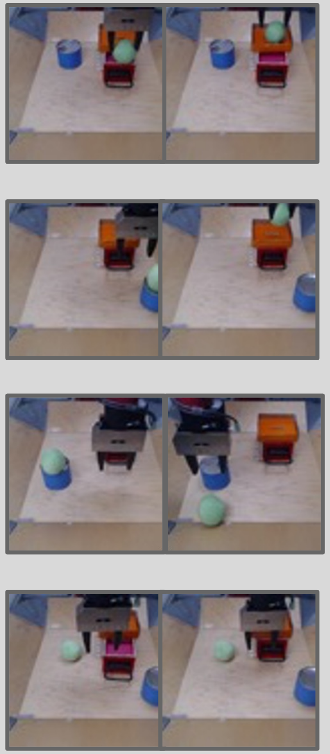
$$\pi(a \mid s, g)$$

goal-conditioned policy

Fang*, Yin*, Nair, Levine. IROS 2022

Planning-to-Practice (PTP)

prior experience



affordance

$$m(s' \mid s, u)$$

future state
H steps away

current state

noise

$$\pi(a \mid s, g)$$

goal-conditioned policy

Model the distribution of feasible future states $p(s'|s)$ within H steps.

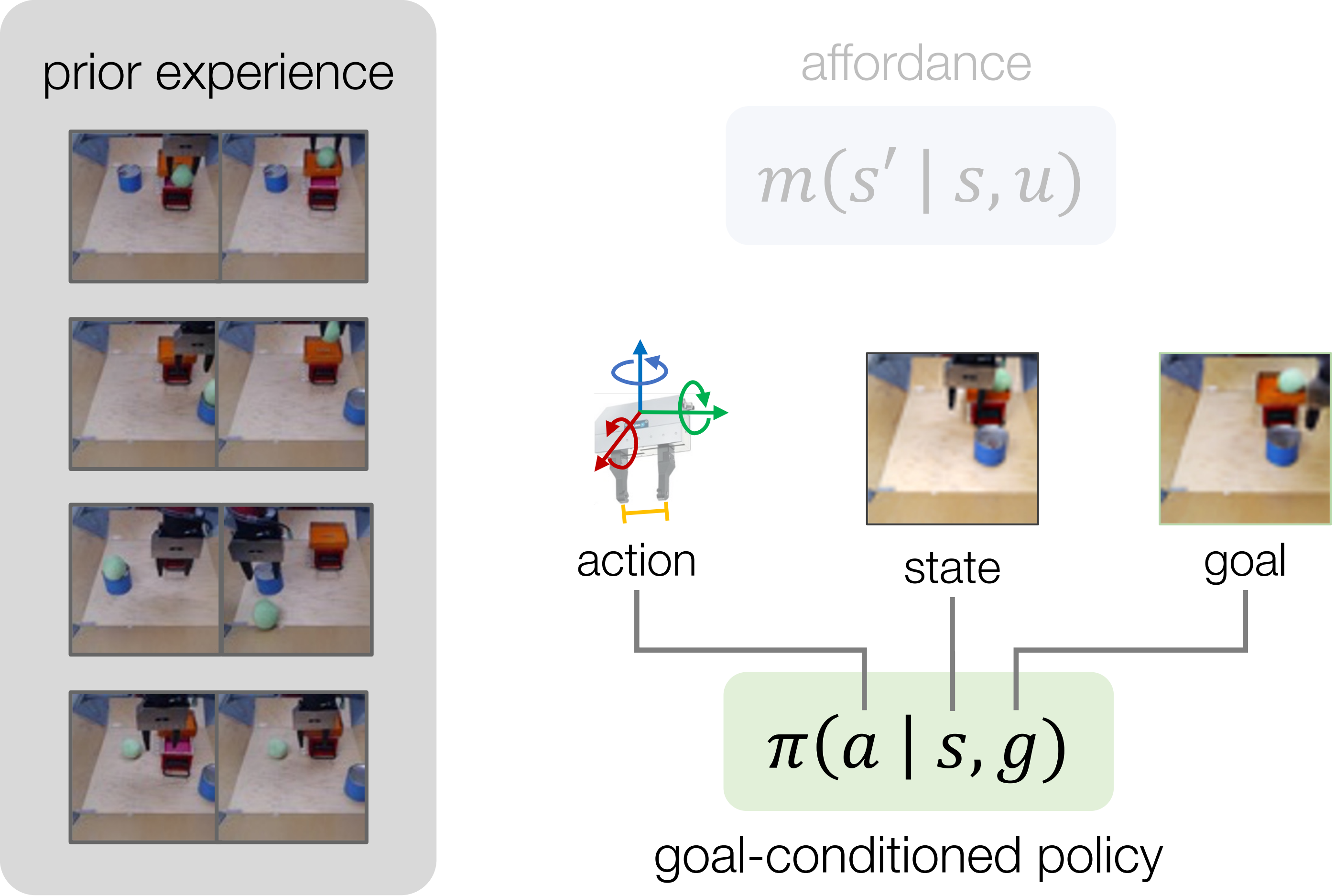
Implemented with a **Conditional Variational Autoencoder (CVAE)**[1].

[1][Sohn et al., 2015]

Trained using transitions sampled from the prior experience.

Fang*, Yin*, Nair, Levine. IROS 2022

Planning-to-Practice (PTP)

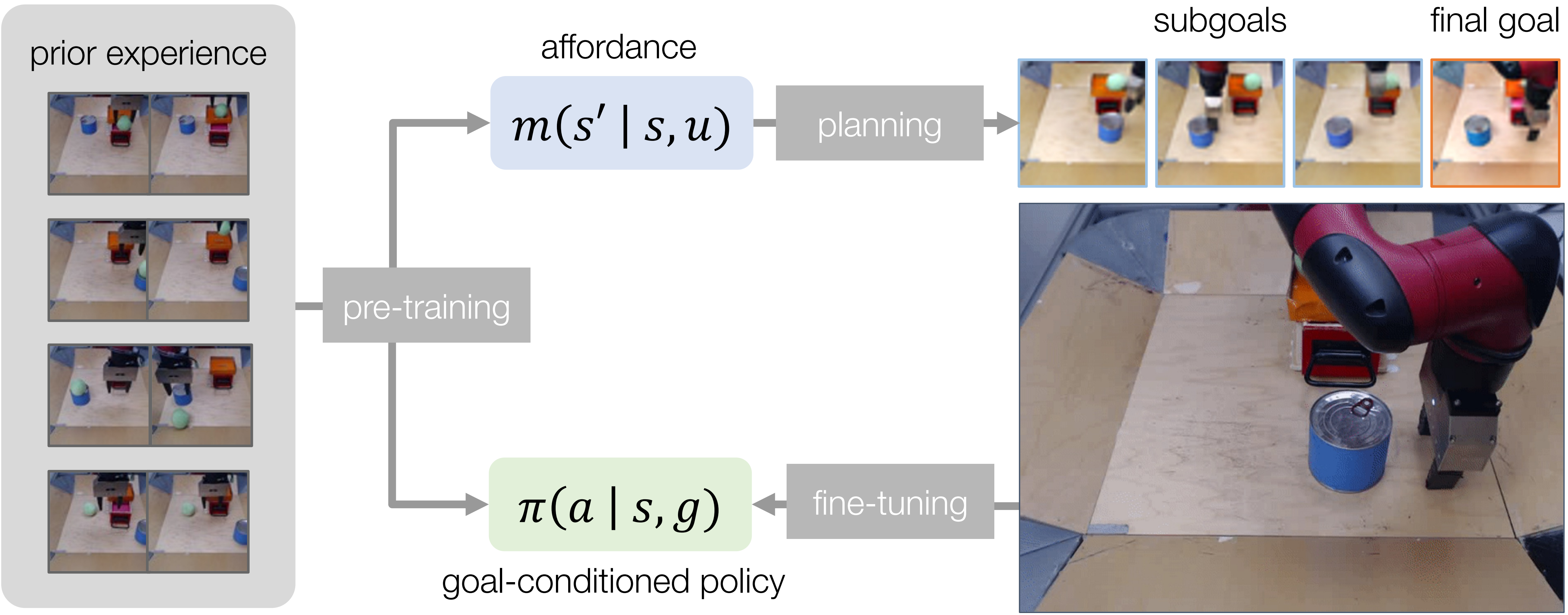


Select the action a to reach the goal g from the current state s.

Trained with the goal-reaching reward in a **self-supervised** manner.

Fang*, Yin*, Nair, Levine. IROS 2022

Planning-to-Practice (PTP)

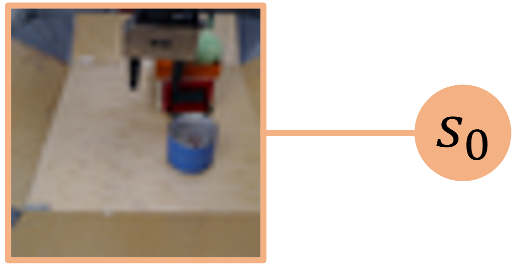


Fang*, Yin*, Nair, Levine. IROS 2022

Recursive generation for subgoal planning

Recursively generate K-step subgoals $\hat{s}_{1:K}$ using the sampled noise $u_{1:K}$.

initial state

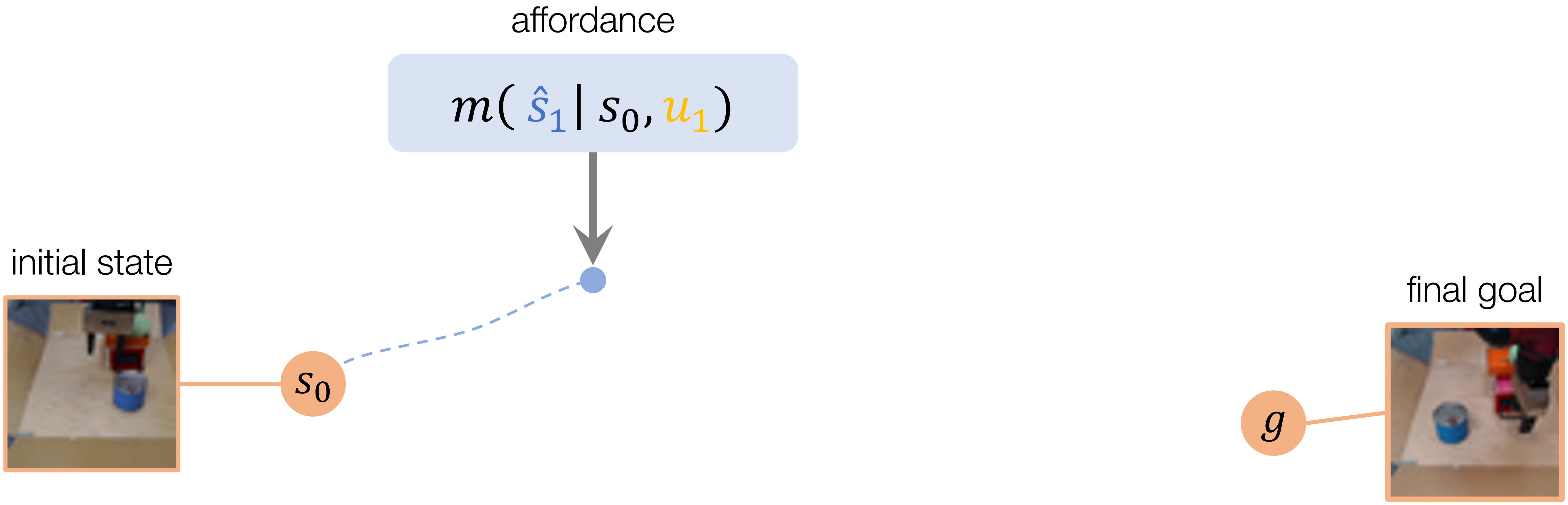


final goal

Fang*, Yin*, Nair, Levine. IROS 2022

Recursive generation for subgoal planning

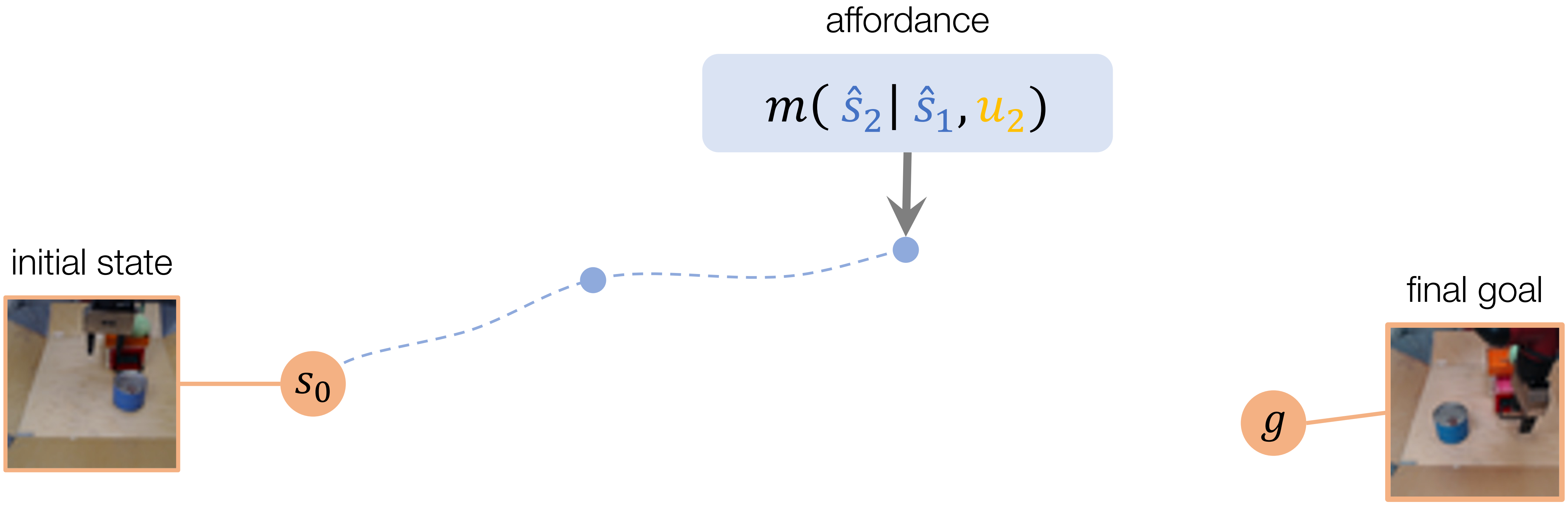
Recursively generate K-step subgoals $\hat{s}_{1:K}$ using the sampled noise $u_{1:K}$.



Fang*, Yin*, Nair, Levine. IROS 2022

Recursive generation for subgoal planning

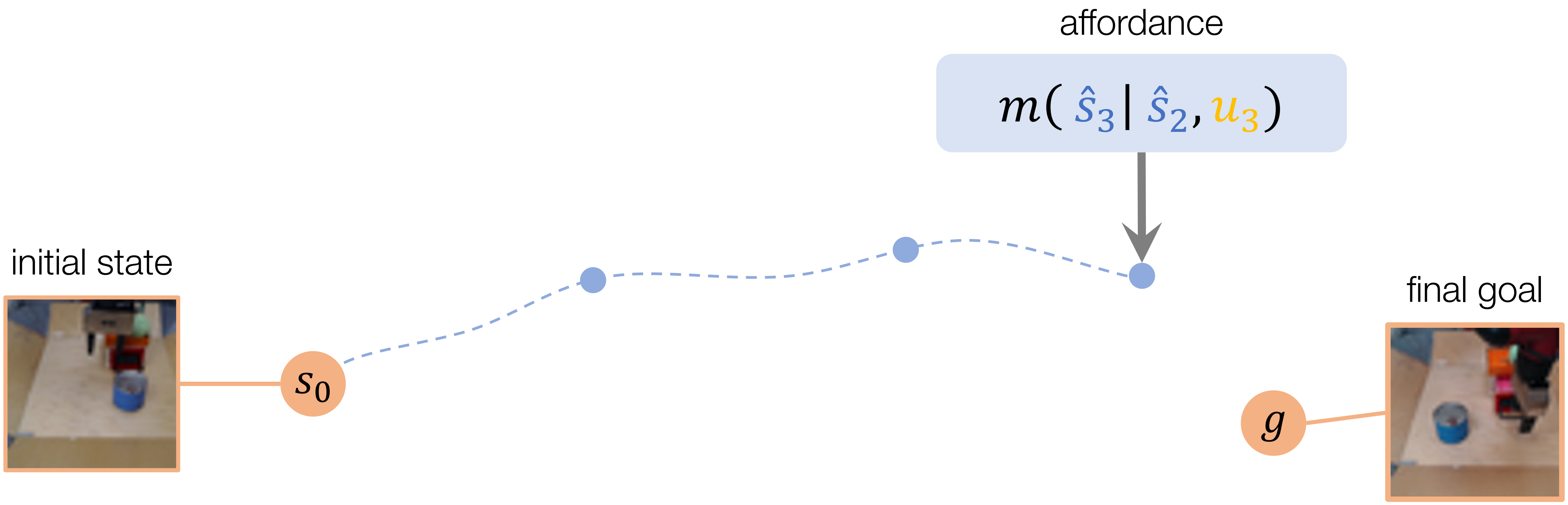
Recursively generate K-step subgoals $\hat{s}_{1:K}$ using the sampled noise $u_{1:K}$.



Fang*, Yin*, Nair, Levine. IROS 2022

Recursive generation for subgoal planning

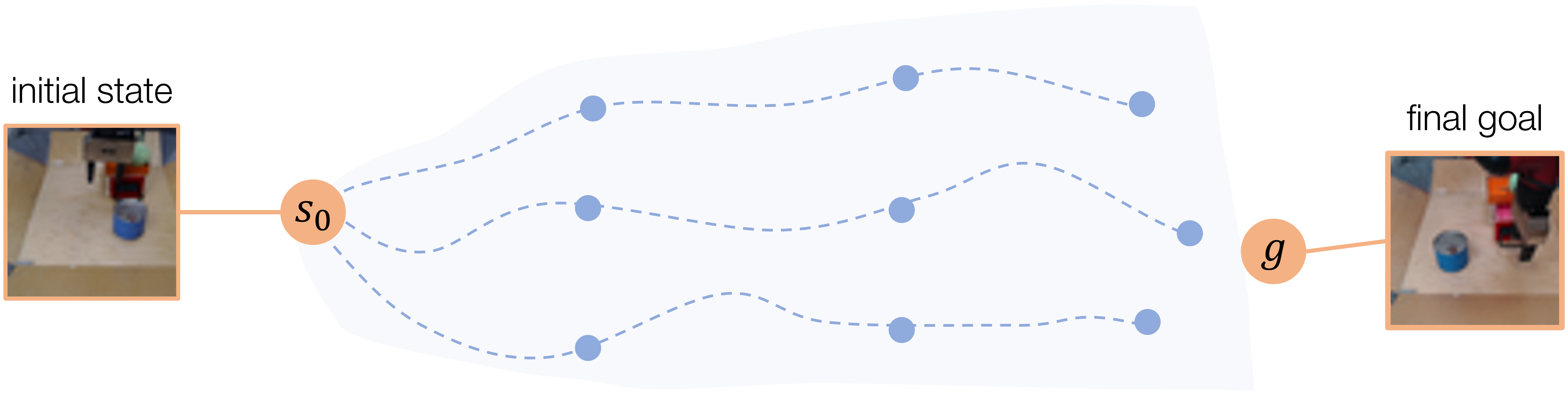
Recursively generate K-step subgoals $\hat{s}_{1:K}$ using the sampled noise $u_{1:K}$.



Fang*, Yin*, Nair, Levine. IROS 2022

Recursive generation for subgoal planning

Recursively generate K-step subgoals $\hat{s}_{1:K}$ using the sampled noise $u_{1:K}$.

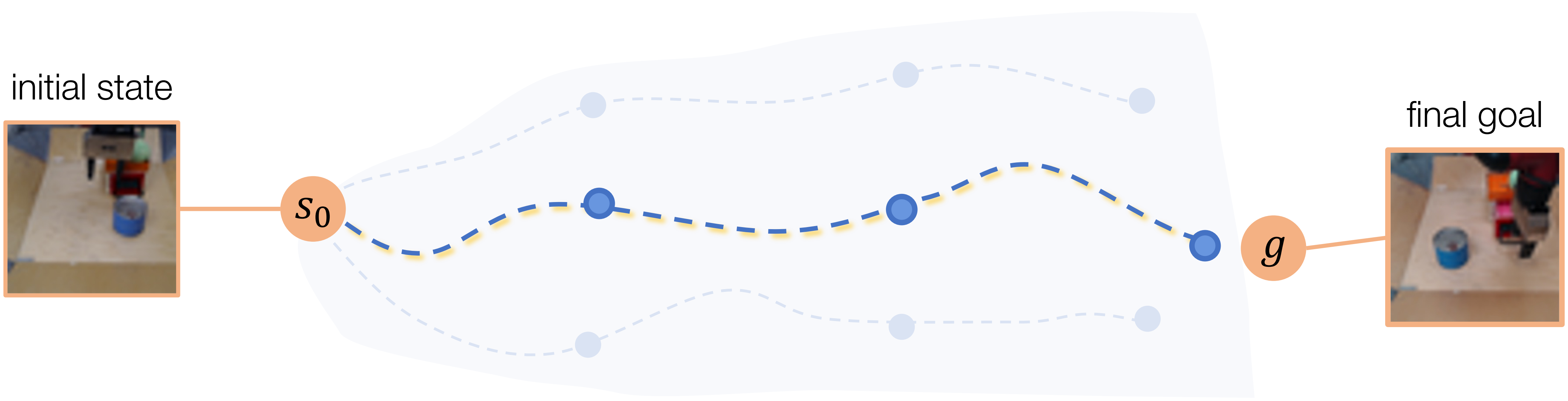


Fang*, Yin*, Nair, Levine. IROS 2022

Recursive generation for subgoal planning

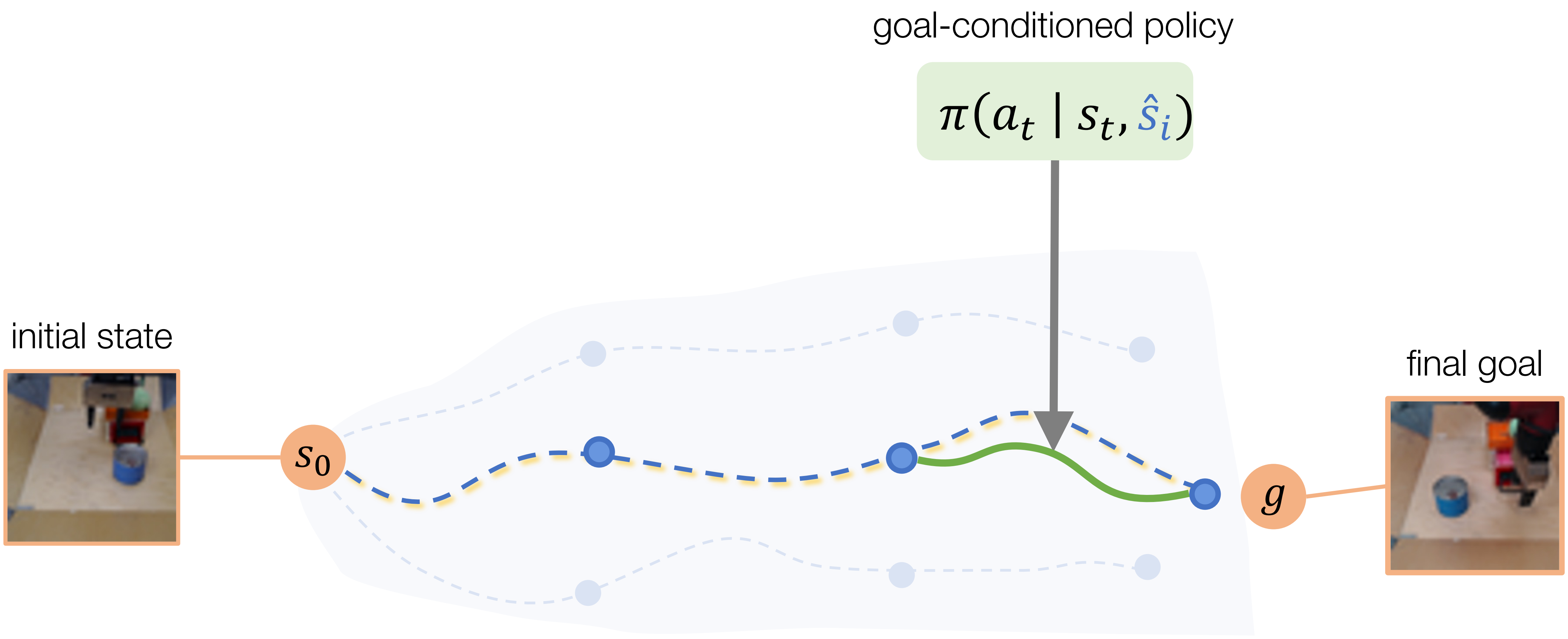
Select the optimal plan by:

$$u^* = \underset{u}{\operatorname{argmin}} \underbrace{\|g - \hat{s}_K\|}_{\text{If the final goal is reached}} - \eta \sum_i \underbrace{\log p(u_i) + V(\hat{s}_{i-1}, \hat{s}_i)}_{\text{If the each subgoal is feasible}}$$



Fang*, Yin*, Nair, Levine. IROS 2022

Recursive generation for subgoal planning



Fang*, Yin*, Nair, Levine. IROS 2022

Comparisons with alternative subgoal generation methods

Recursive generation

PTP (ours)

initial state

$P($  $|$  $)$

Interpolative generation

GCP [Pertsch et al. 2020]

initial state final goal

$P($  $|$  $,$  $)$

Limited generalization

Unconditional generation

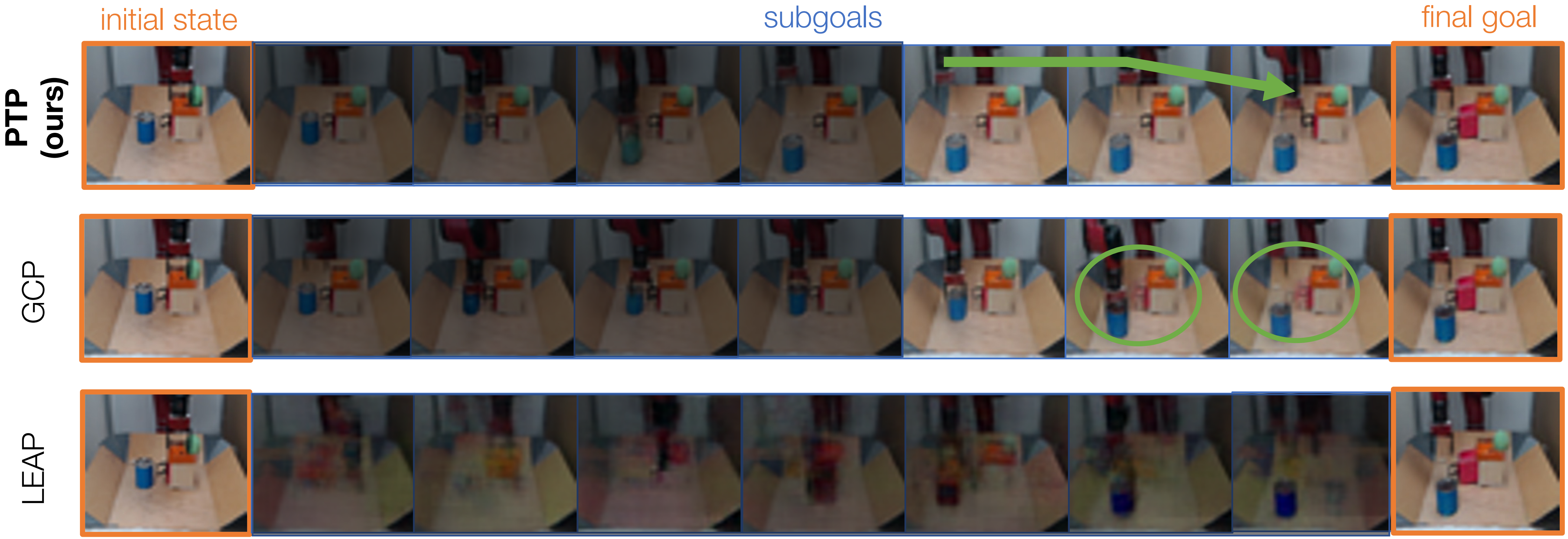
LEAP [Nasiriany et al. 2019]

$P($  $)$

Ignore contextual information

Generated subgoal sequences

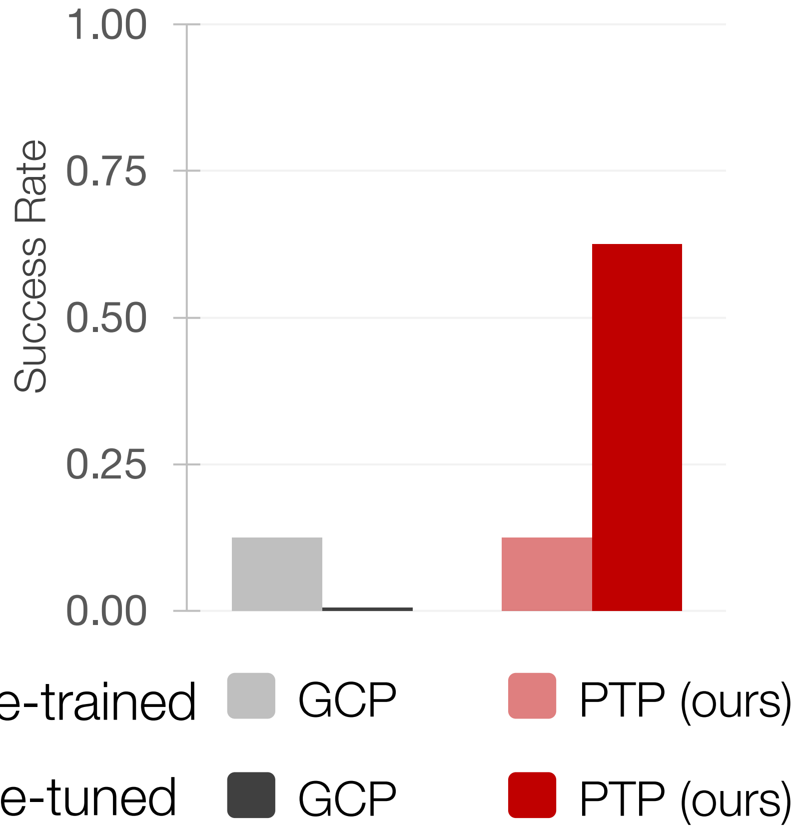
Recursive generation enables generalizable subgoal planning.

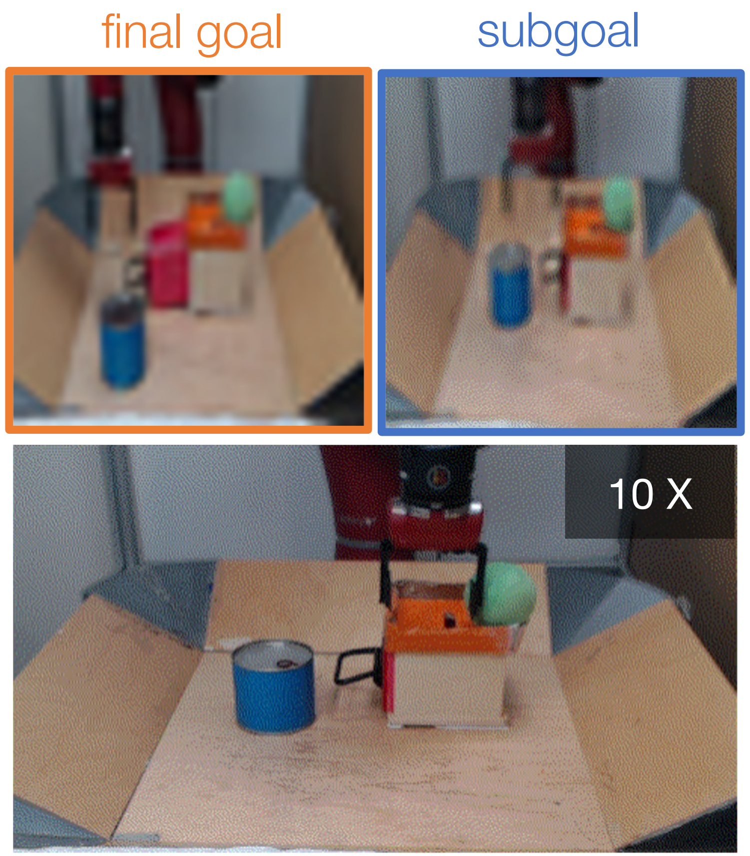


Fang*, Yin*, Nair, Levine. IROS 2022

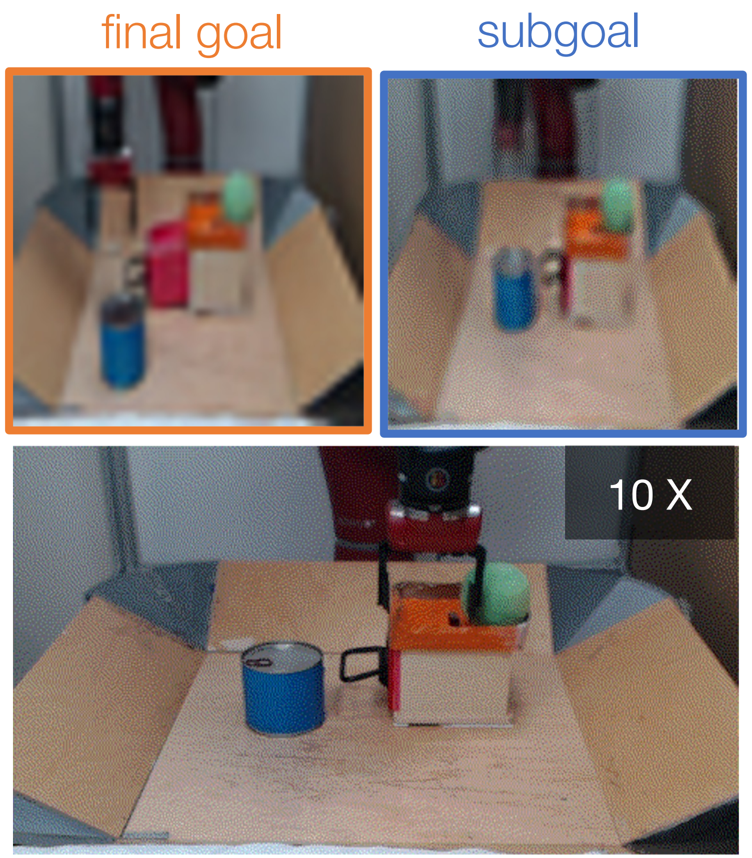
Solving sequential tasks using the fine-tuned skills

Fine-tuning for **80 episodes (3 hour)**





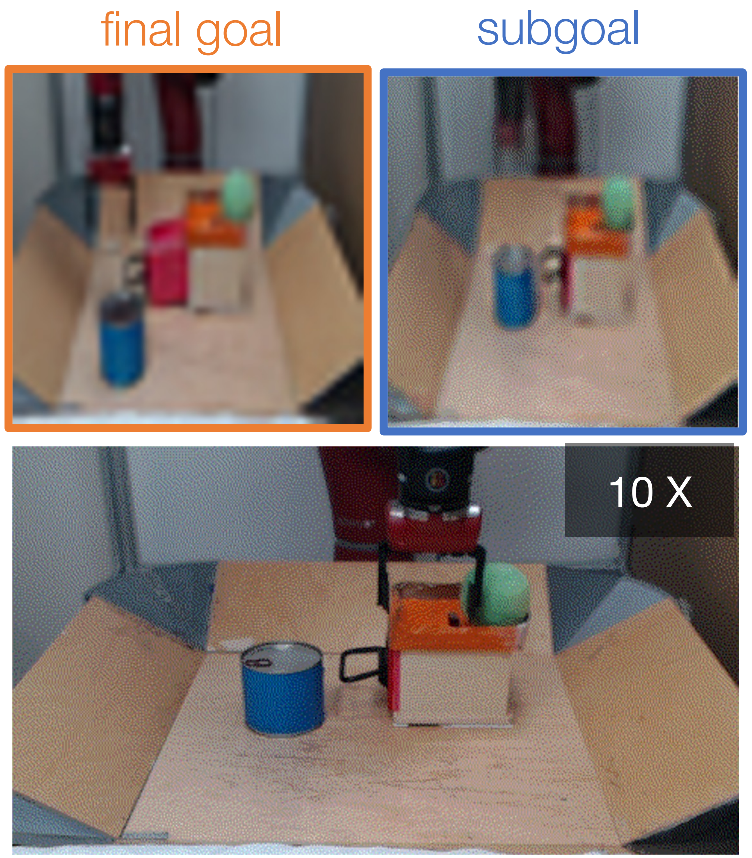
GCP
[Pertsch et al. NeurIPS 2020]



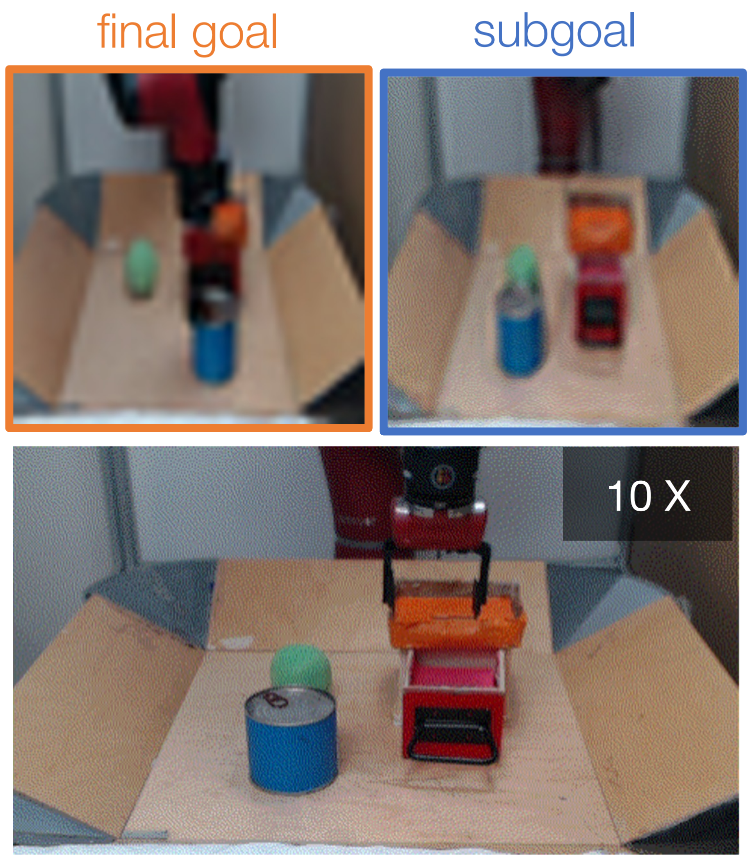
PTP
(Ours)

Fang*, Yin*, Nair, Levine. IROS 2022

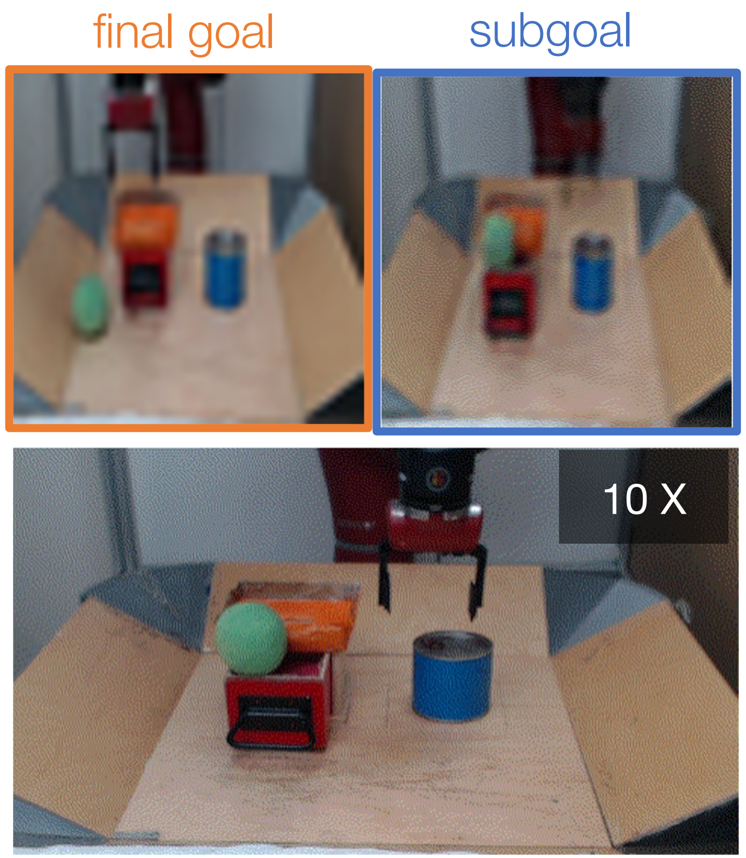
Solving long-horizon tasks using the fine-tuned skills



Move away the can and then open the drawer.



Close the drawer then move the can in front of it.



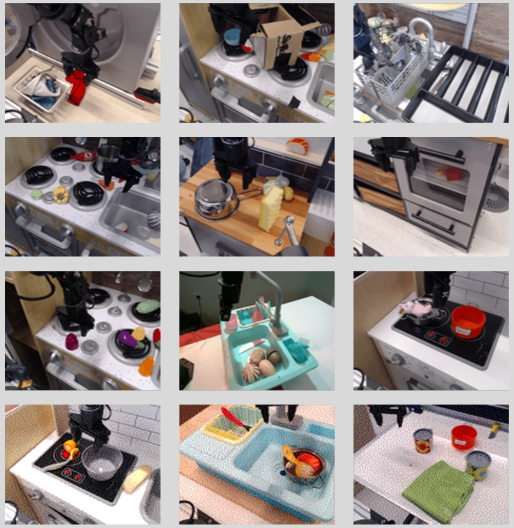
Poke the ball out of the drawer and then close it

Fang*, Yin*, Nair, Levine. IROS 2022

Leveraging broad prior experiences across tasks and environments

12k trajectories collected through teleoperation

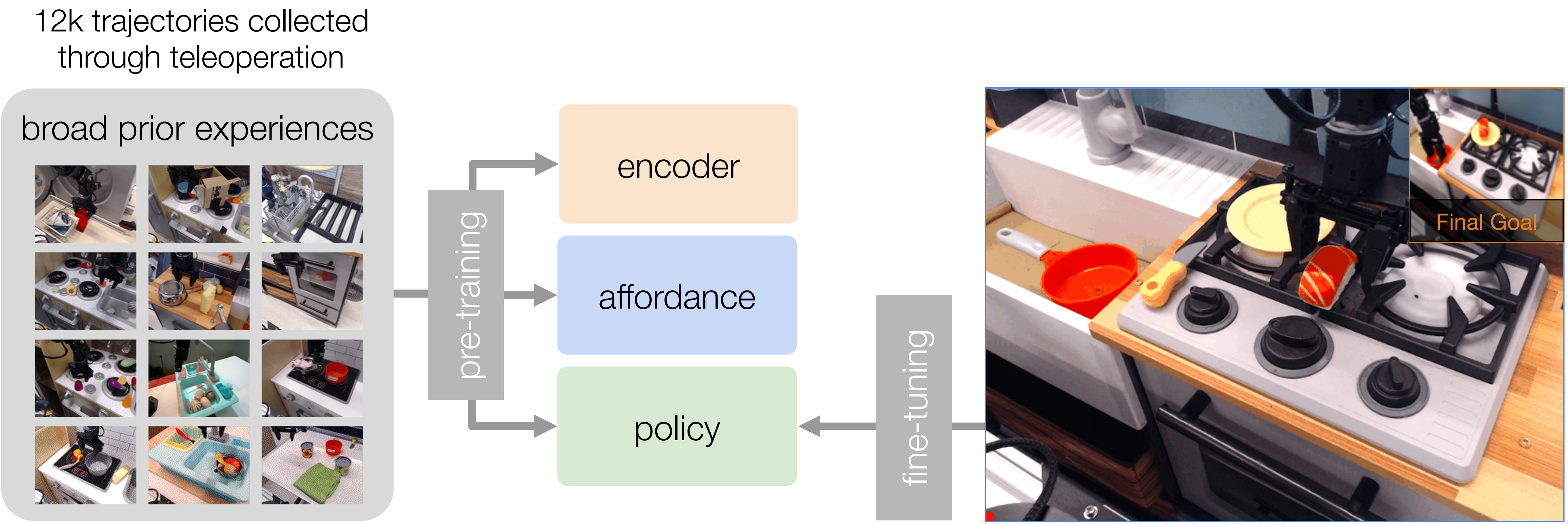
broad prior experiences



adaptation

Fang, Yin, Nair, Walke, Yan, Levine. CoRL 2022

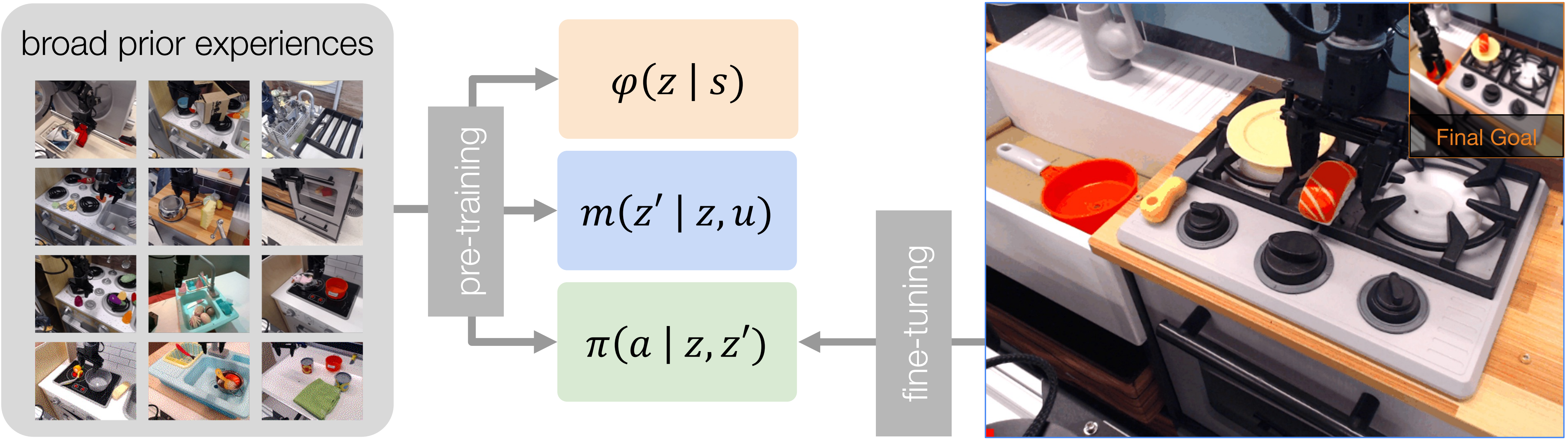
Fine-tuning with Lossy Affordance Planner (FLAP)



Fang, Yin, Nair, Walke, Yan, Levine. CoRL 2022

Fine-tuning with Lossy Affordance Planner (FLAP)

The encoder learns to project the initial states and the final goal.



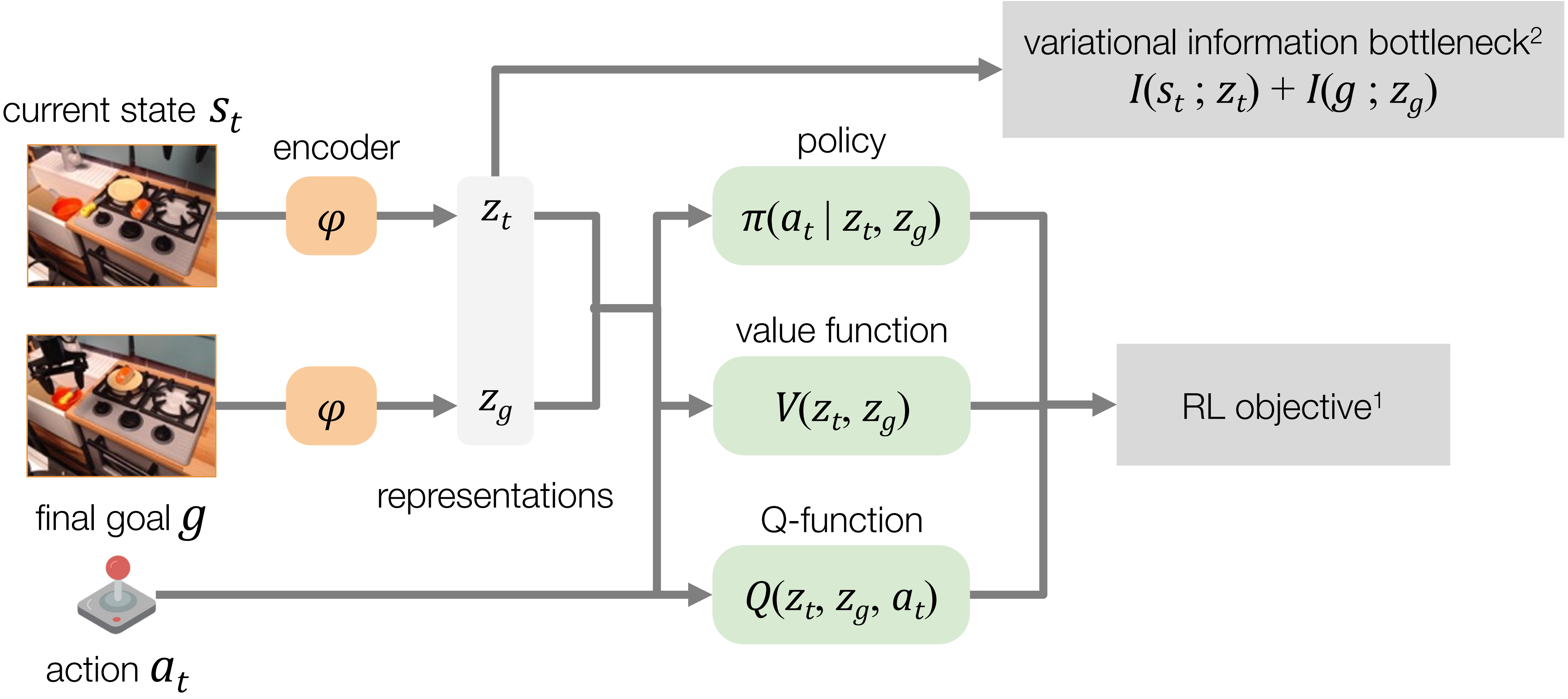
Both the affordance model and policy are defined in the lossy representation space.

Fang, Yin, Nair, Walke, Yan, Levine. CoRL 2022

Pre-training lossy representations

Highlight task-relevant information.

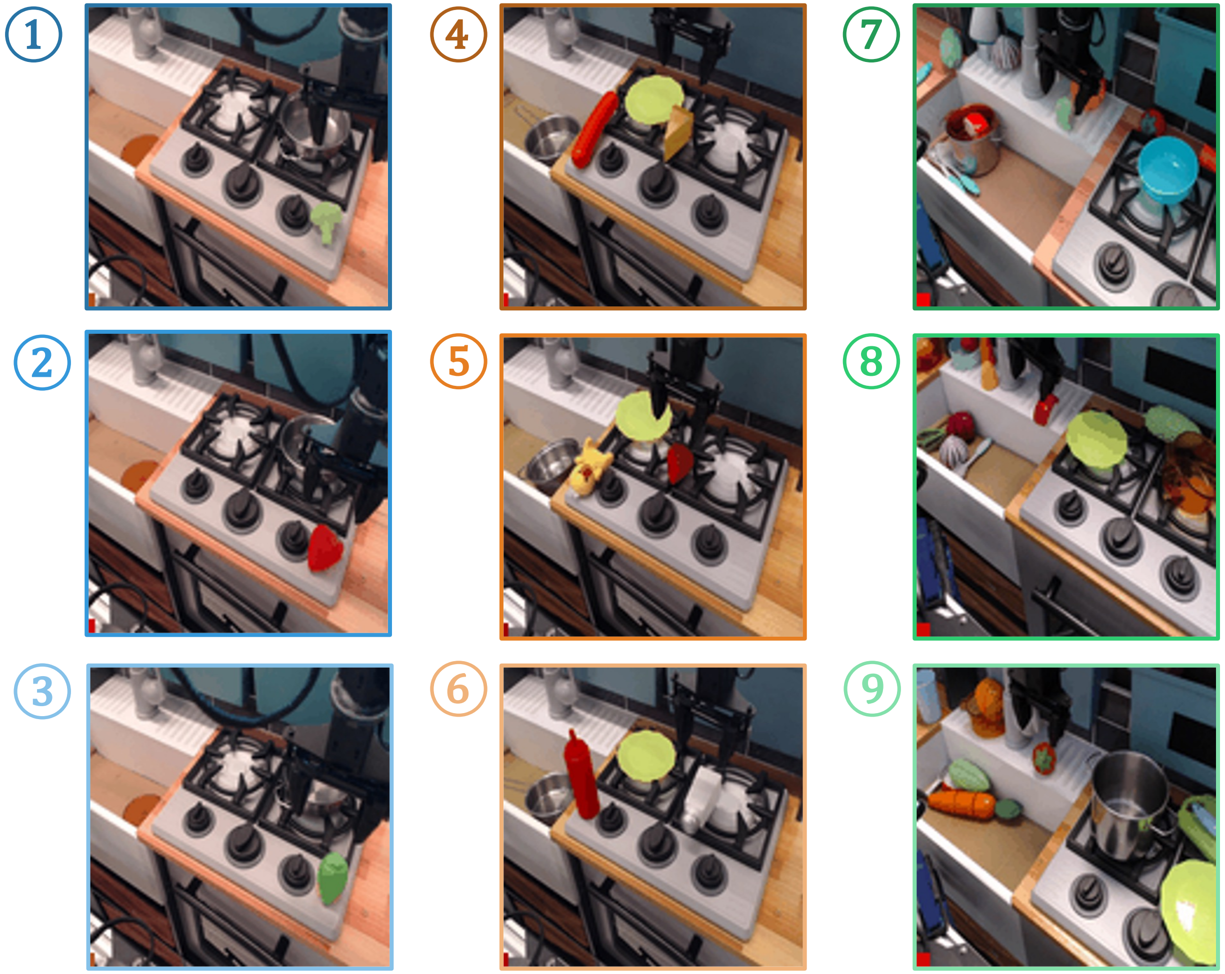
Remove environment-specific details



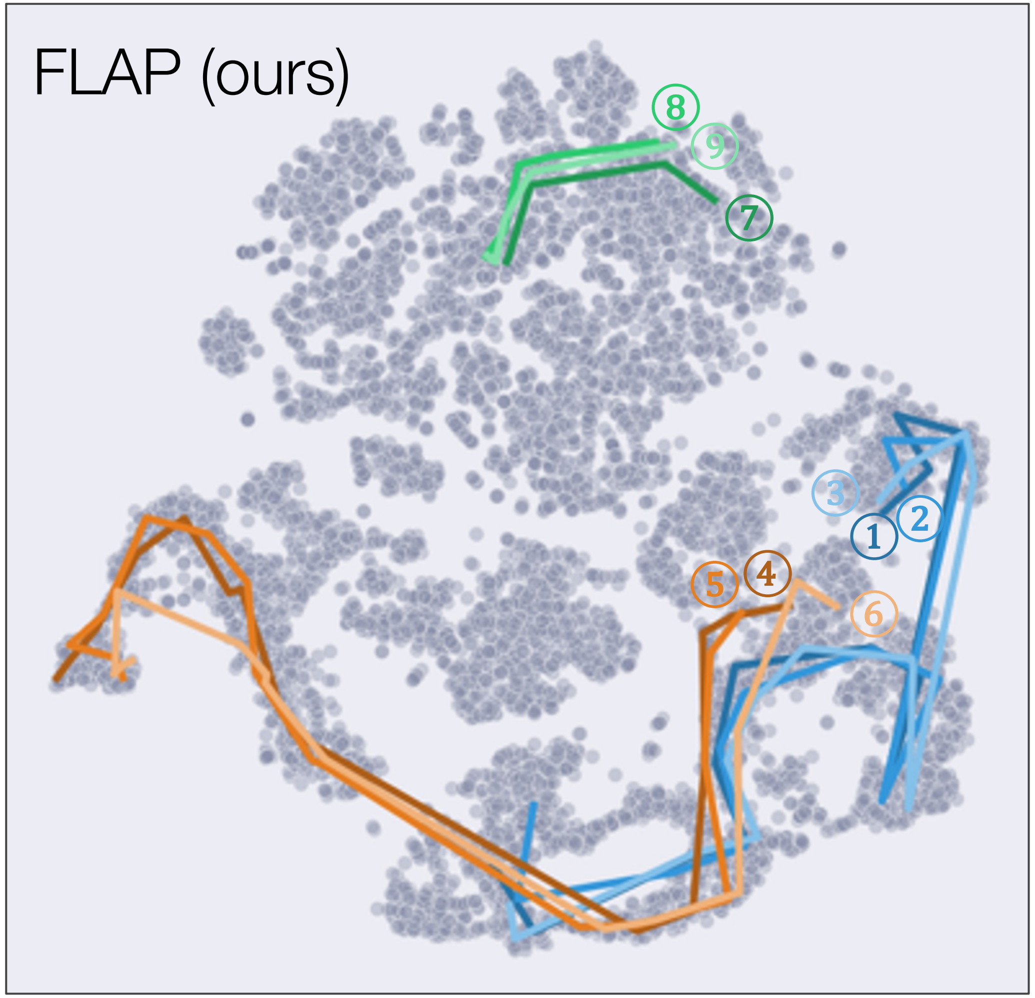
[1][Kostrikov et al. 2021]; [2][Alemi et al. 2016]

Fang, Yin, Nair, Walke, Yan, Levine. CoRL 2022

Learned lossy representations



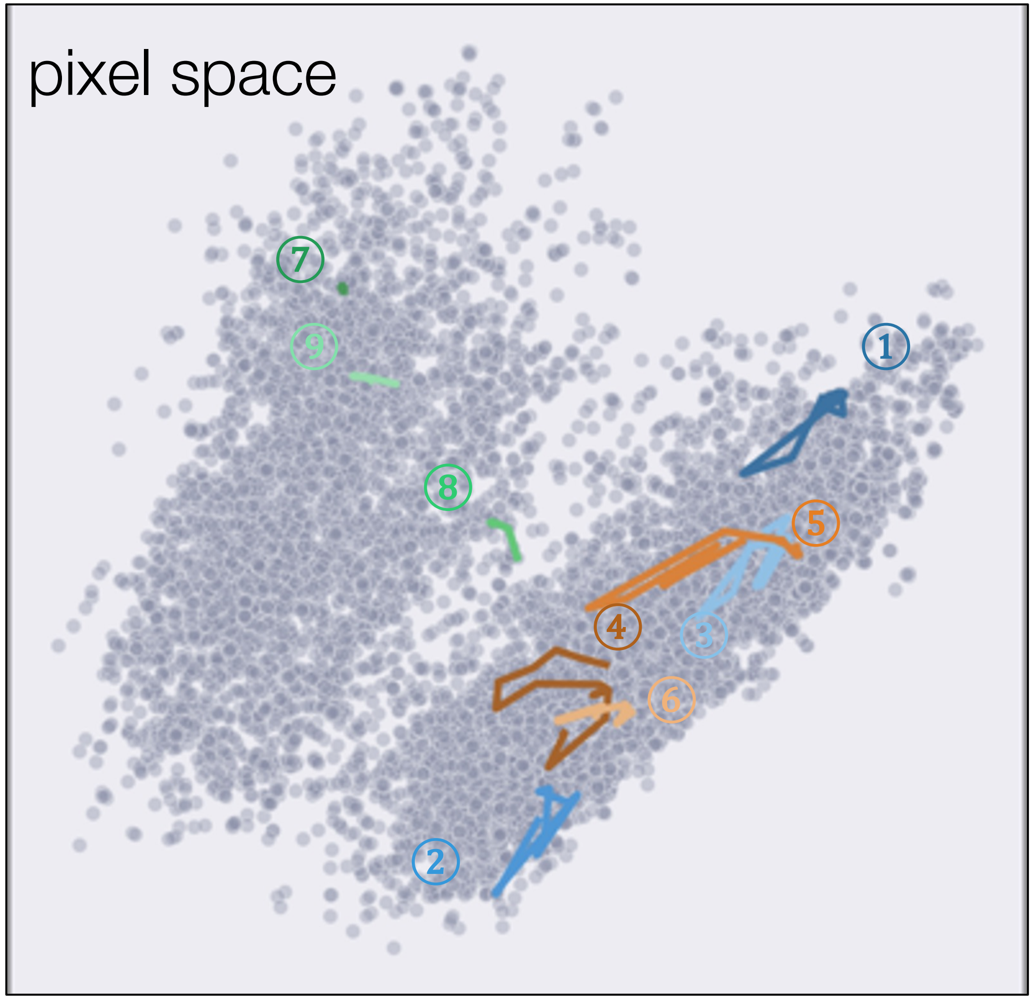
Visualize trajectories using t-SNE[1].

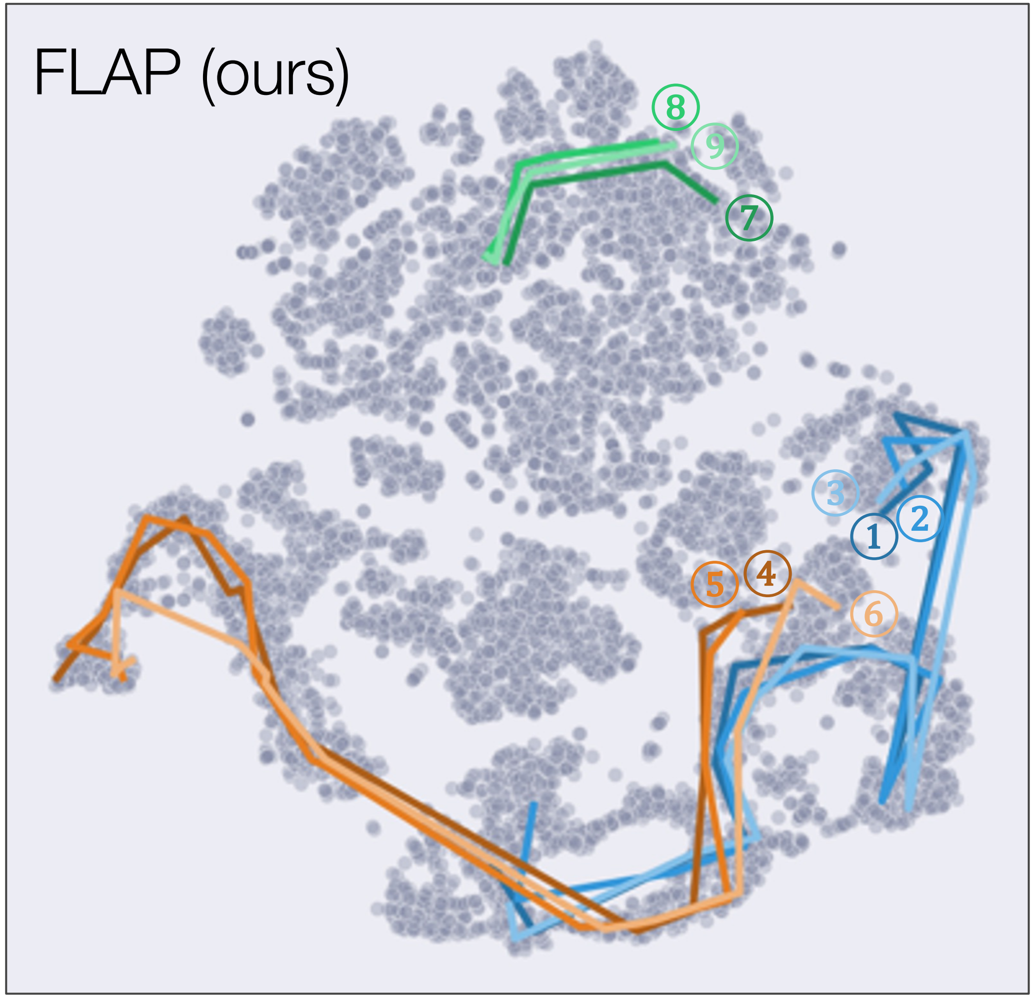


[1][van der Maaten and Hinton 2008]

Fang, Yin, Nair, Walke, Yan, Levine. CoRL 2022

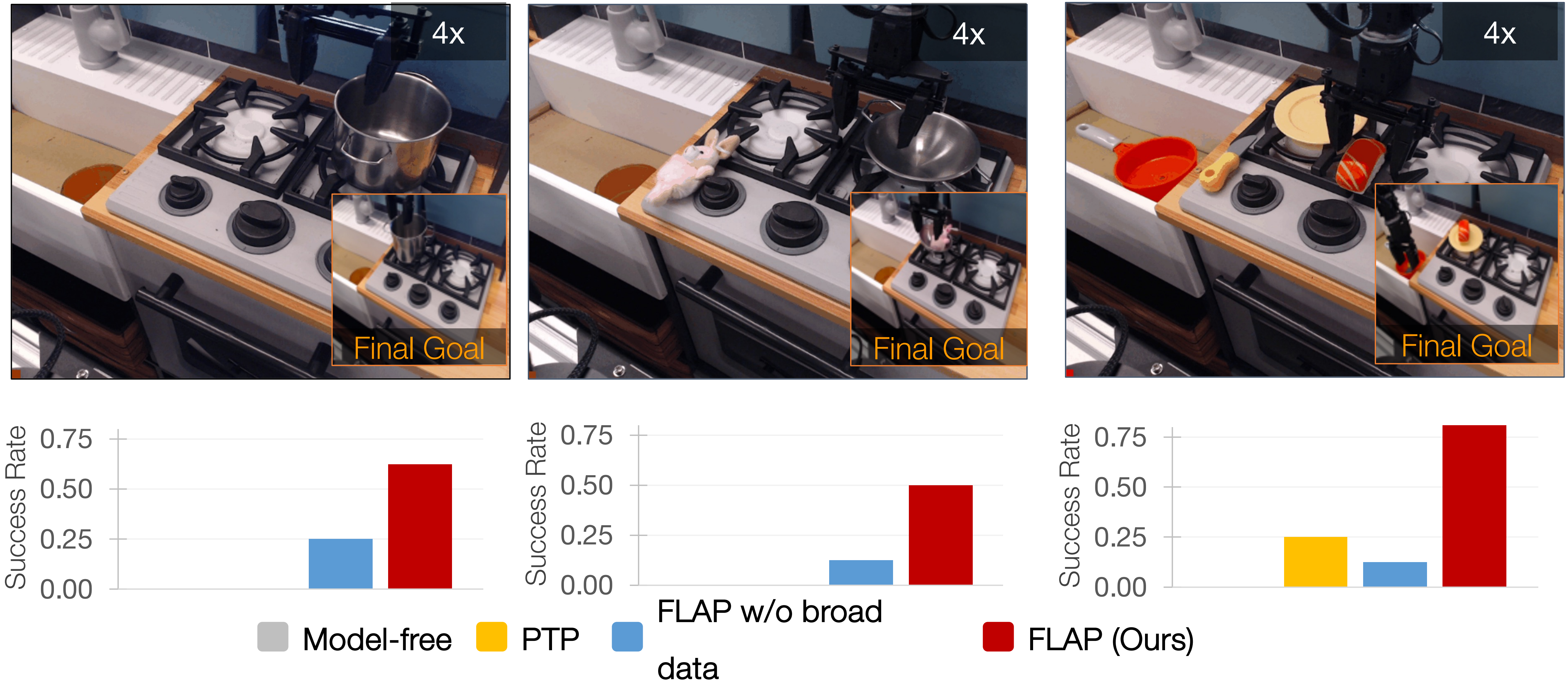
Learned lossy representations





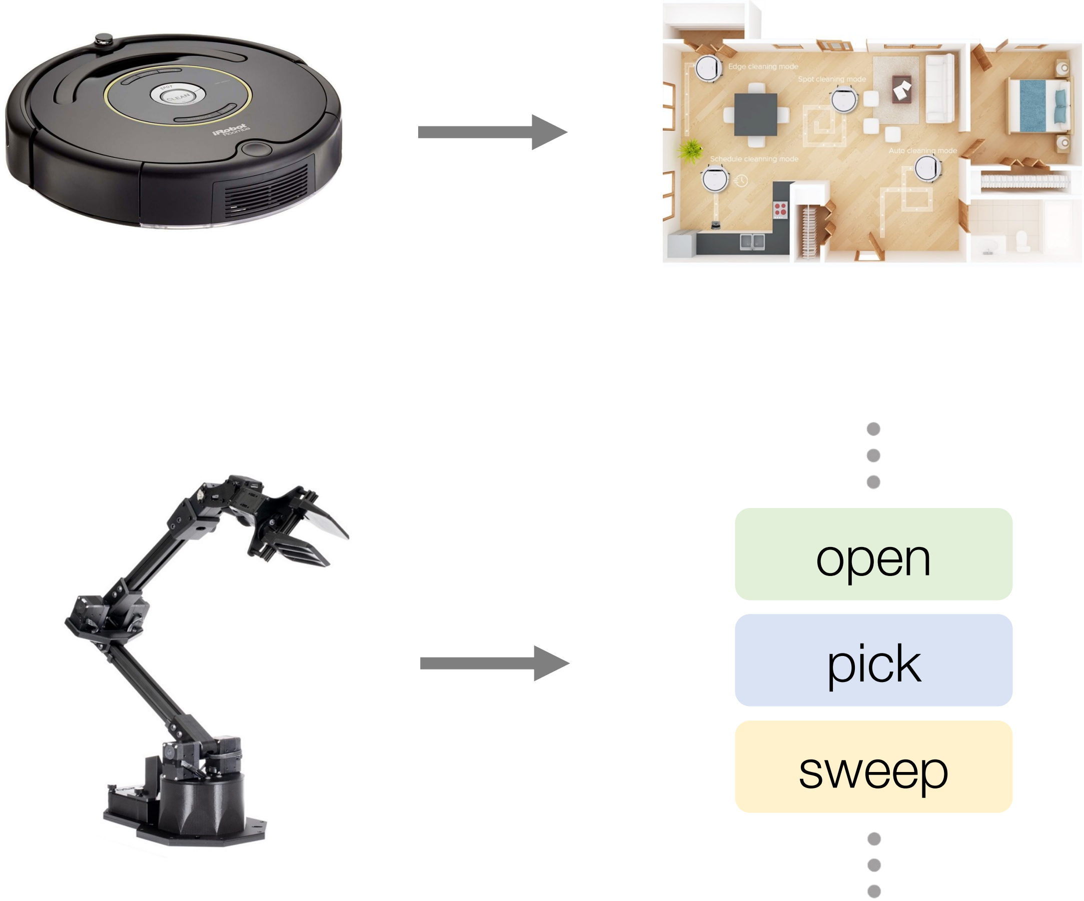
Fang, Yin, Nair, Walke, Yan, Levine. CoRL 2022

Fast adaptation to novel tasks in the target environments



Fang, Yin, Nair, Walke, Yan, Levine. CoRL 2022

Discovering skills in novel environments



Discovering skills in novel environments

Challenge: how to discover **diverse** and **useful** skills in unknown environments?

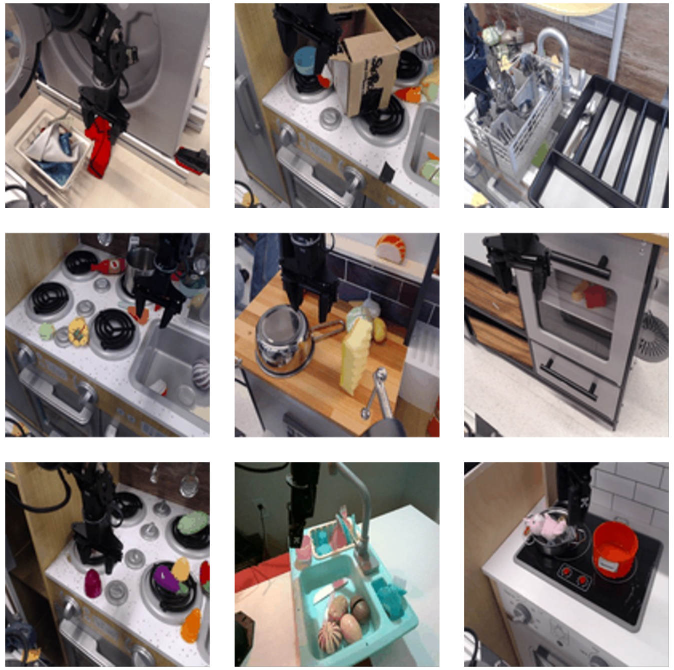
Repeating the same skill

Attempting useless behaviors

Diverse and useful skills
(What we want)

Goal-directed exploration by leveraging broad prior experiences

pre-training with
broad prior experiences

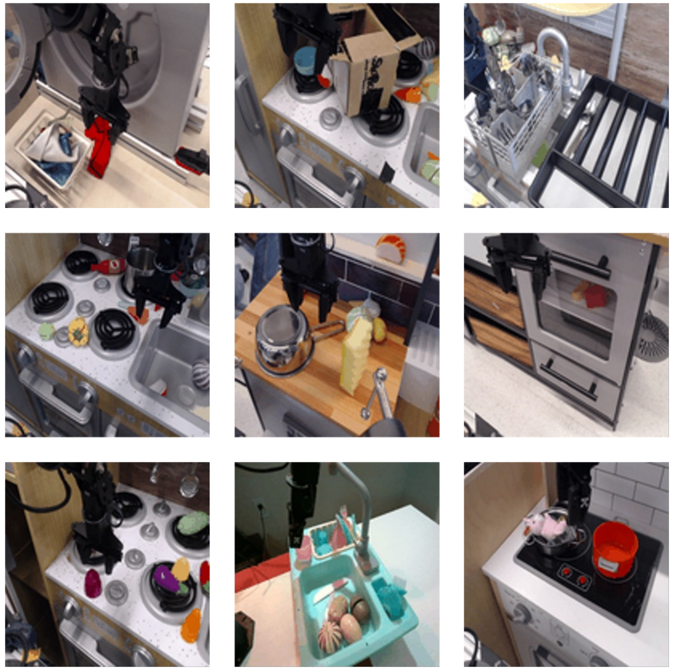


current image

goal image

Goal-directed exploration by leveraging broad prior experiences

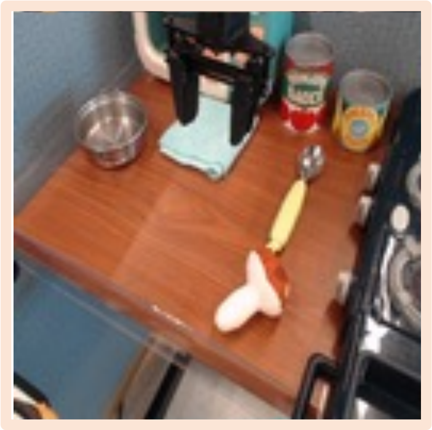
pre-training with
broad prior experiences



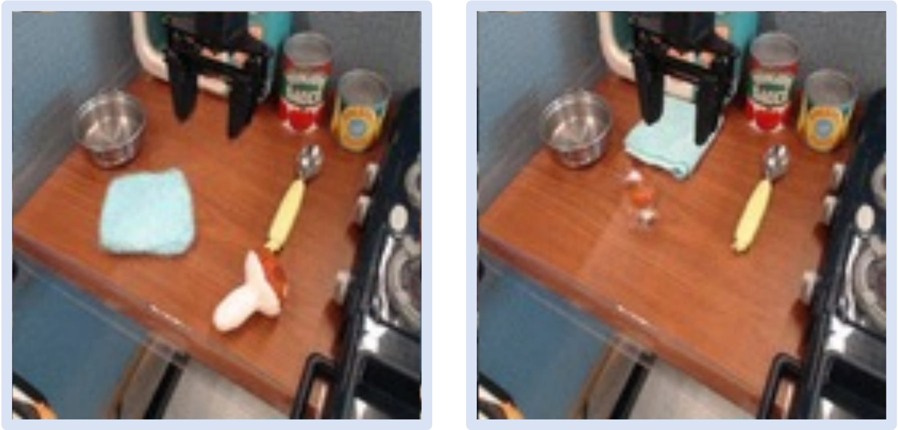
conditional goal generation using diffusion models

[Sohl-Dickstein et al. ICML 2015; Song et al. NeurIPS 2019; Ho et al. NeurIPS 2020]

current image



generated goals



Exploration with generated goals

Autonomously collecting 1,000 trajectories (10 hours) directed by generated goals.

Discovered skills in novel environments

Move the mushroom into the pot

success rate: 6.7% → 86.7%

Move the pot to the corner

success rate: 13.3% → 73.3%

Sweep the table with the cloth

success rate: 53.3% → 80.0%

Solving sequential tasks using the discovered skills

Subgoal 1 Subgoal 2 Subgoal 3

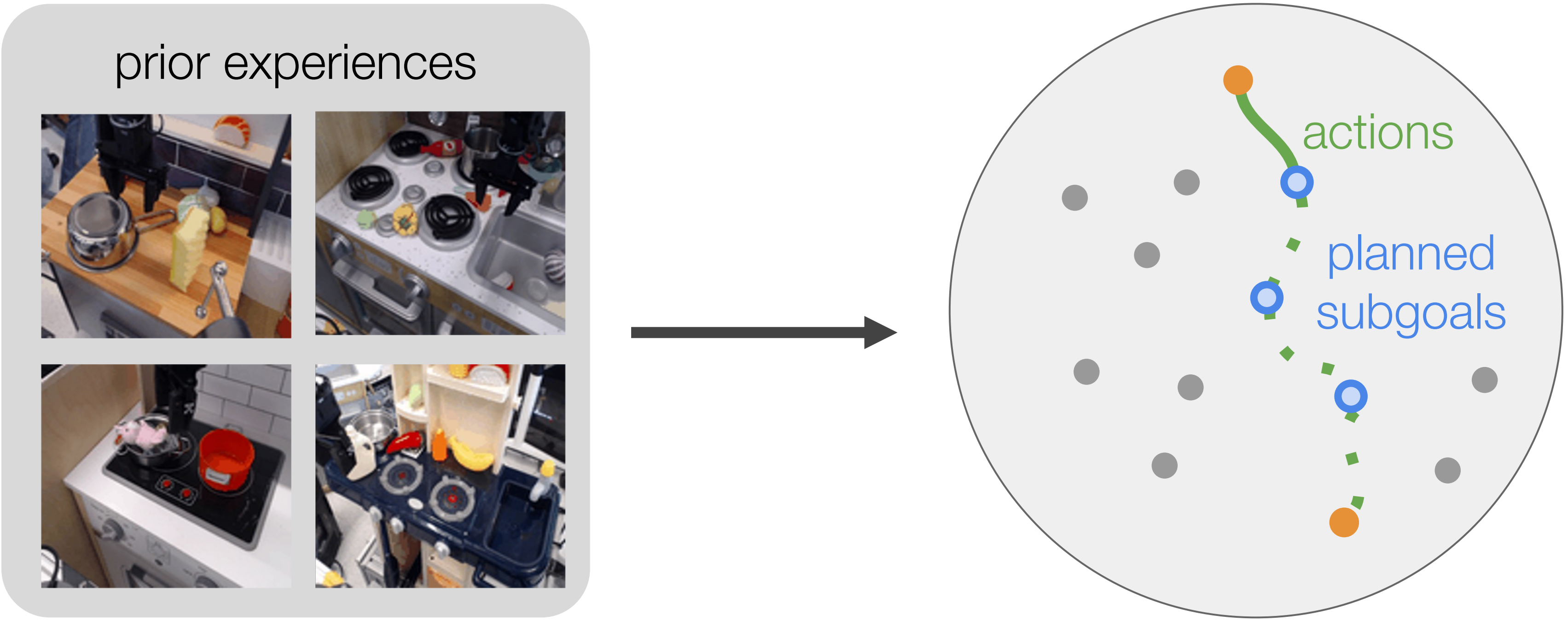
Setting up the table

Subgoal 1 Subgoal 2 Subgoal 3

Cleaning up the table

Summary

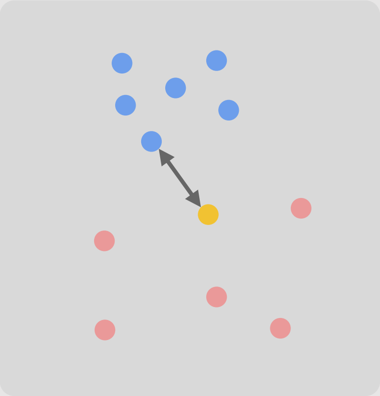
- Compose and adapt prior skills using recursively generated subgoals.
- Generalize across environments using learned lossy representations.

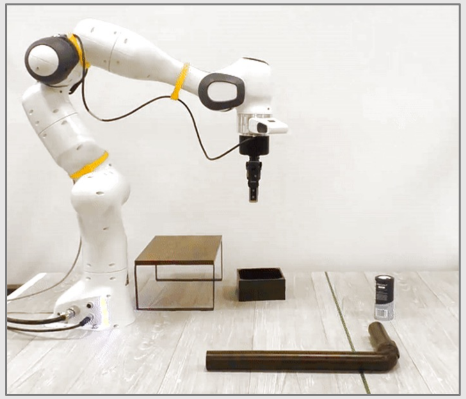


Generalization via Generation:

Learning Long-Horizon Tasks with Limited Supervision

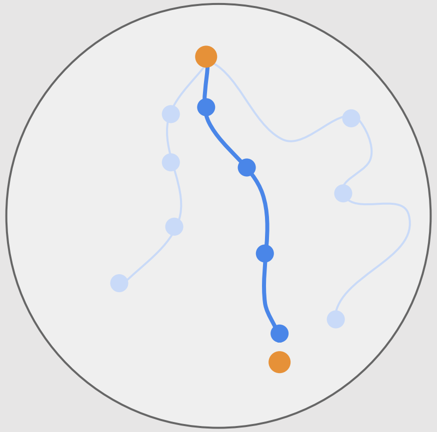
Learning Robust Skill via Environment Generation





[**Fang** et al. IJRR 2019]
[**Fang***, Migimatsu* et al. 2023]

Adapting Prior Skills via Goal Generation



[**Fang** et al. CoRL 2019]
[**Fang***, Yin* et al. IROS 2022]
[**Fang** et al. CoRL 2022]

Acknowledgement


Fei-Fei Li


Silvio Savarese


Sergey Levine

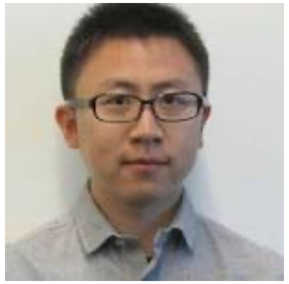

Jeannette Bohg


Animesh Garg


Yuke Zhu


Joseph Lim


Mrinal Kalakrishnan


Yunfei Bai


Stefan Hinterstoisser


Alexander Toshev


Andrey Kurenkov

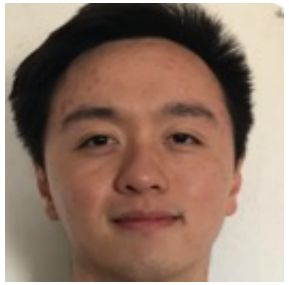

Viraj Mehta


Ajay Mandelekar


Ashvin Nair


Homer Walke


Toki Migimatsu


Patrick Yin


Gengchen Yan

Acknowledgement

Fei-Fei Li

Silvio Savarese

Sergey Levine

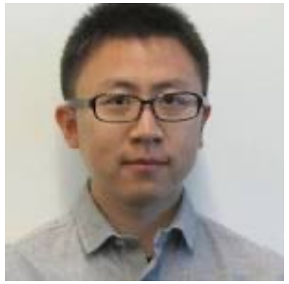
Jeannette Bohg

Animesh Garg

Yuke Zhu

Joseph Lim

Mrinal Kalakrishnan



Yunfei Bai

Stefan Hinterstoisser

Alexander Toshev

Andrey Kurenkov

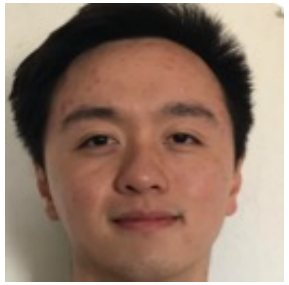
Viraj Mehta

Ajay Mandelekar

Ashvin Nair

Homer Walke

Toki Migimatsu



Patrick Yin

Gengchen Yan

Opportunities at

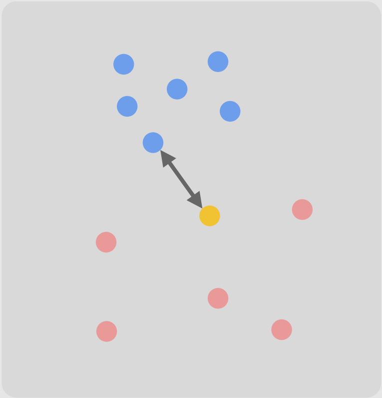
Cornell Bowers C·IS
Computer Science

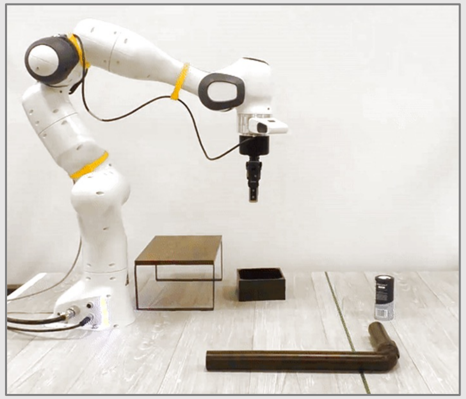
18+ robotics faculty across departments
continuously growing

Generalization through Generation:

Learning Long-Horizon Tasks with Limited Supervision

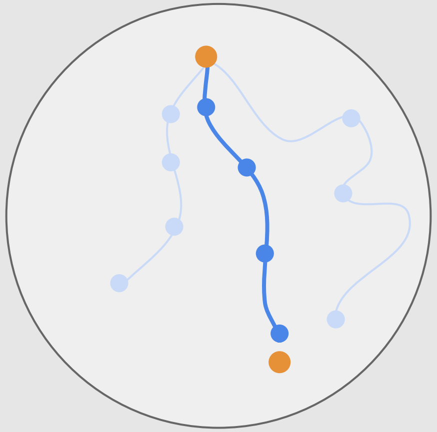
Learning Robust Skill via Environment Generation





[**Fang** et al. IJRR 2019]
[**Fang***, Migimatsu* et al. 2023]

Adapting Prior Skills via Goal Generation



[**Fang** et al. CoRL 2019]
[**Fang***, Yin* et al. IROS 2022]
[**Fang** et al. CoRL 2022]

Questions?