Scaling Robot Learning with Skills

Towards Furniture Assembly and Beyond

Youngwoon Lee

USC

Dissertation Defense

October 27, 2022

Robots work great in factory!

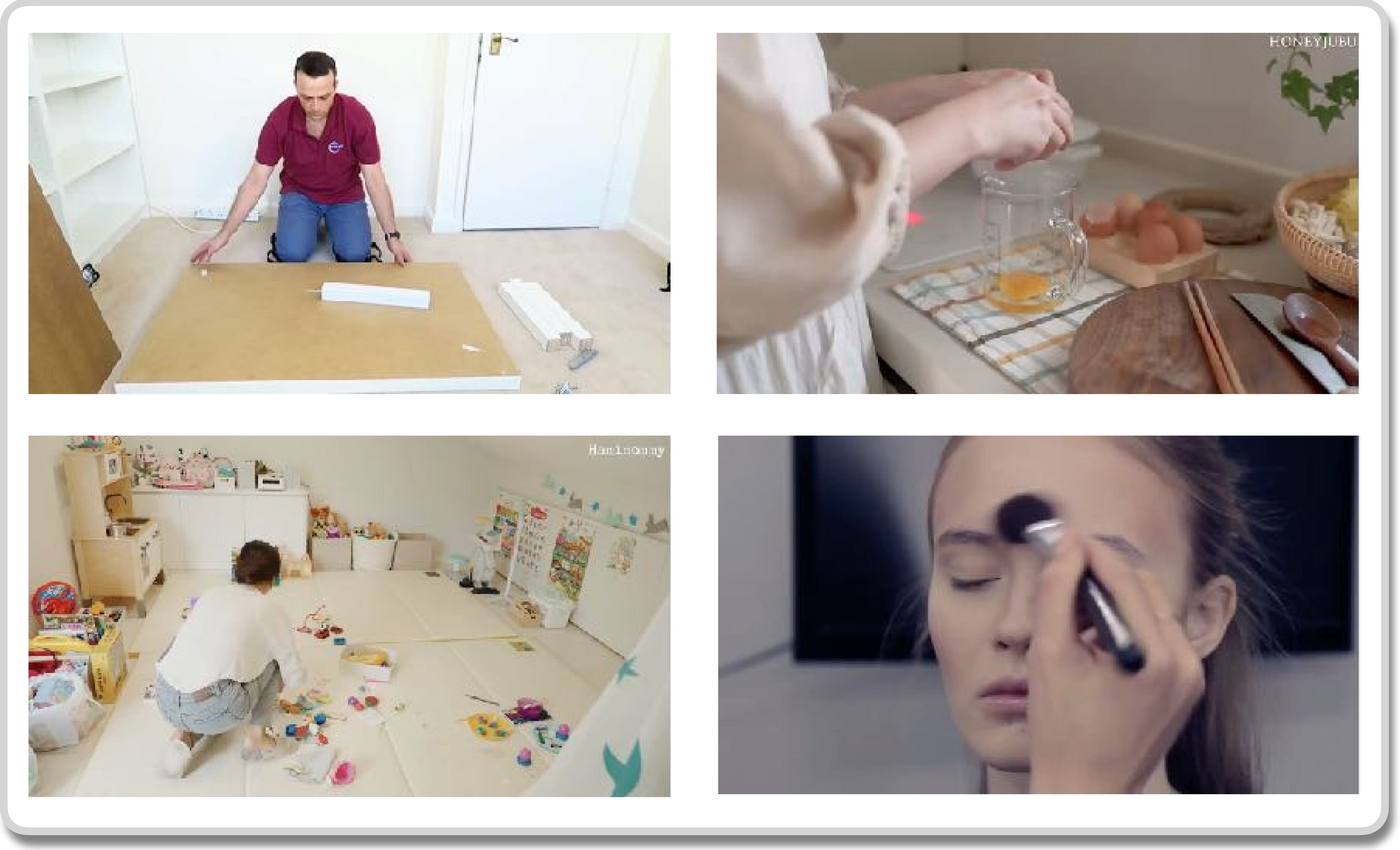
BMW Car Factory ROBOTS - Fast Manufacturing: https://www.youtube.com/watch?v=P7fi4hP_y80

Why not at home?

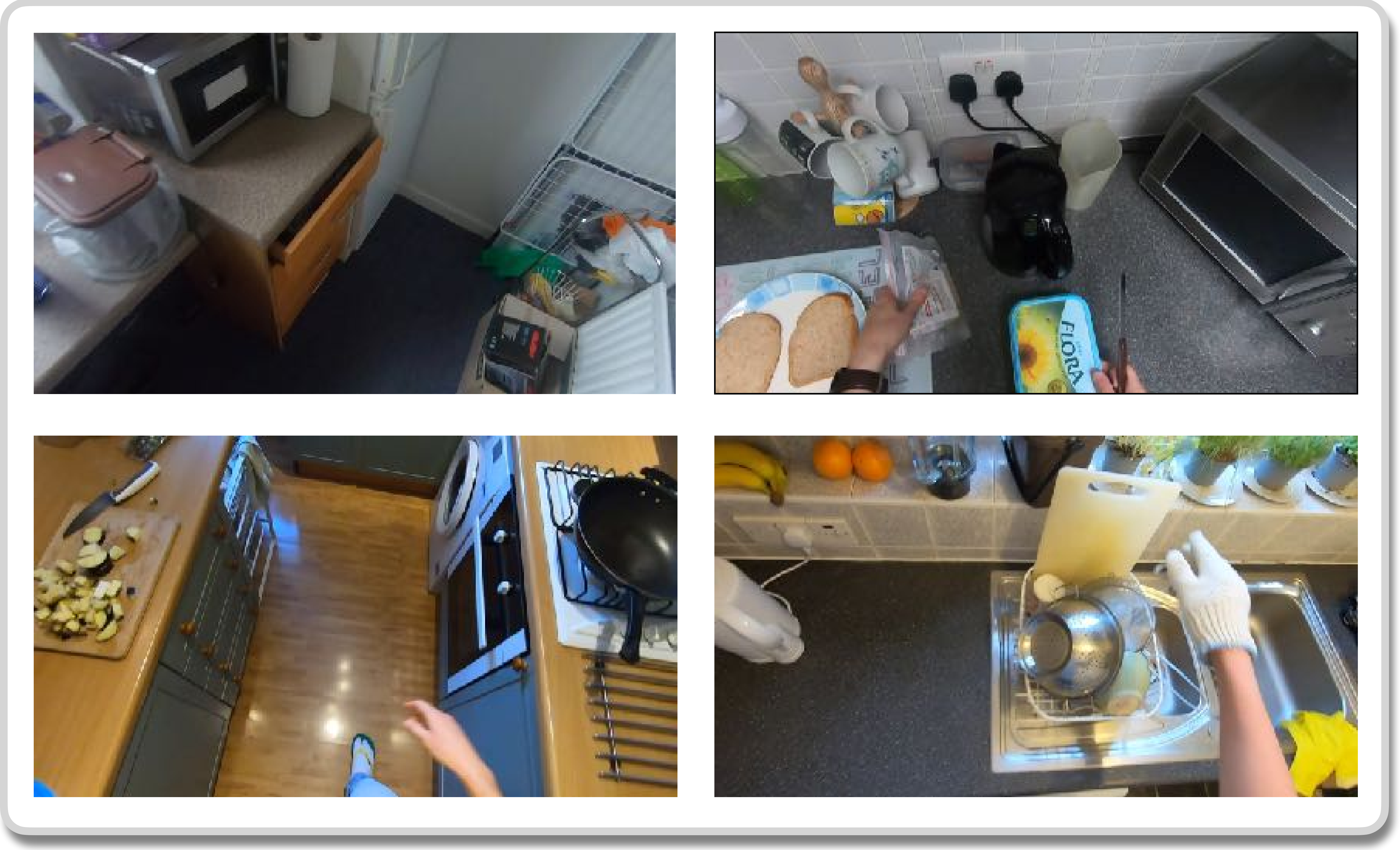
[CES 2021] Next Generation Robotics | Samsung: https://www.youtube.com/watch?v=qrPsa7JsPBU

Challenges

Complexity of tasks



Diversity of environments



Robot Learning — data-driven approach

If you know how to solve,

Imitation Learning (IL)

Learning from demonstrations

Requiring a lot of demonstrations

If you don't know how to solve,

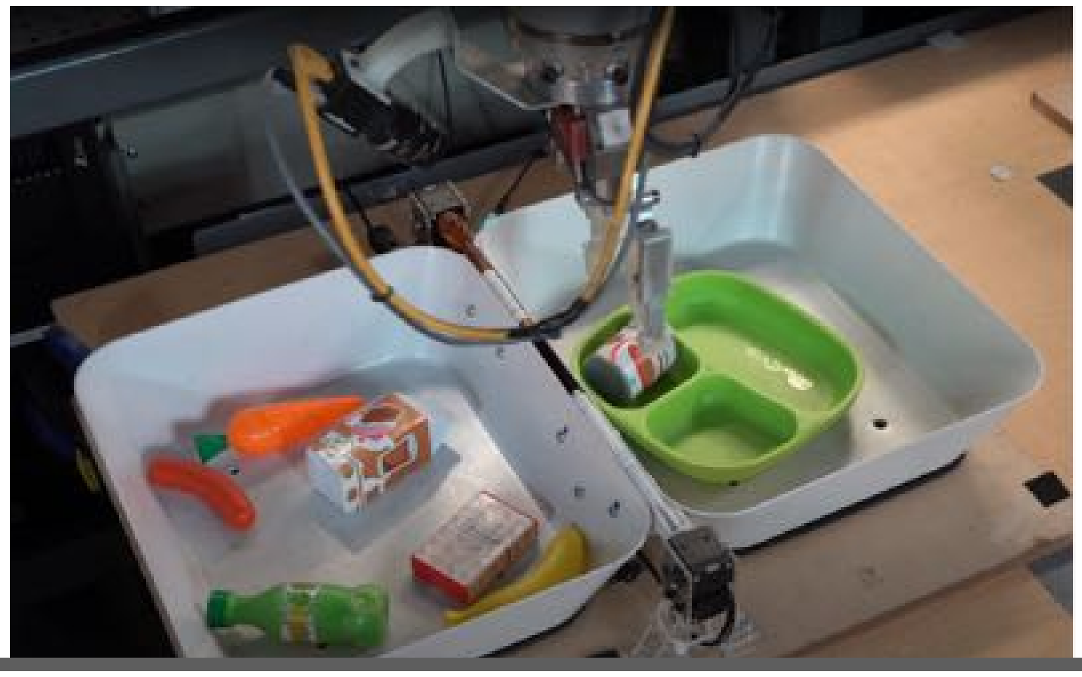
Reinforcement Learning (RL)

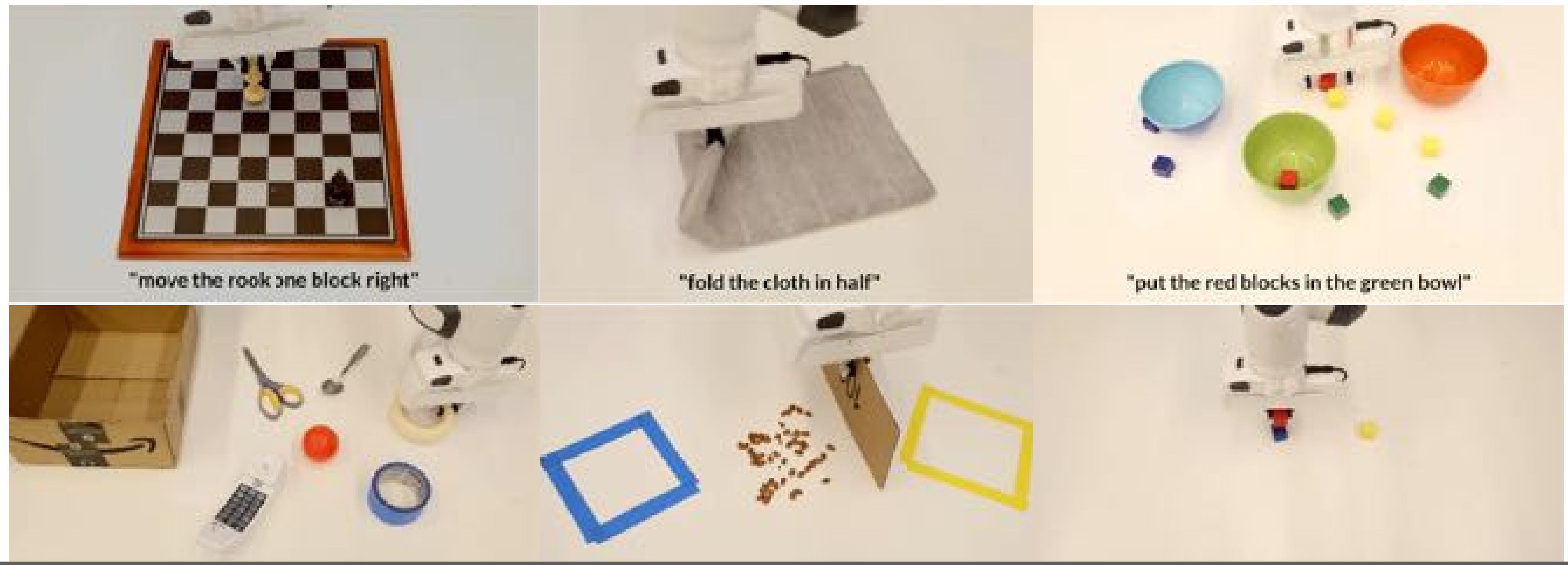
Learning through trial-and-error

Requiring a lot of trials

Credit: Yuke Zhu's slides

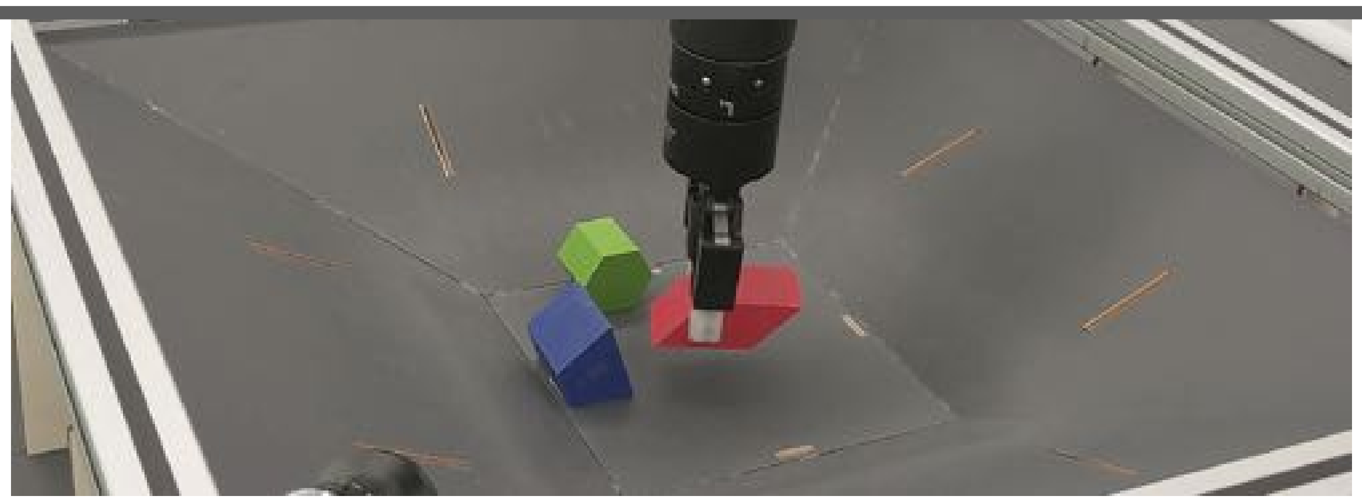
Robot Learning — data-driven approach





Limited to **simple** and **short** tasks in **controlled** environments

BC-Z [Jang et al. CoRL 2021]



RGB-Stack [Lee et al. CoRL 2021]

My research goal:

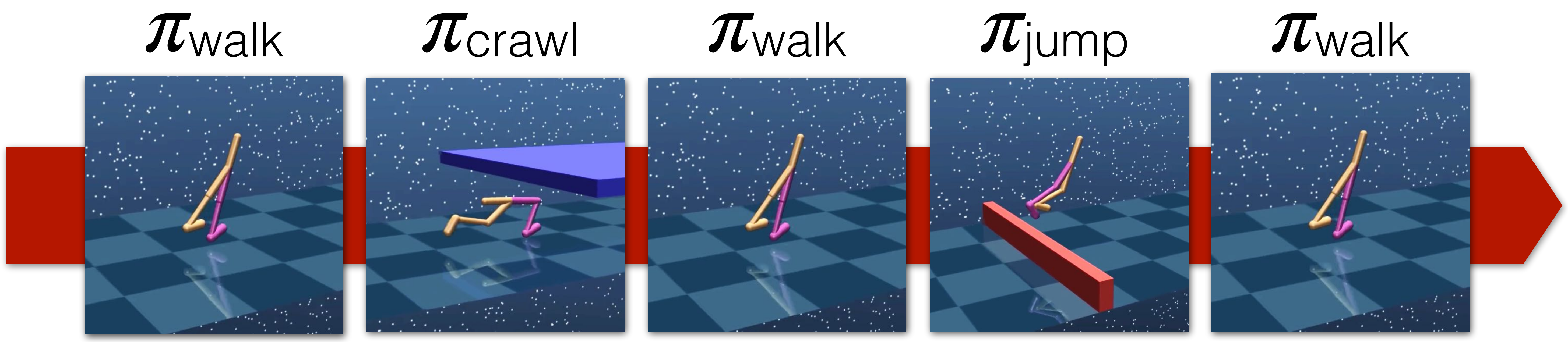
Scaling robot learning to real-world tasks

complex long-horizon real-world

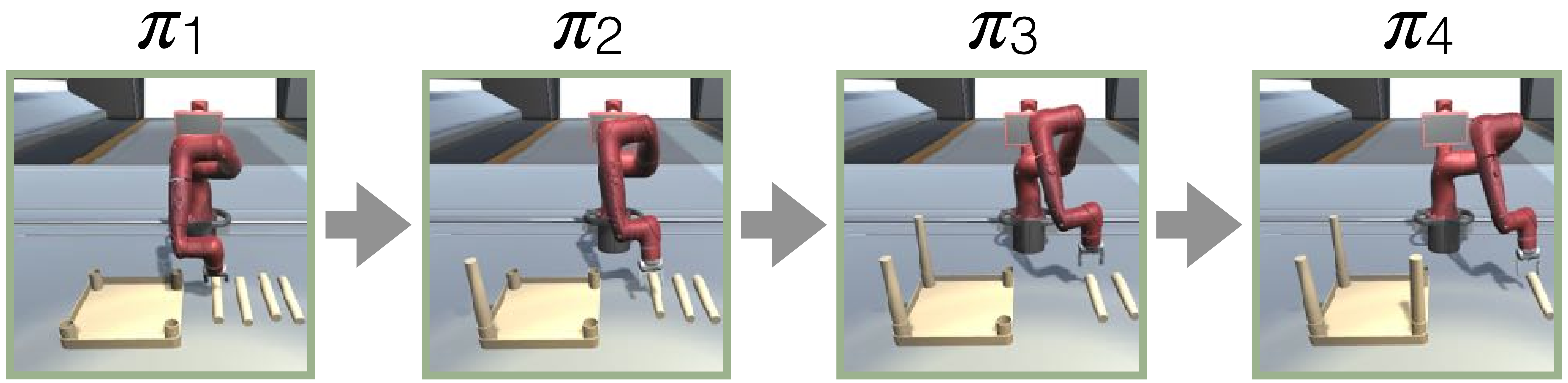
~~Limited to simple~~ and ~~short~~ tasks in ~~controlled~~ environments

Can RL solve complex long-horizon tasks?

Obstacle course

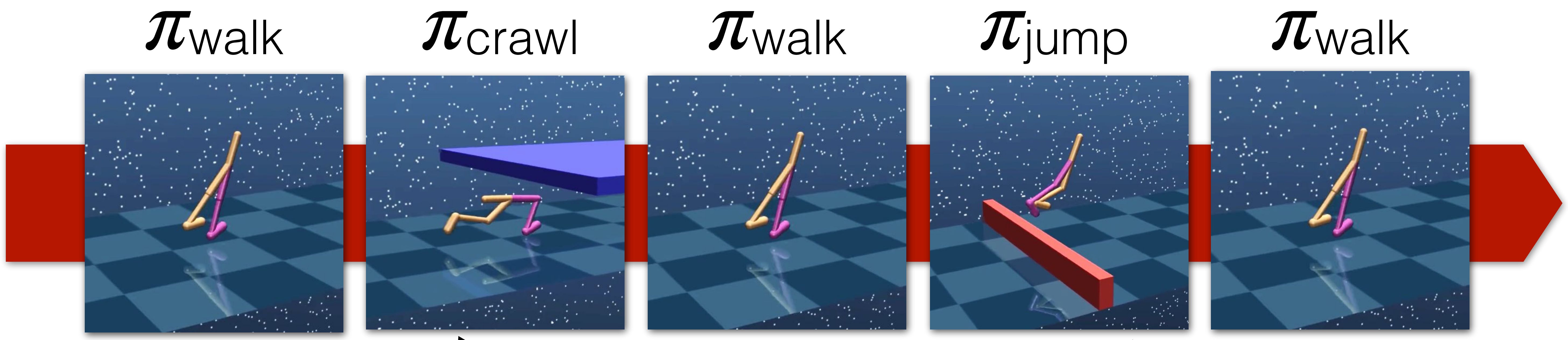


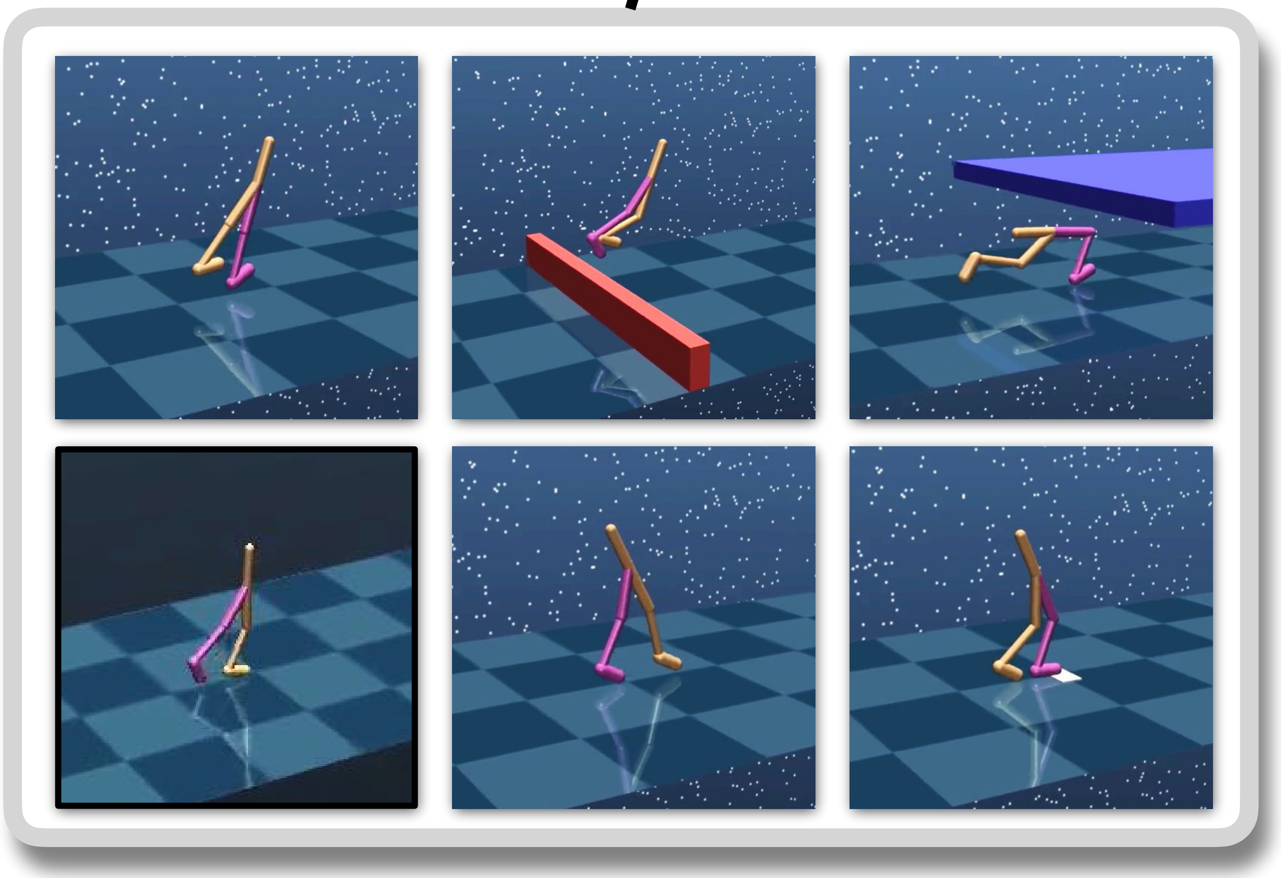
Furniture assembly



… sample efficiency 

Can RL solve complex long-horizon tasks?





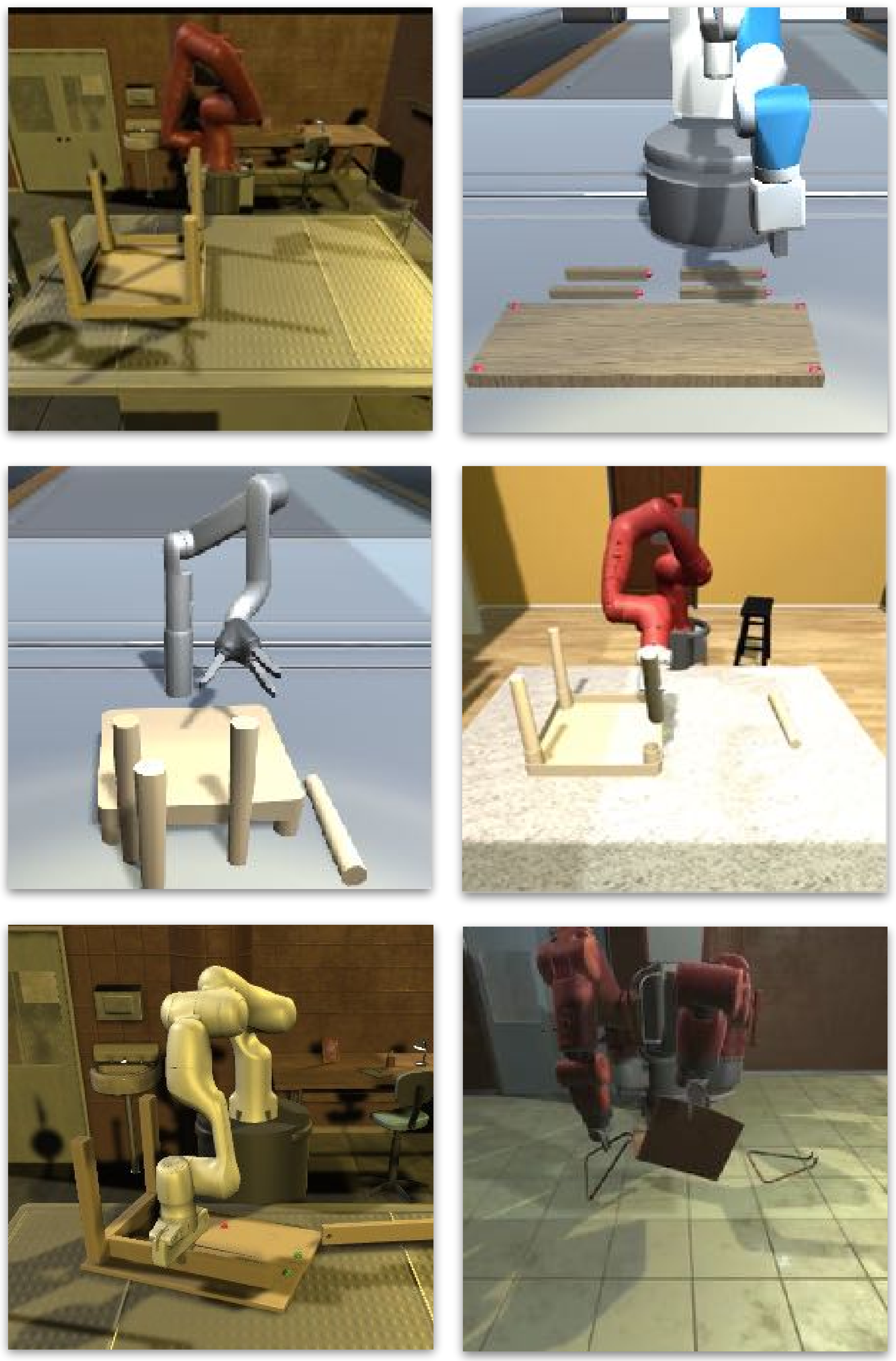
Skills

Temporally extended actions

Prior experience

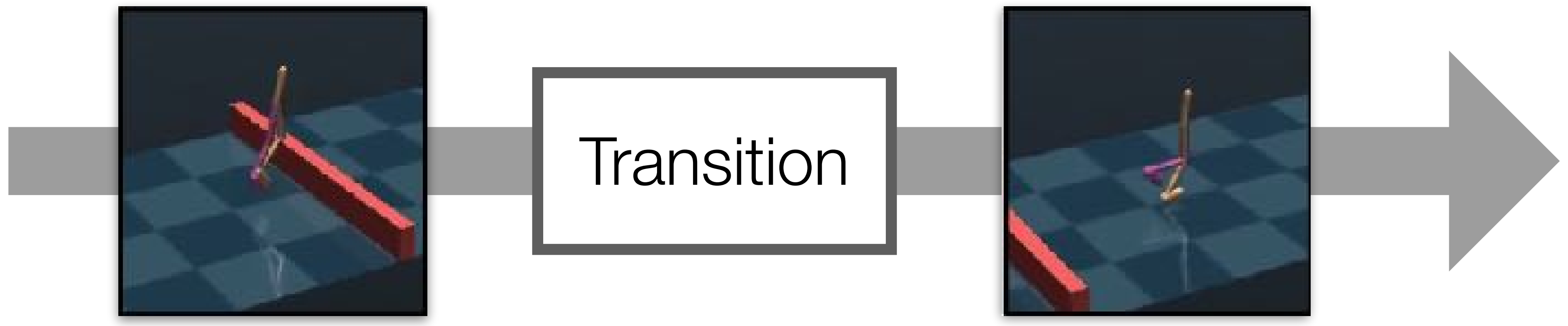
Similar experience, prior knowledge, etc

Scaling robot learning to long-horizon tasks

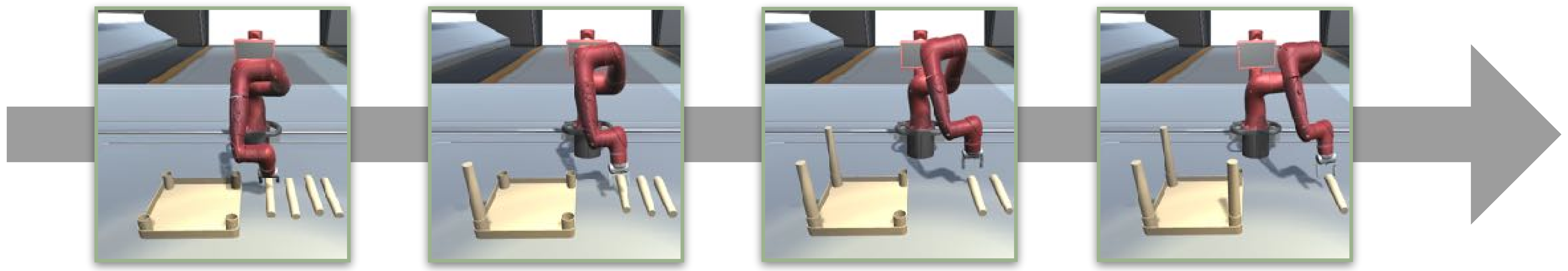


IKEA furniture assembly [ICRA’21]

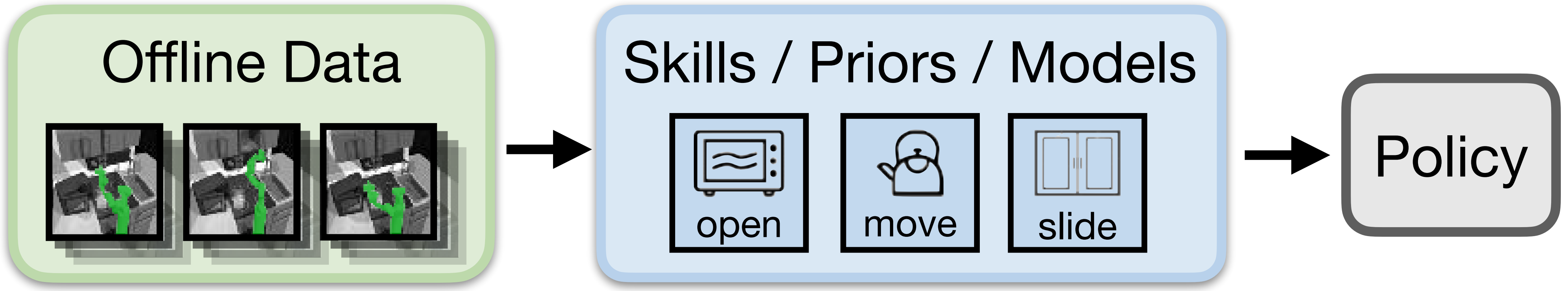
Benchmark



Transition policy [ICLR’19]



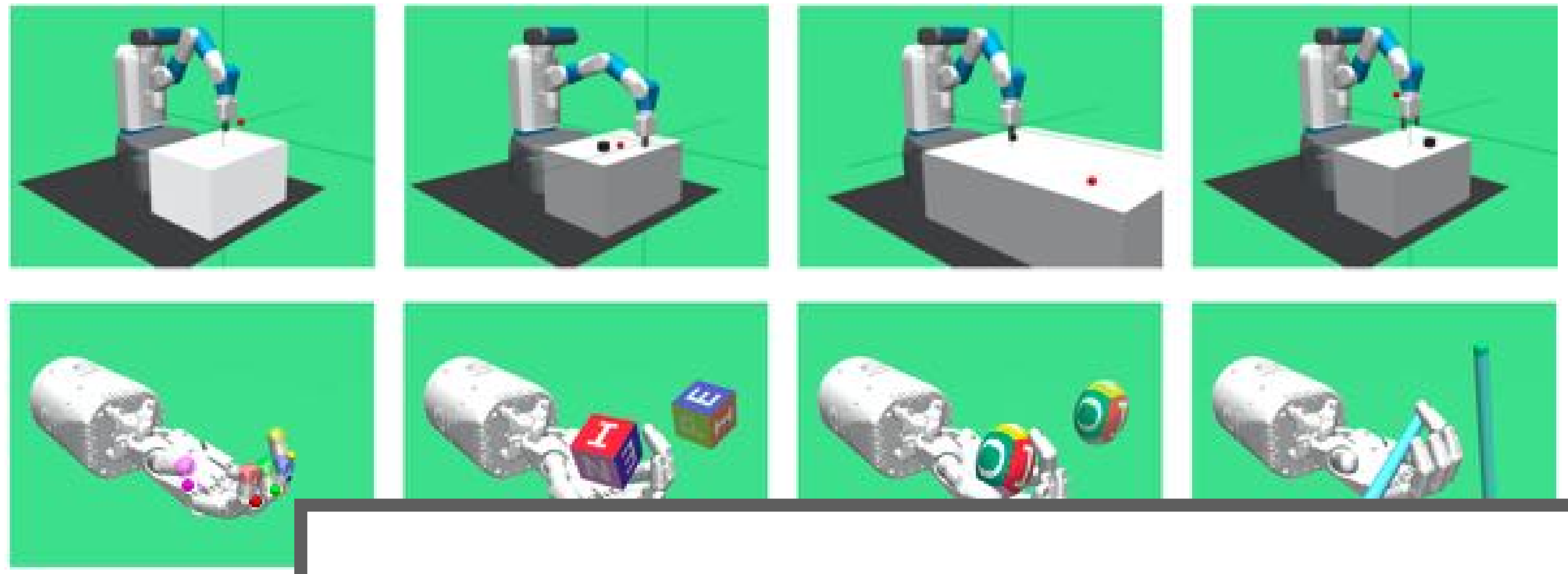
Skill chaining via fine-tuning [CoRL’21]

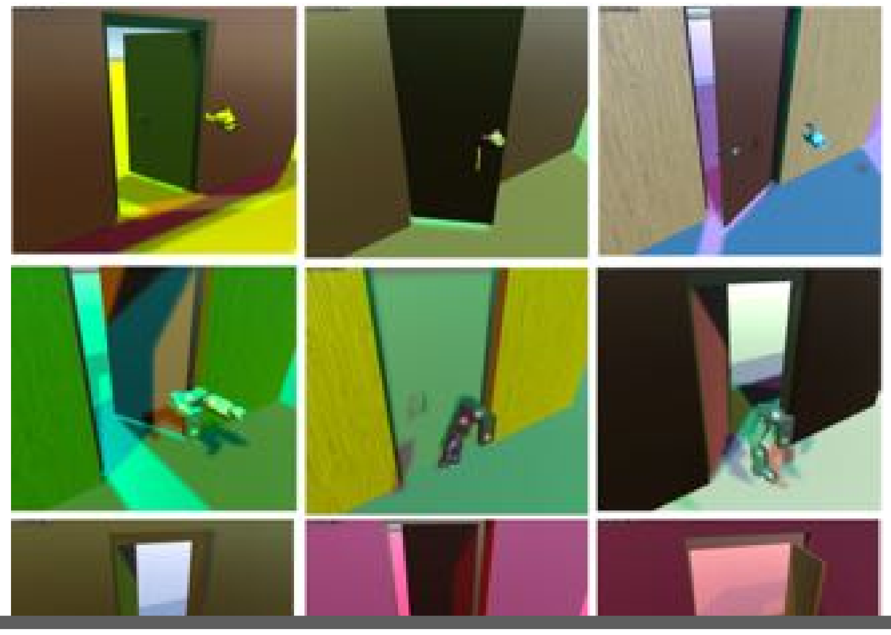


Skill-based RL [CoRL’20, CoRL’21, CoRL’22]

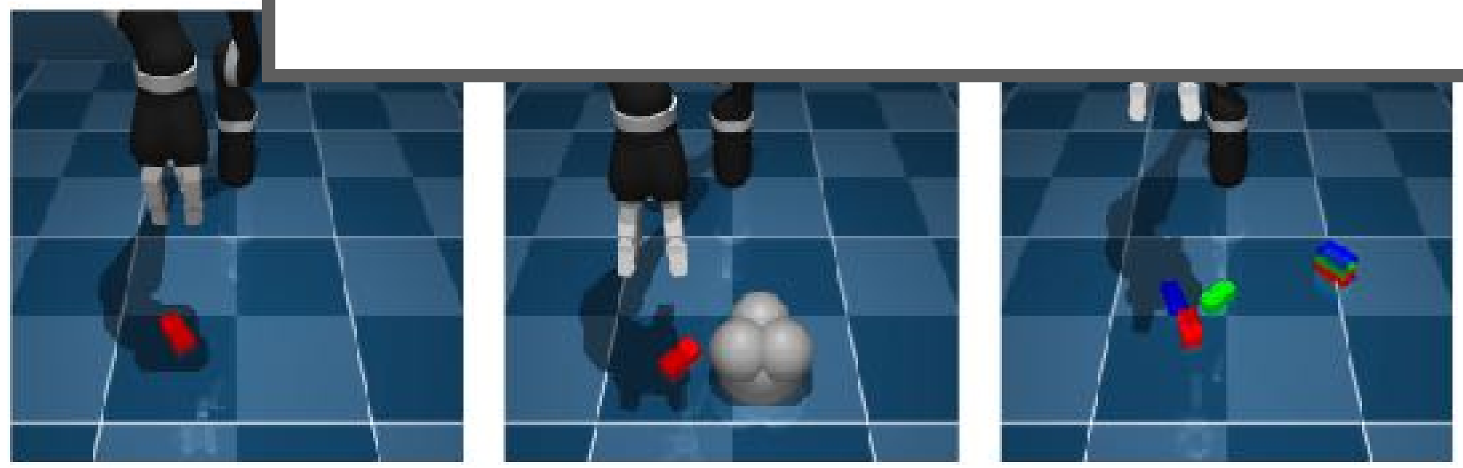
Algorithm

Standard manipulation benchmarks

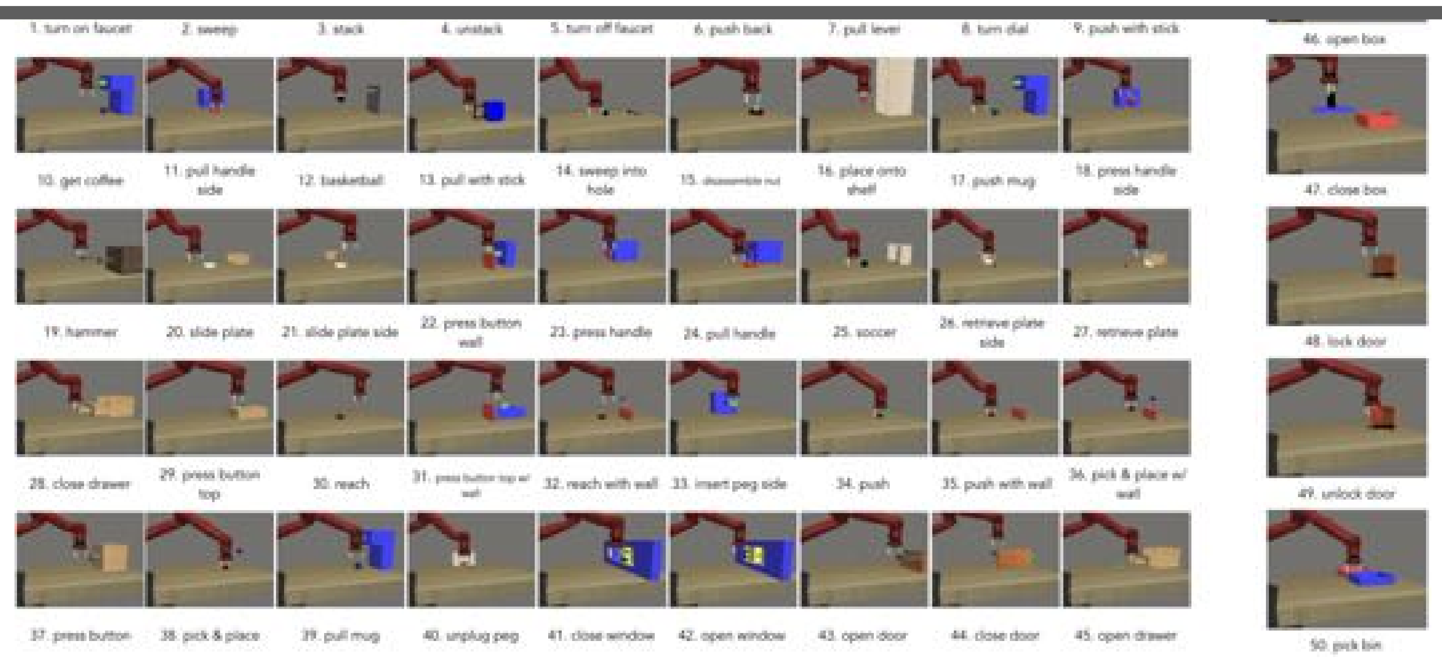




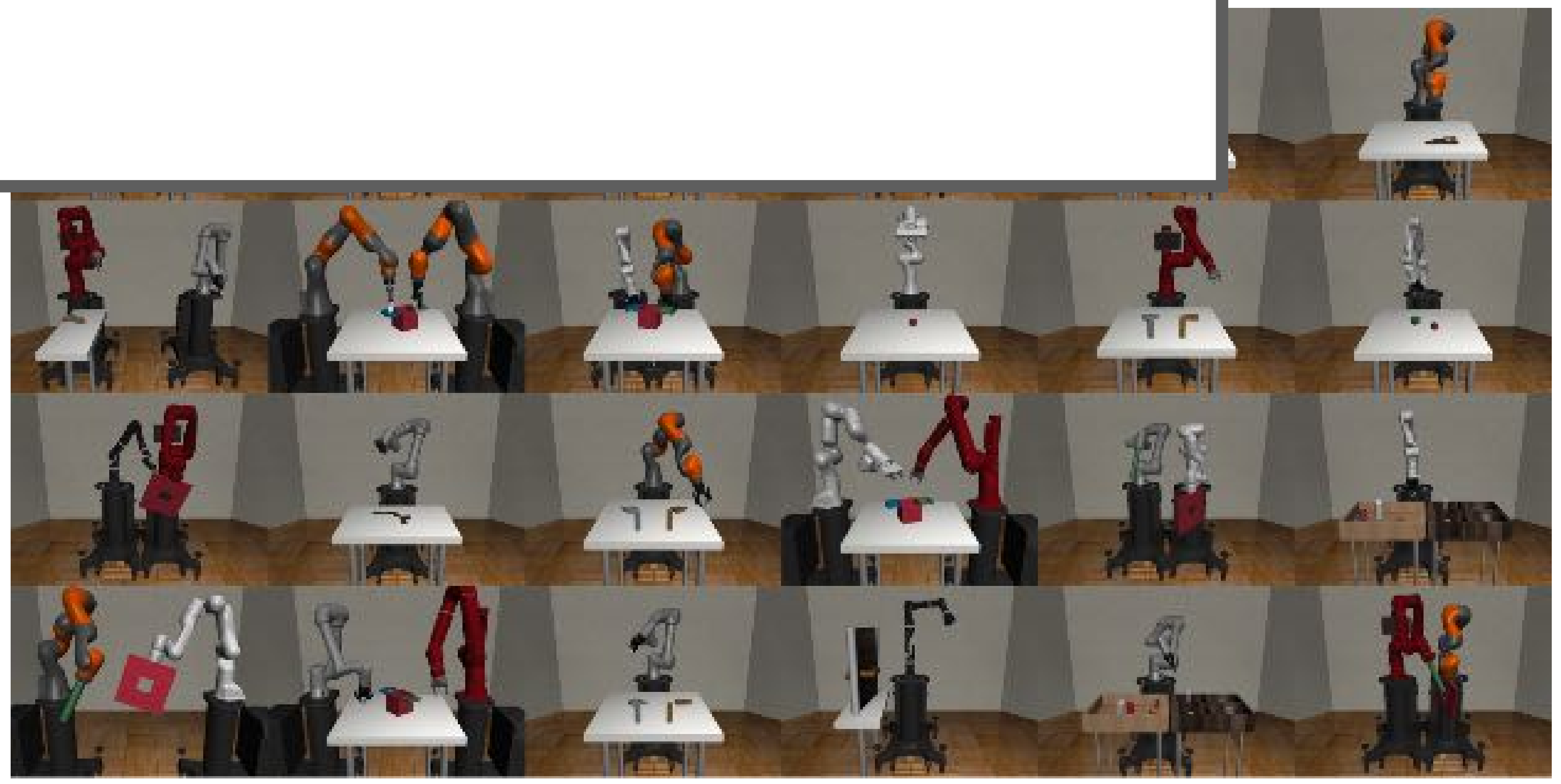
Limited to **simple** and **short** tasks



dm_control

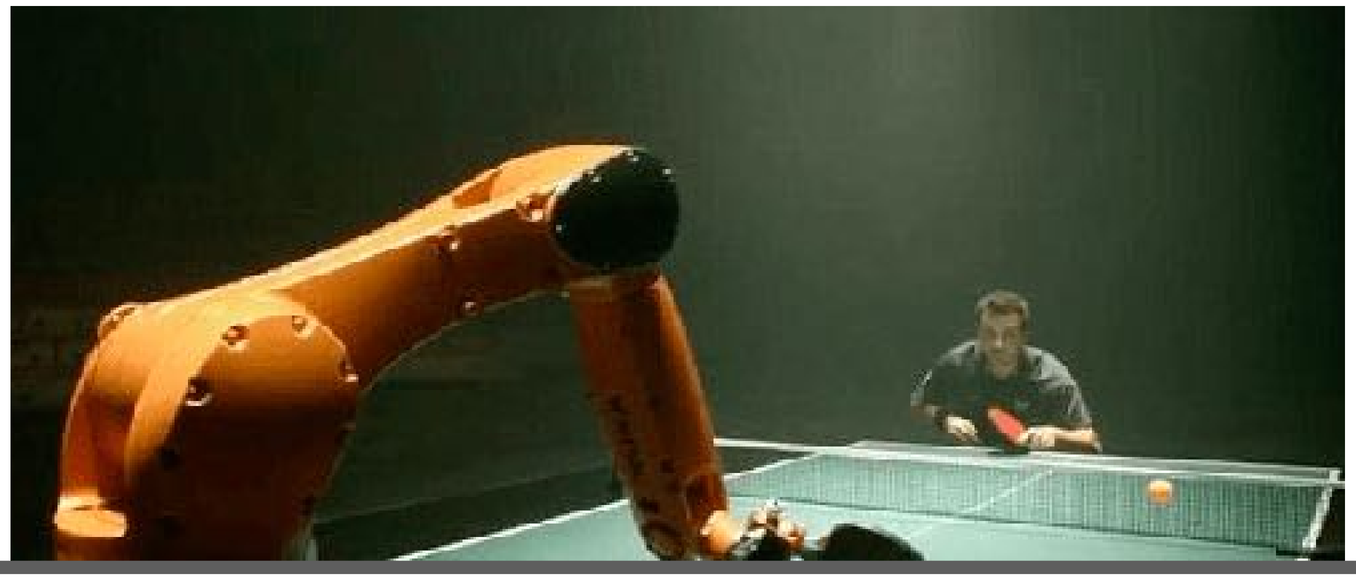


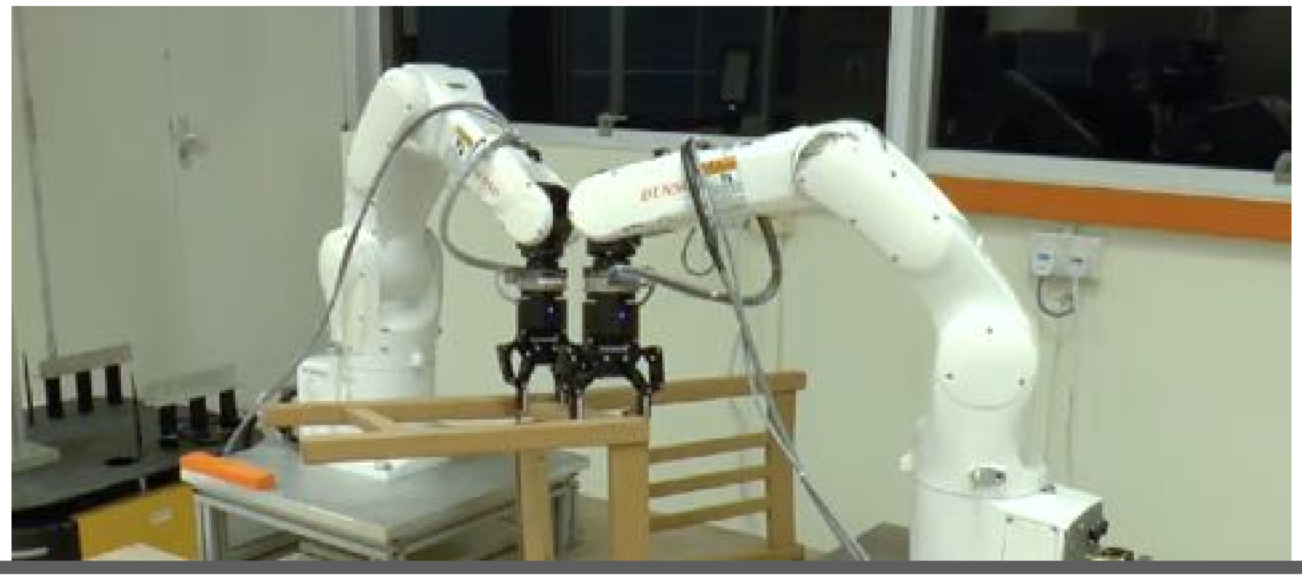
Meta-World



Robosuite

Scaling robot learning to “realistic tasks”





We need a benchmark for **complex** and **long-horizon** tasks

IKEA Furniture Assembly Simulator

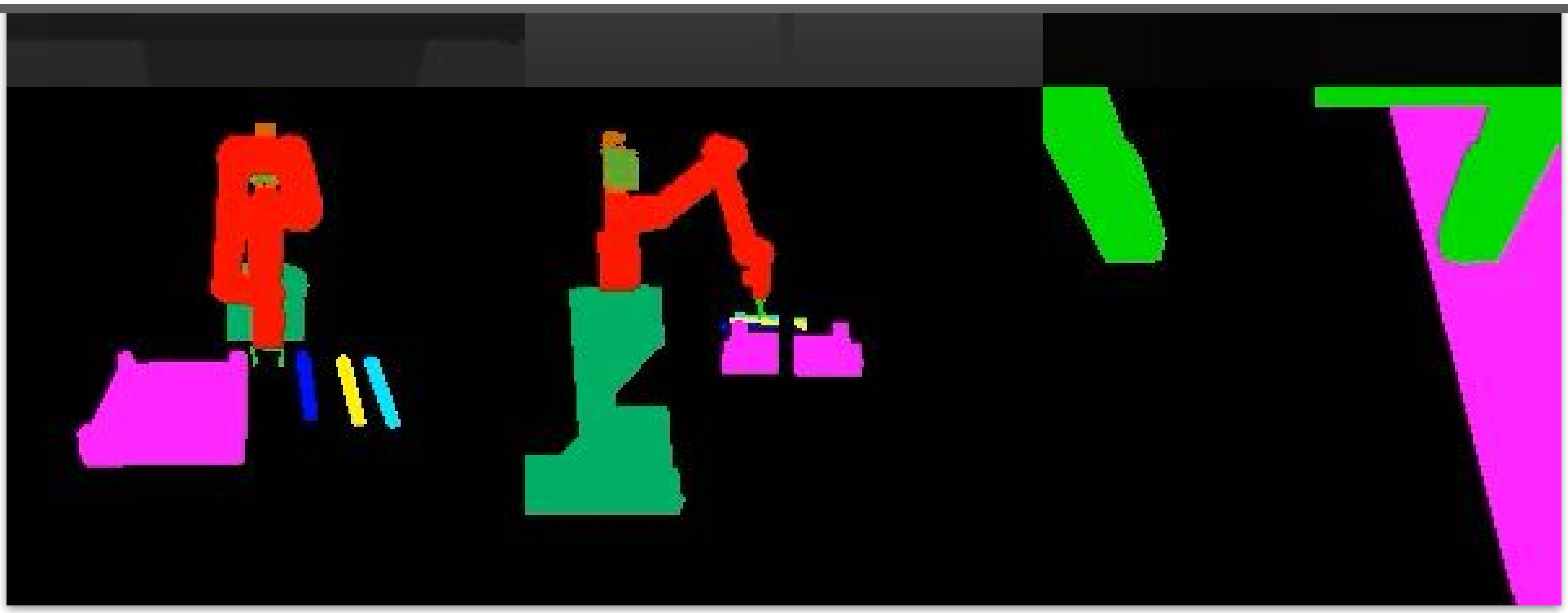
Lee et al. *"IKEA Furniture Assembly Environment"* ICRA 2021

Realistic 3D environment

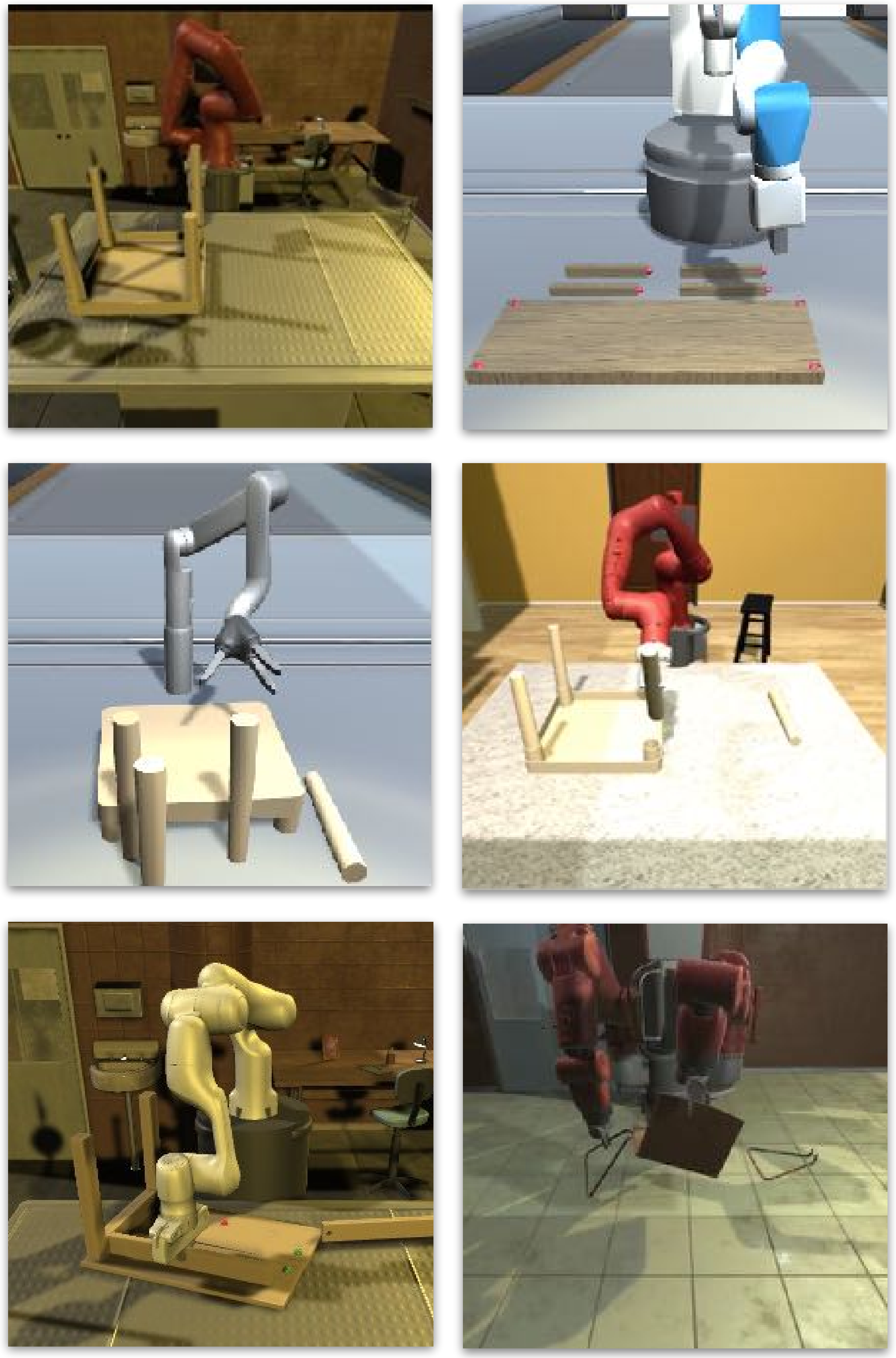
Camera

We propose the first benchmark for **complex, long-horizon tasks**

Segmentation map

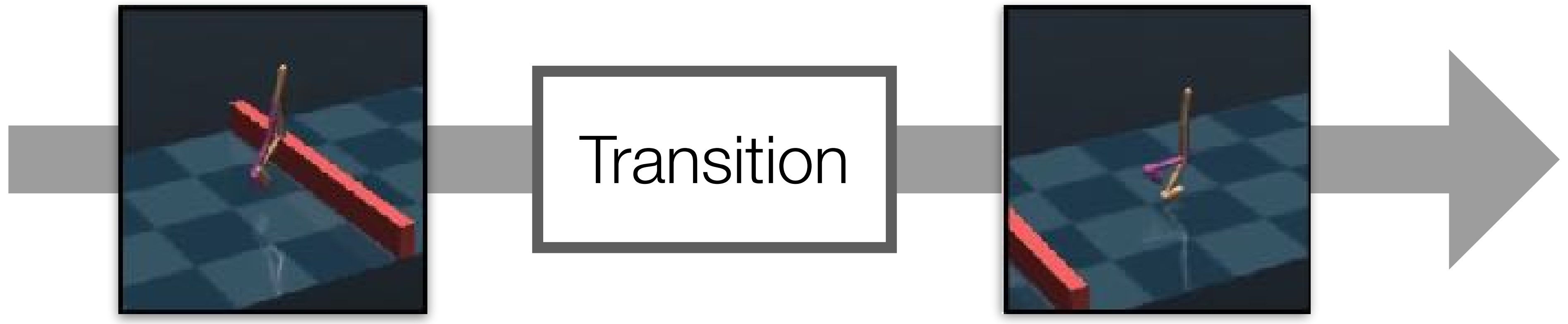


Scaling robot learning to long-horizon tasks

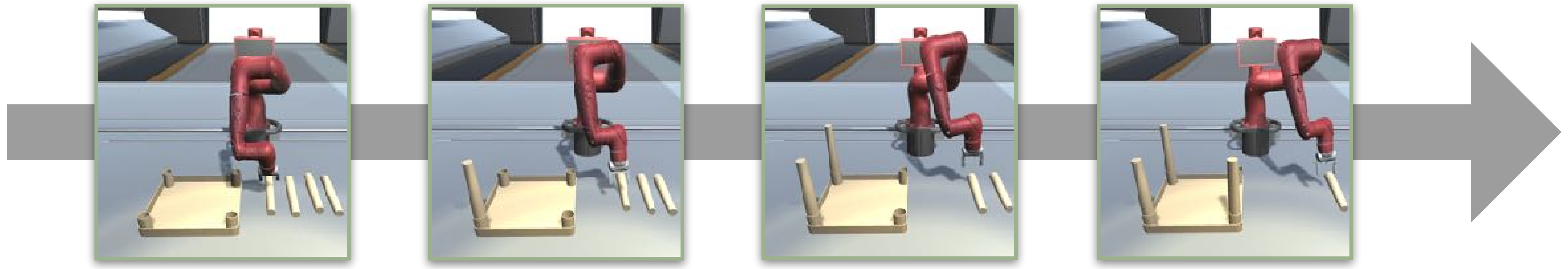


IKEA furniture assembly [ICRA'21]

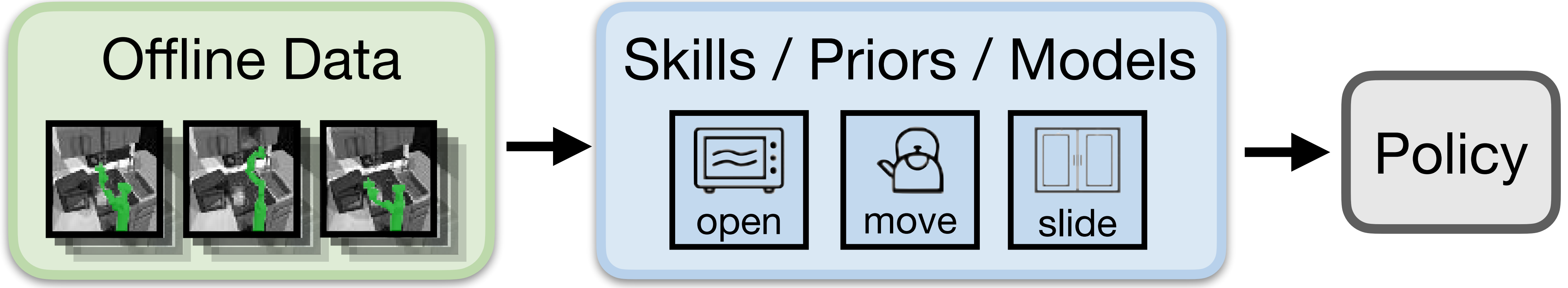
Benchmark



Transition policy [ICLR'19]



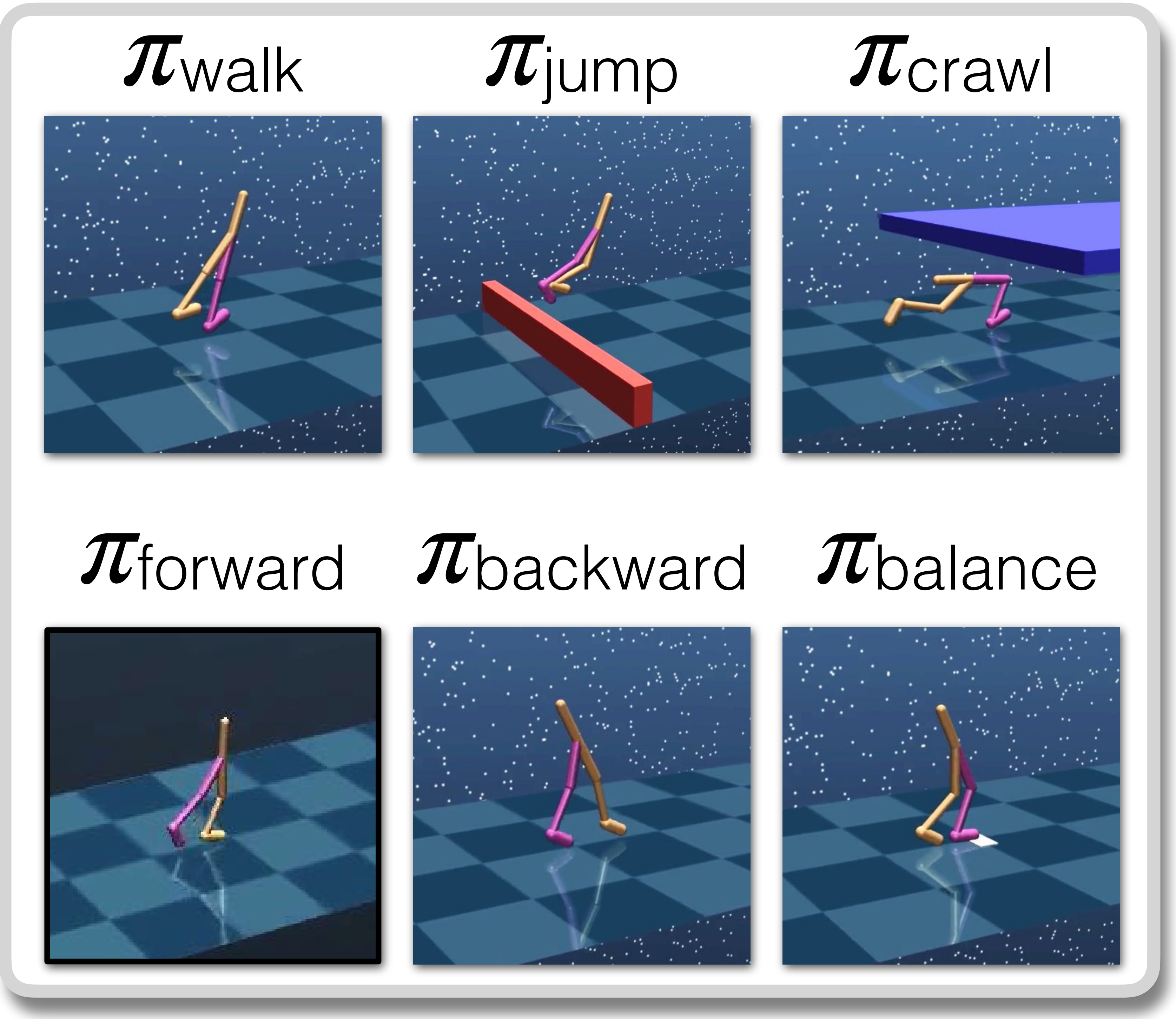
Skill chaining via fine-tuning [CoRL'21]



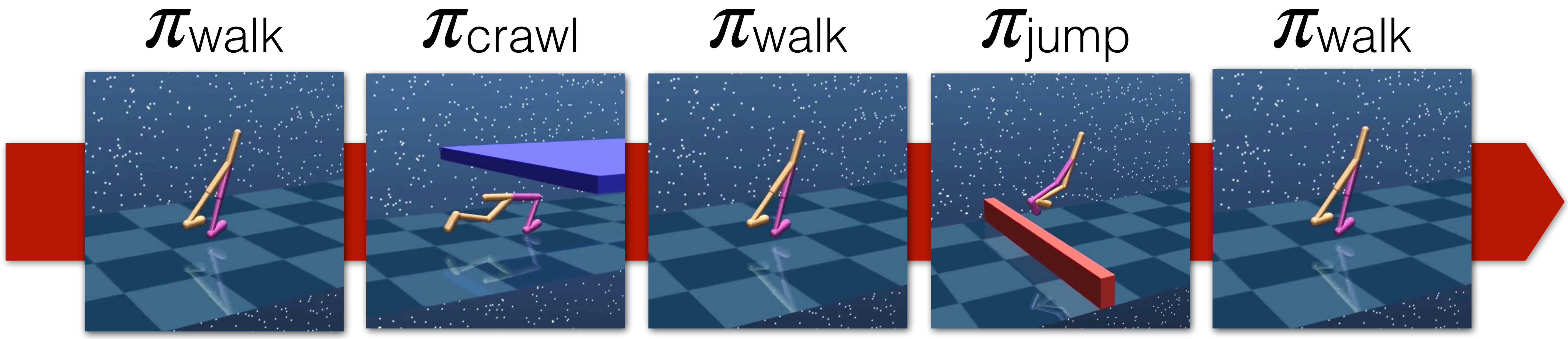
Skill-based RL [CoRL'20, CoRL'21, CoRL'22]

Algorithm

Leverage “skills”



Reusable Skills



Compositional Task

Lee et al. *“Composing Complex Skills by Learning Transition Policies”* ICLR 2019

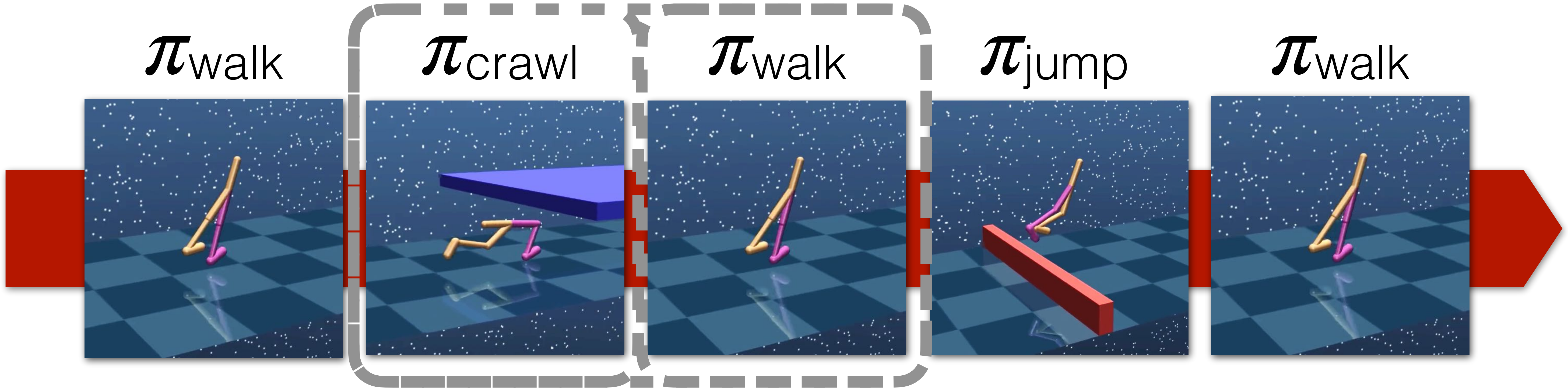

Catch-Dribble-Shoot


Catch-~~Dribble~~-Shoot

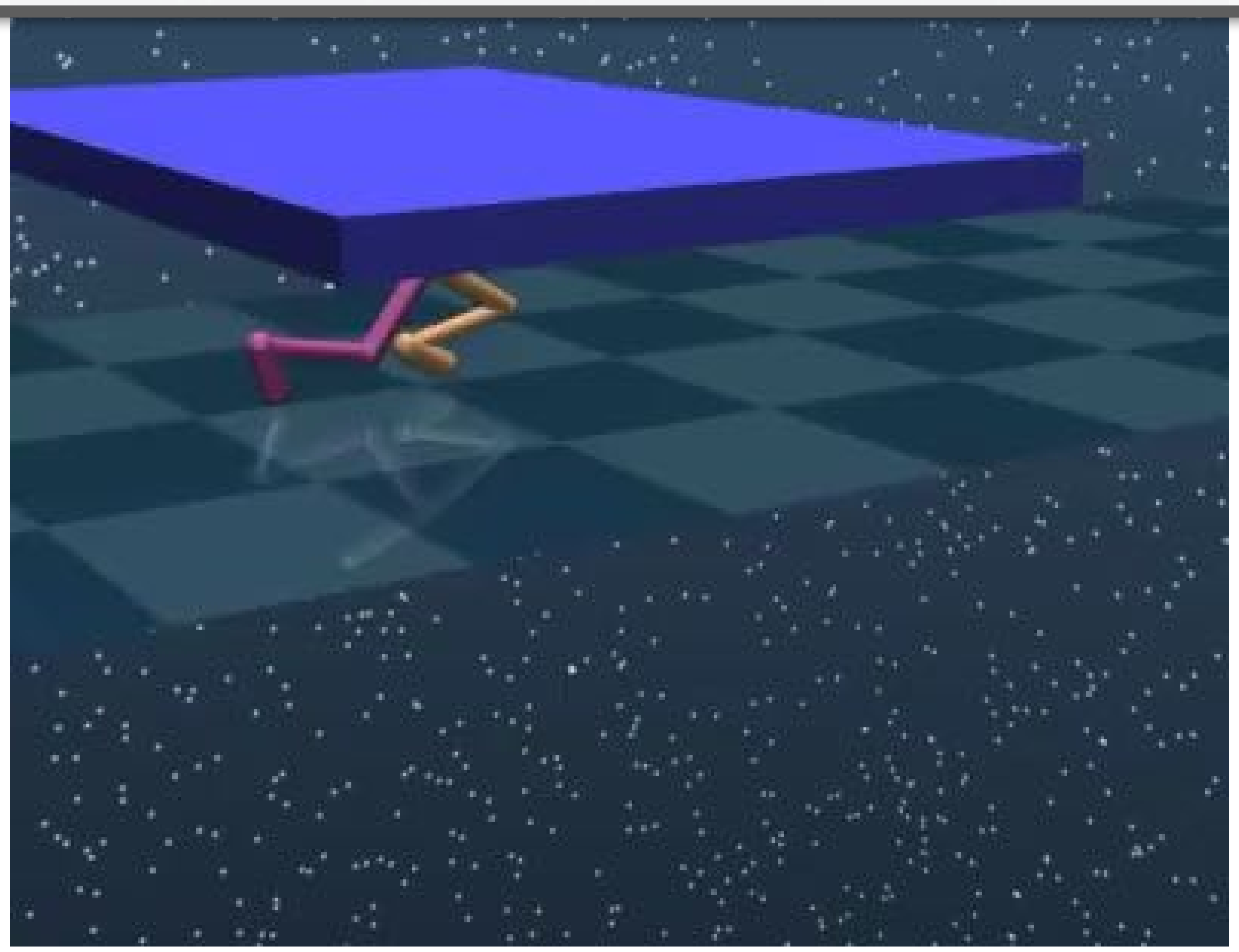
Learning a complex task efficiently:

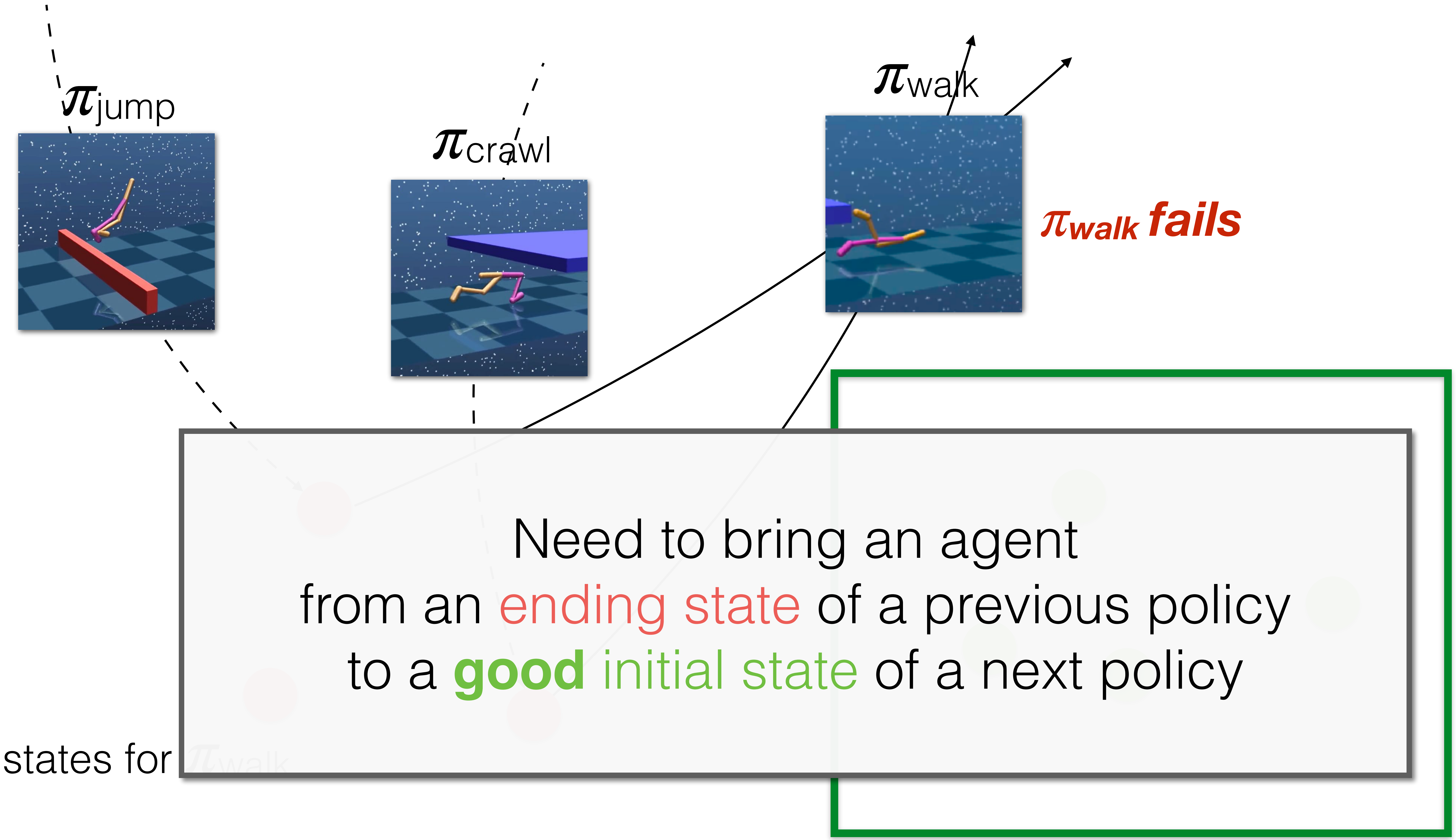
(1) learning **skills in isolation**

(2) learning **“transitions”** between skills



Fail since these skills never learned to connect



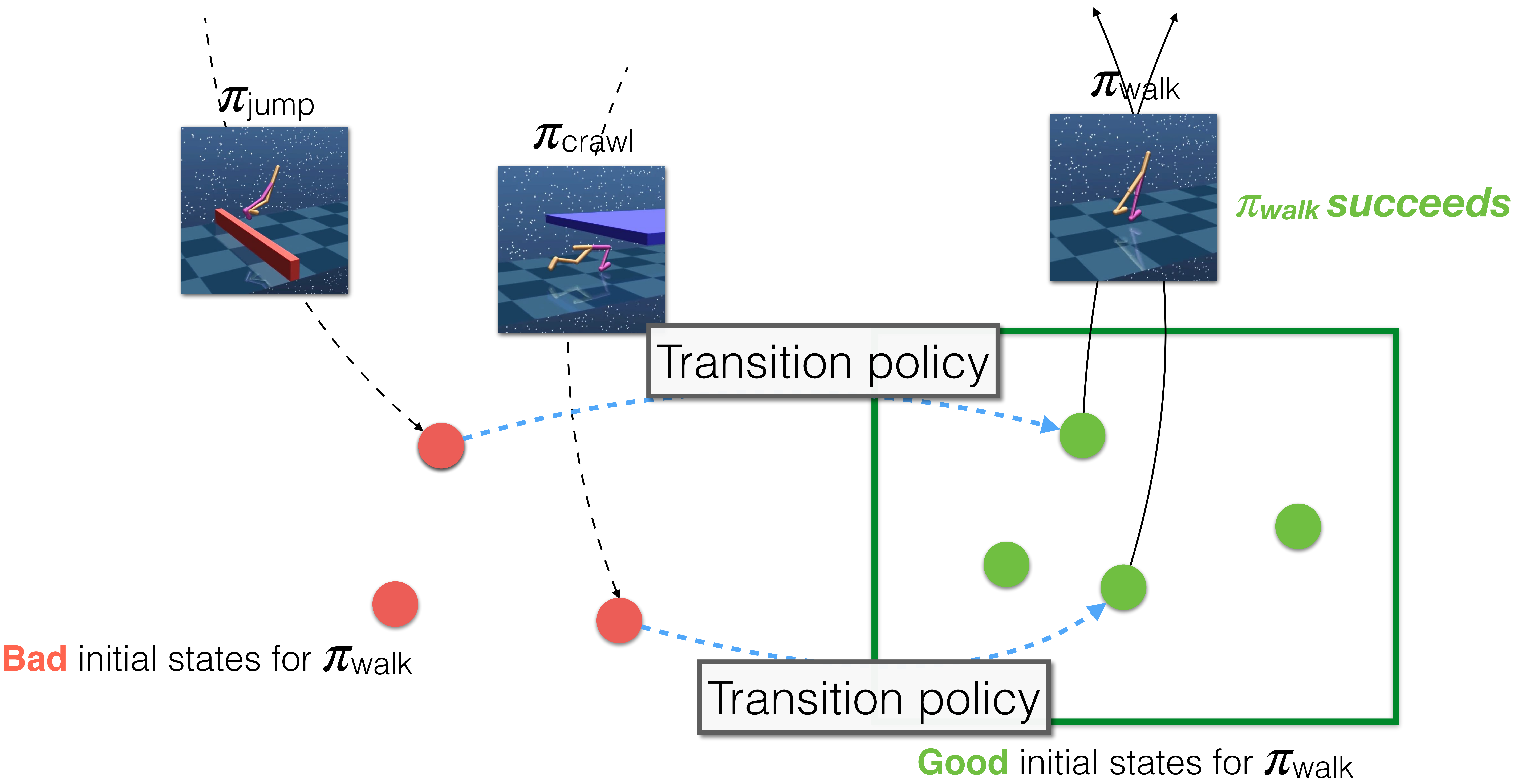


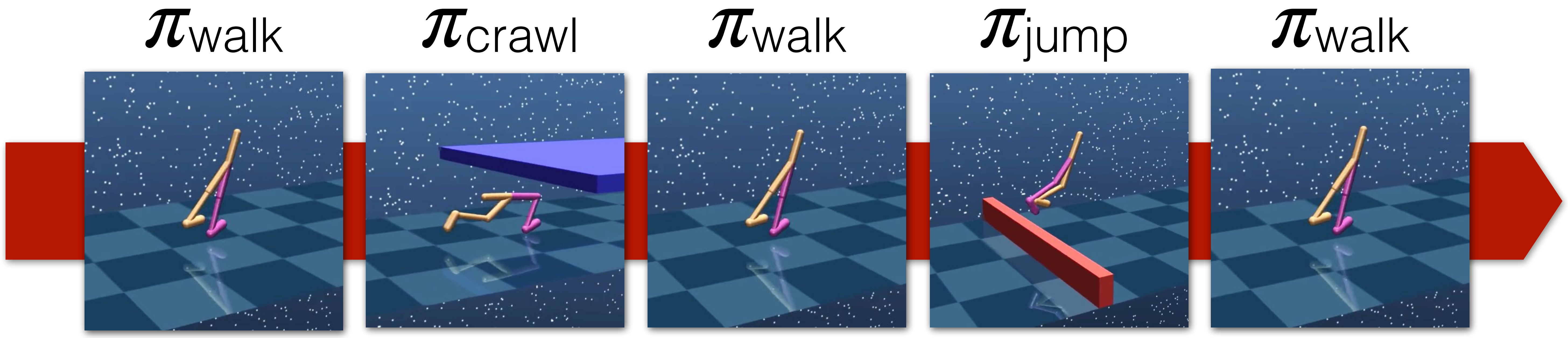
Need to bring an agent
from an ending state of a previous policy
to a **good** initial state of a next policy

Bad initial states for π_{walk}

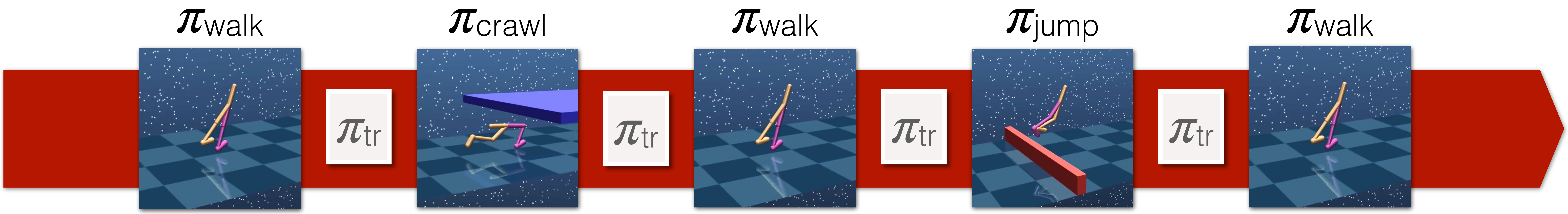
Good initial states for π_{walk}

"Initiation set"


π_{jump}
π_{crawl}
π_{walk}
π_{walk} succeeds
Transition policy
Transition policy
Bad initial states for π_{walk}
Good initial states for π_{walk}



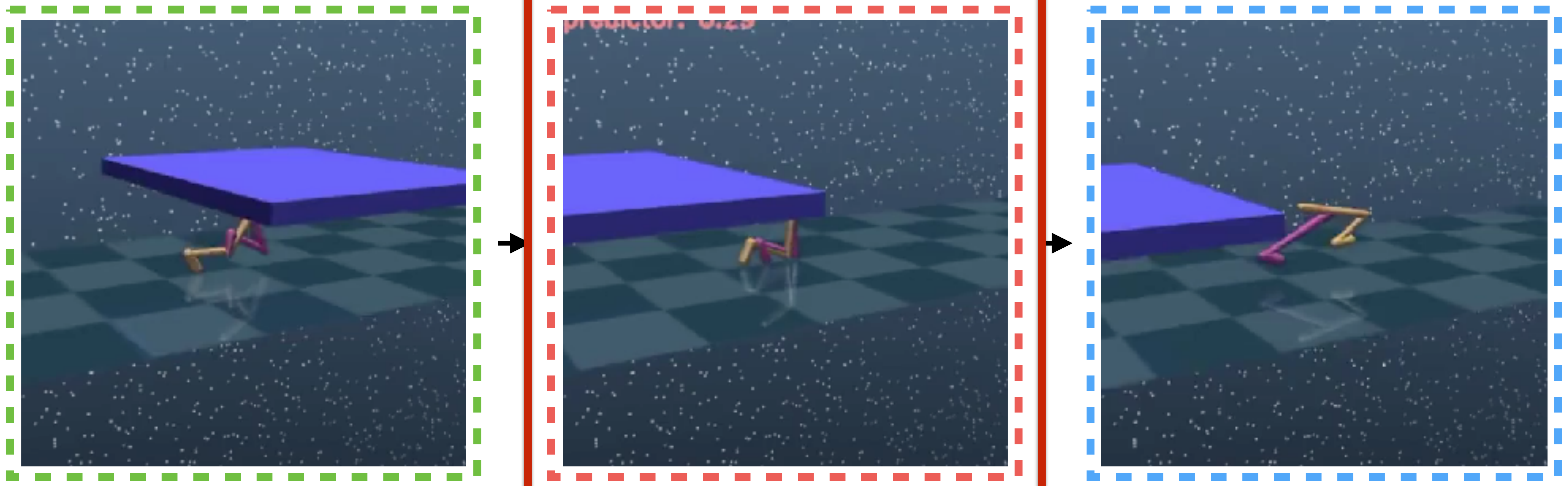
π_{walk}
π_{crawl}
π_{walk}
π_{jump}
π_{walk}


π_{walk}
π_{tr}
π_{crawl}
π_{tr}
π_{walk}
π_{tr}
π_{jump}
π_{tr}
π_{walk}

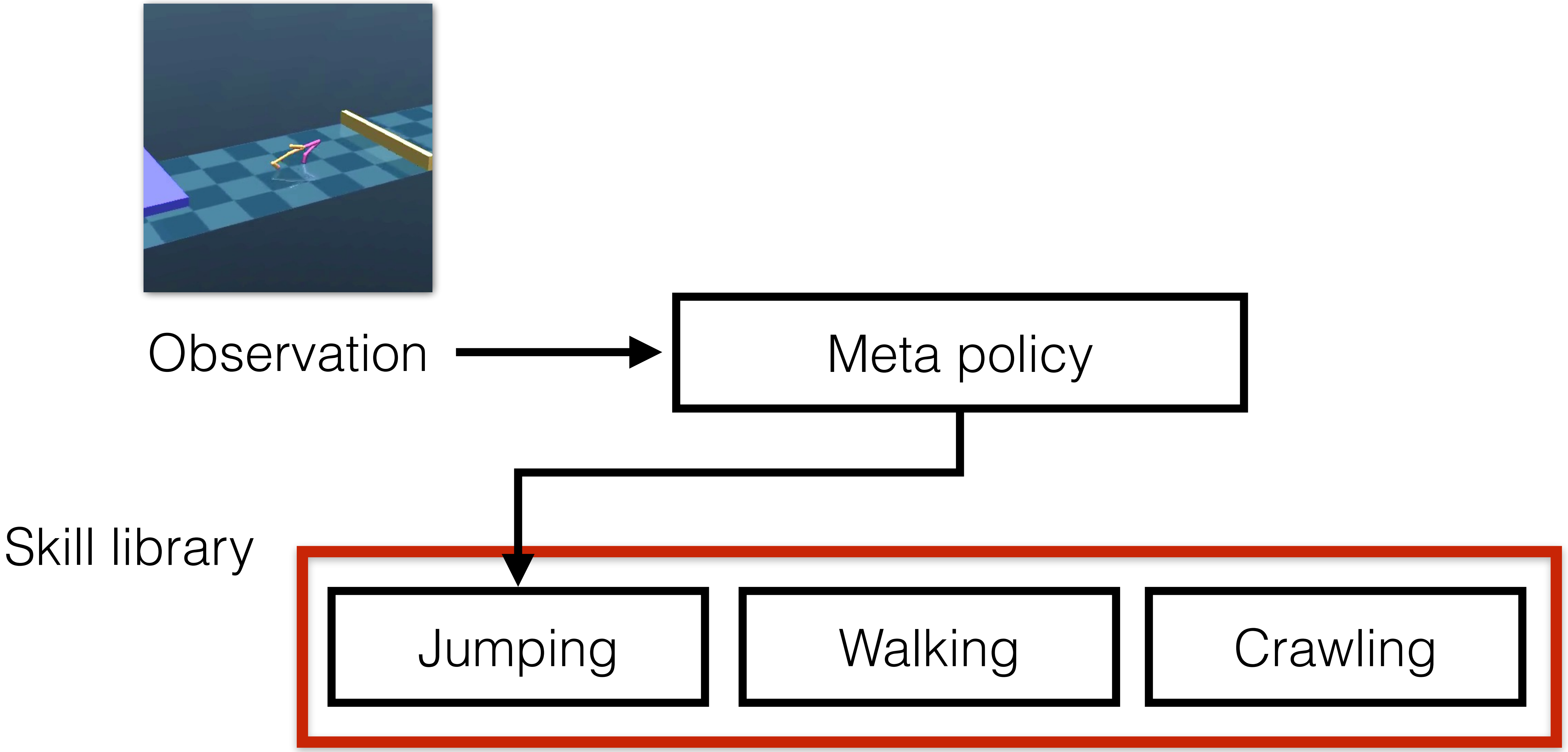
Smoothly connect skills

Obstacle course

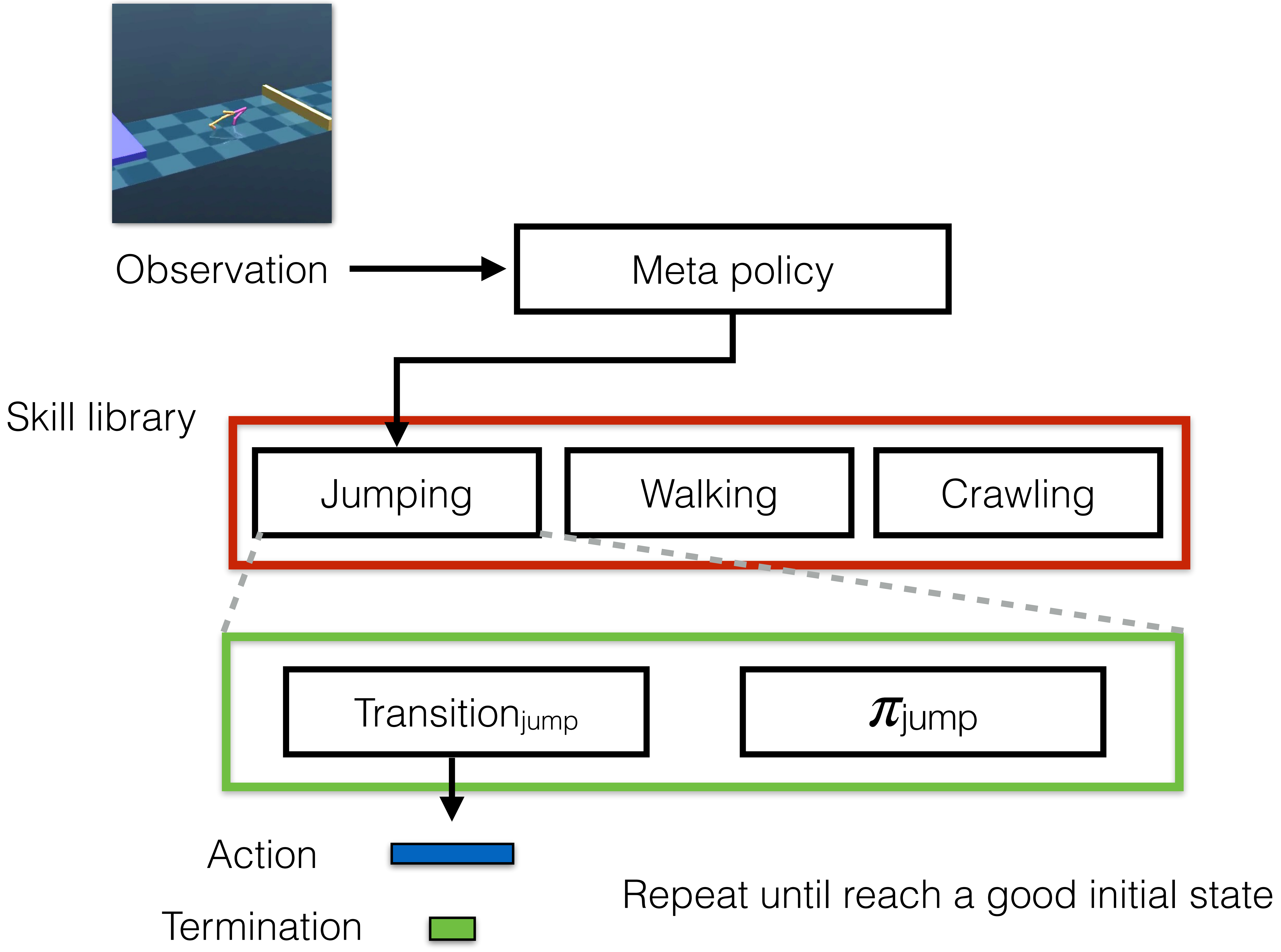
π_{crawl} → **Transition policy** → π_{walk}



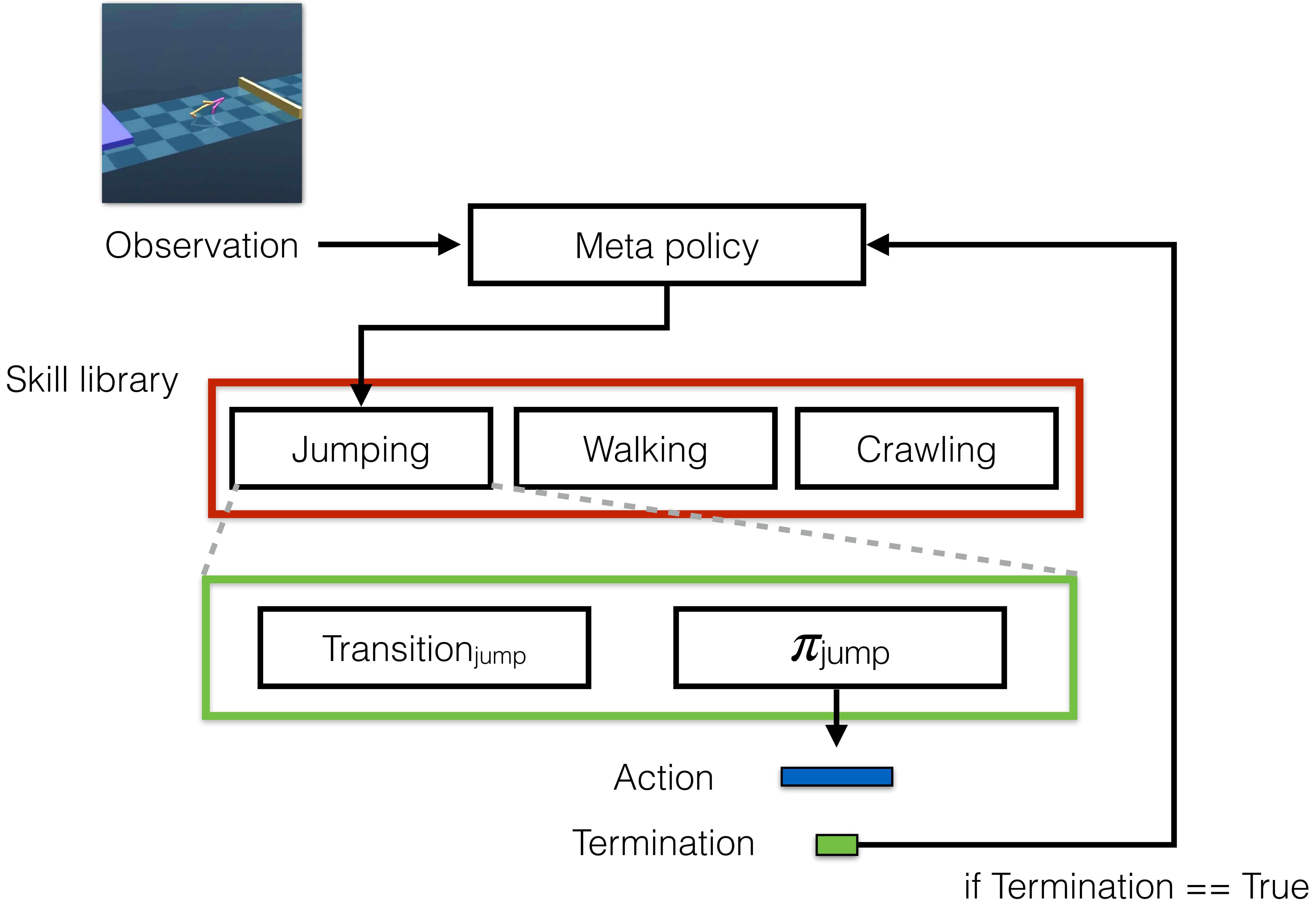
Model


Observation
Meta policy
Skill library
Jumping
Walking
Crawling

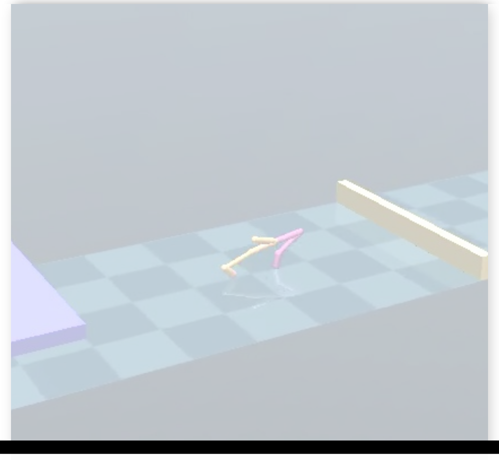
Model


Observation
Meta policy
Skill library
Jumping
Walking
Crawling
Transition_{jump}
π_{jump}
Action
Termination
Repeat until reach a good initial state

Model

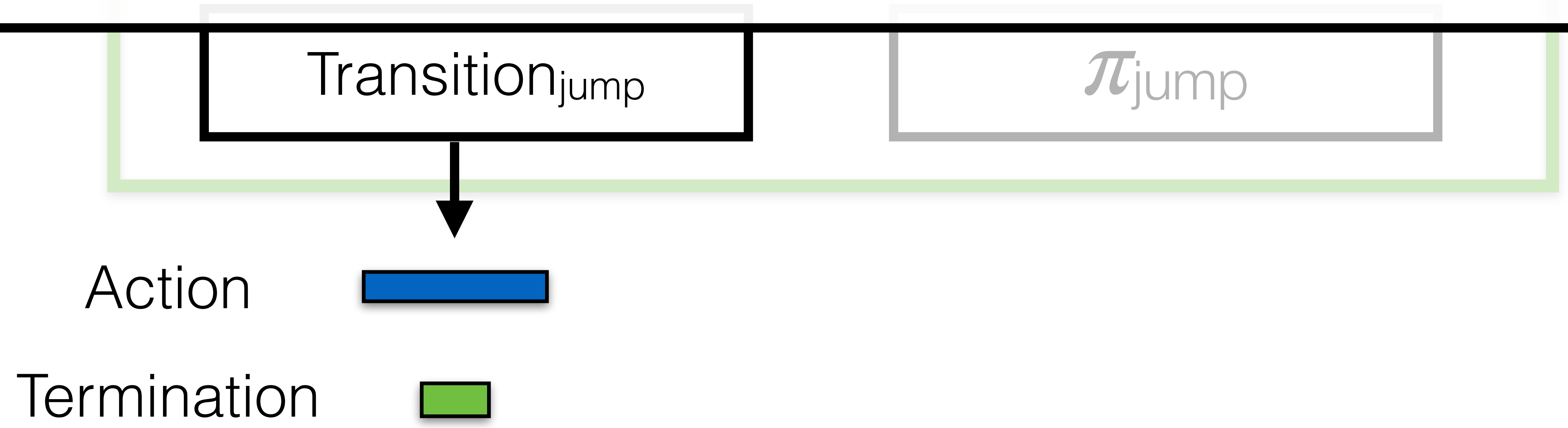

Observation
Meta policy
Skill library
Jumping
Walking
Crawling
Transition$_{jump}$
π_{jump}
Action
Termination
if Termination == True

Model

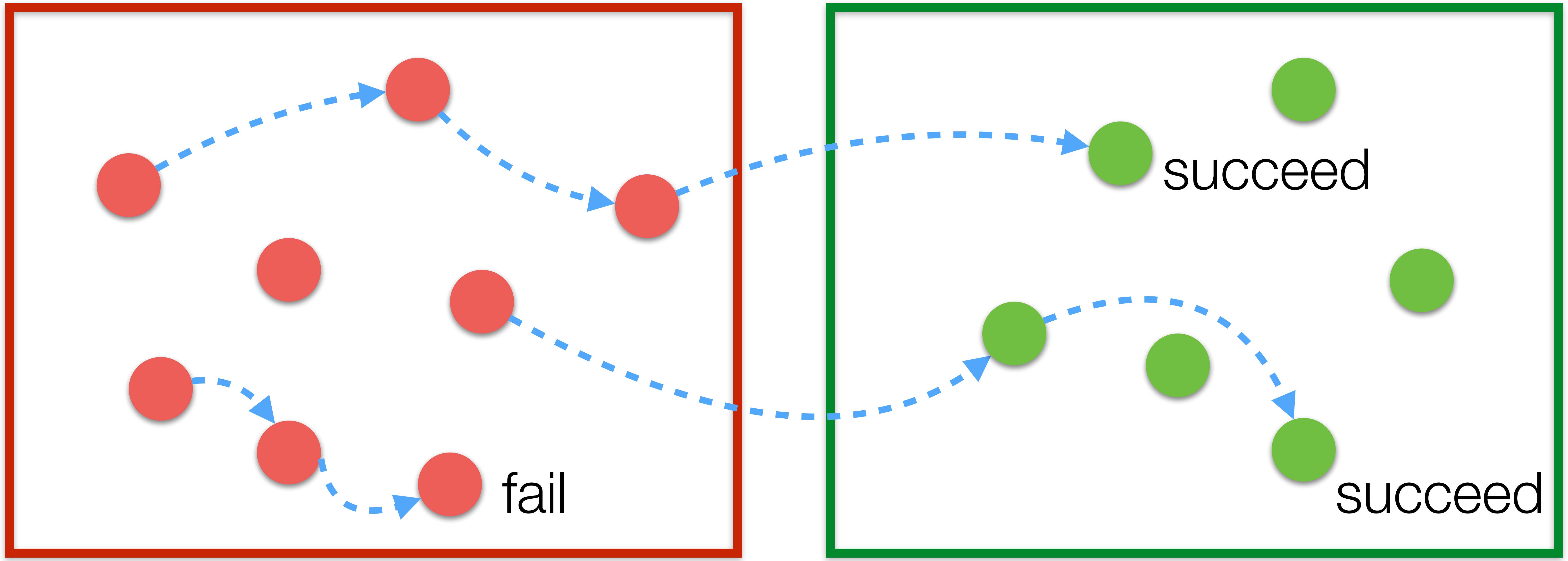


What is reward for learning a transition policy?

Success of the following skill



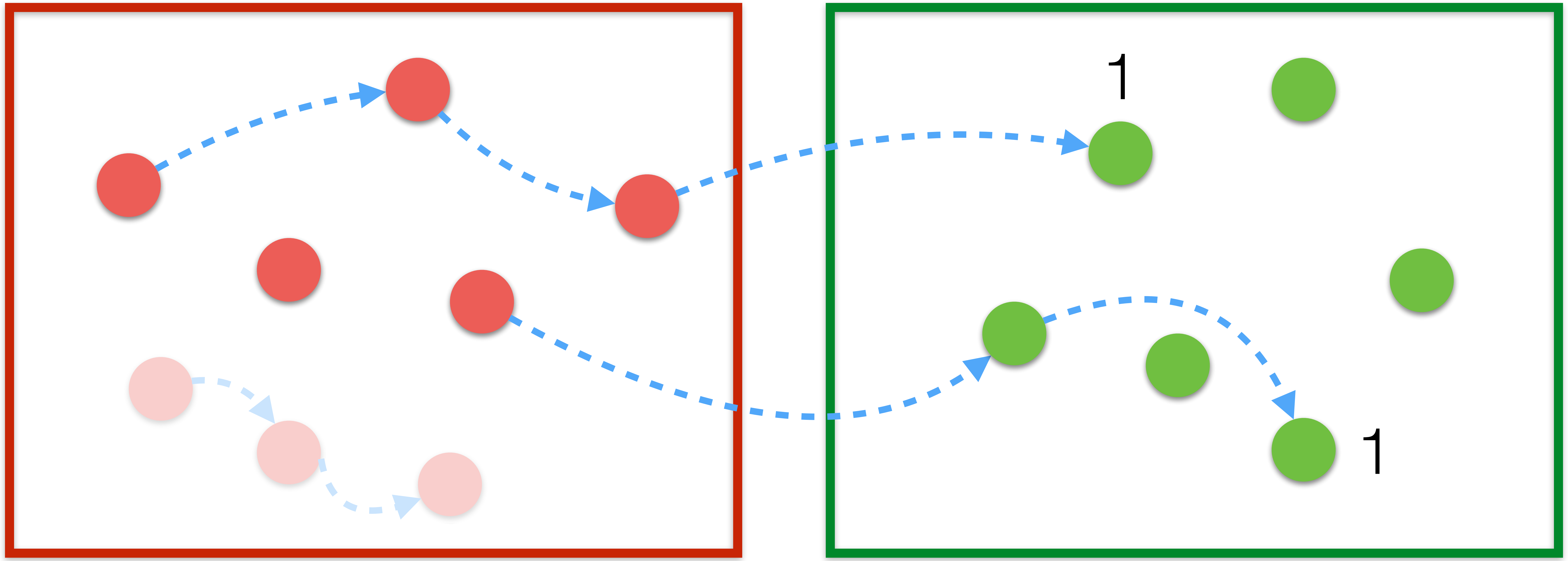
Learning transition policy



Bad initial states for π_{walk} **Good** initial states for π_{walk}

Learning transition policy

Successful execution of the following skill: +1

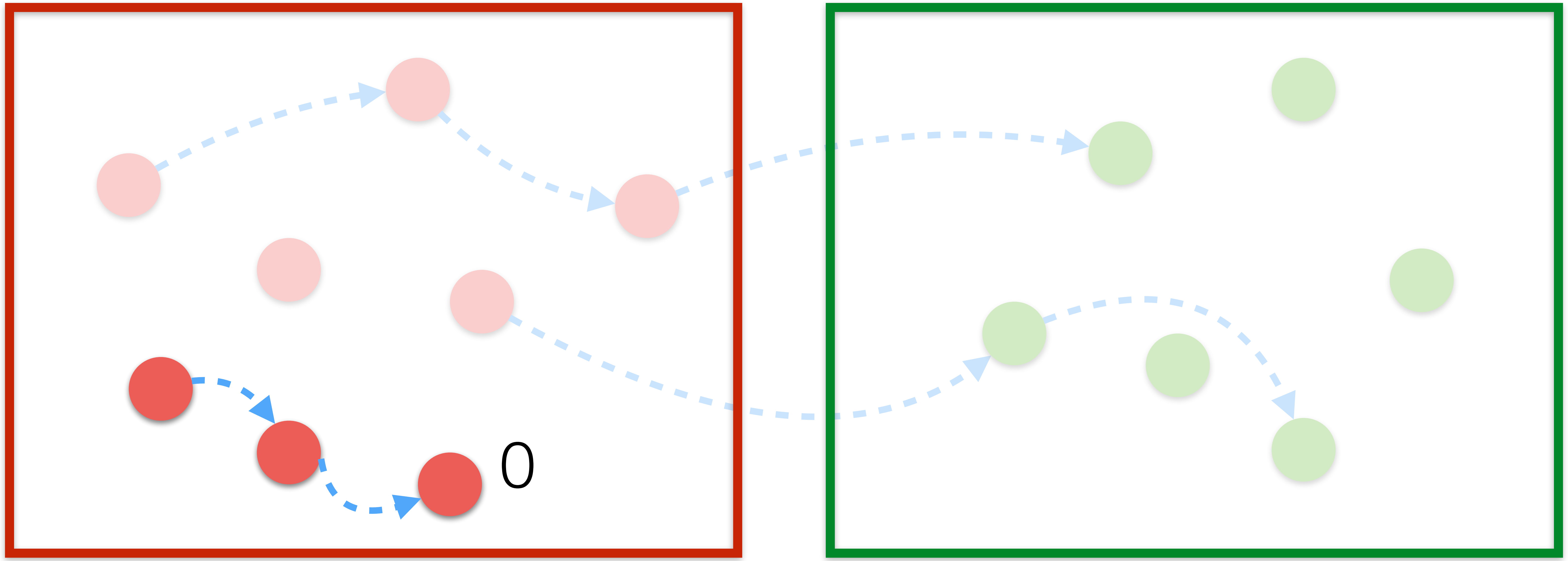


Bad initial states for π_{walk} **Good** initial states for π_{walk}

Learning transition policy

Successful execution of the following skill: +1

Failing execution of the following skill: 0

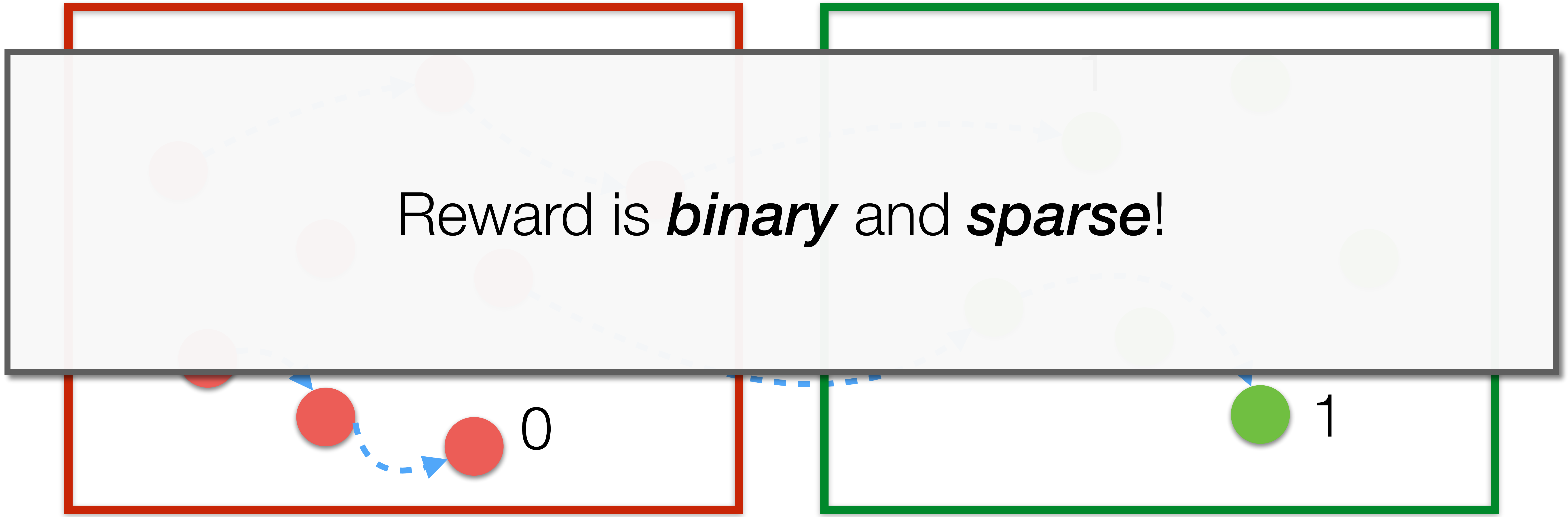


Bad initial states for π_{walk} **Good** initial states for π_{walk}

Learning transition policy

Successful execution of the following skill: +1

Failing execution of the following skill: 0

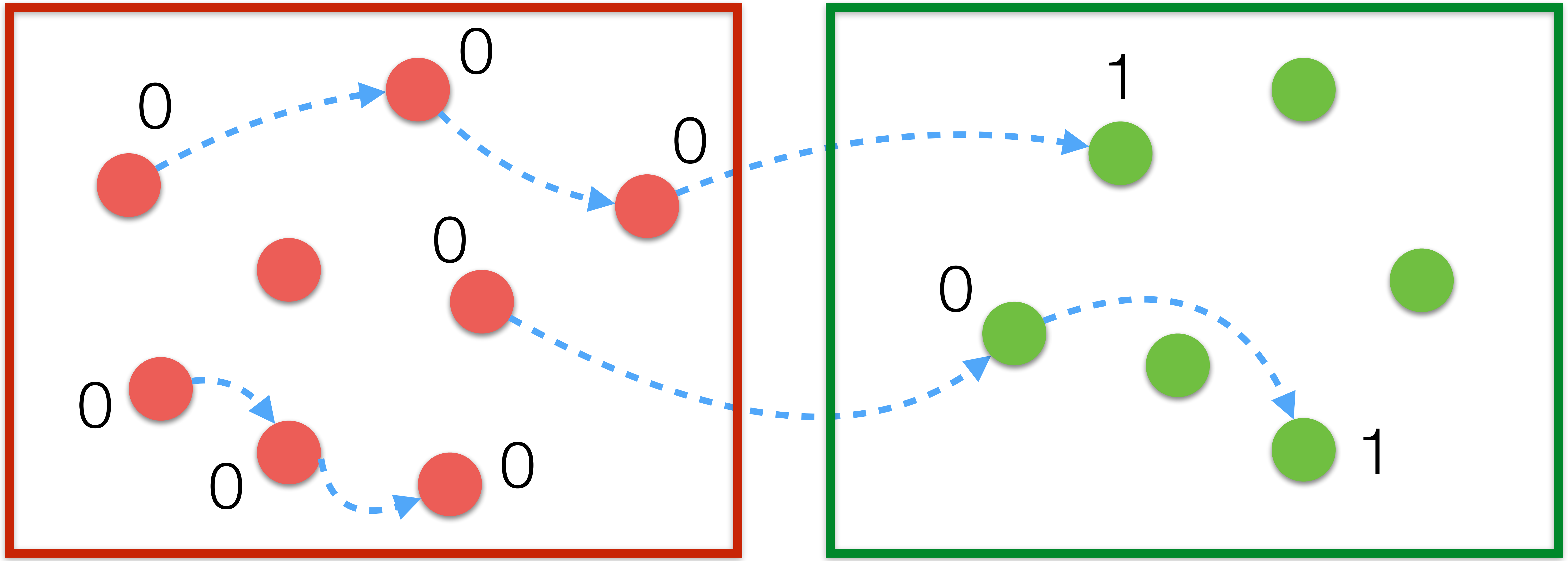


Bad initial states for π_{walk}

Good initial states for π_{walk}

Proximity reward

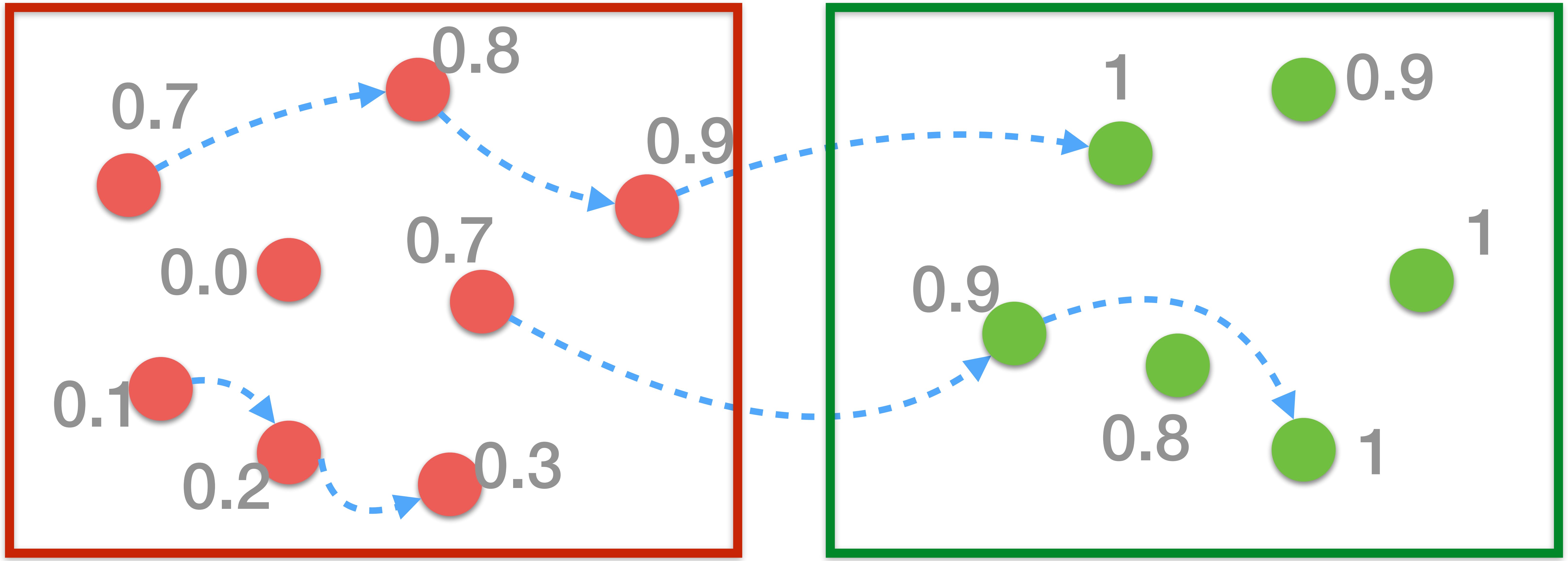
Instead of binary reward



Bad initial states for π_{walk} **Good** initial states for π_{walk}

Proximity reward

Instead of binary reward, use ***"proximity prediction"***,
which estimates how close to good initial states

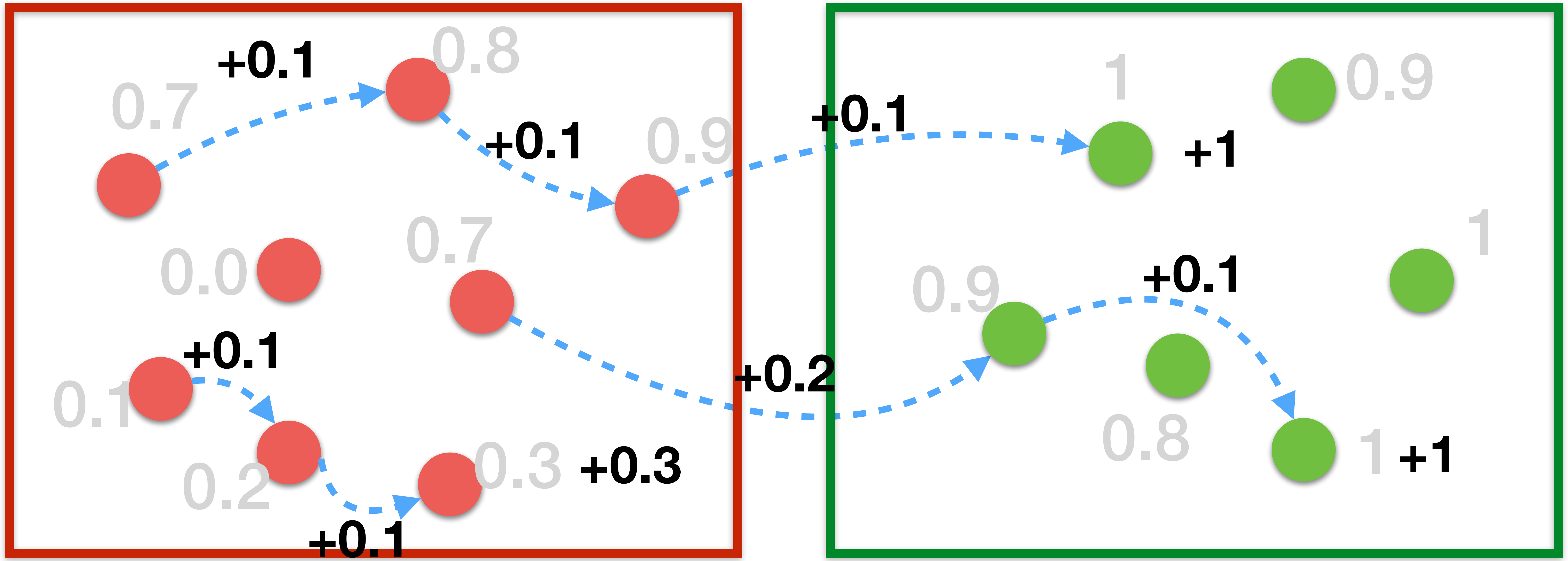


Bad initial states for π_{walk} **Good** initial states for π_{walk}

We define ***proximity*** as: $P(s) = \delta^{step} \quad \delta \in (0,1)$

Proximity reward

Instead of binary reward, use ***"proximity prediction"***,
which estimates <u>how close</u> to good initial states



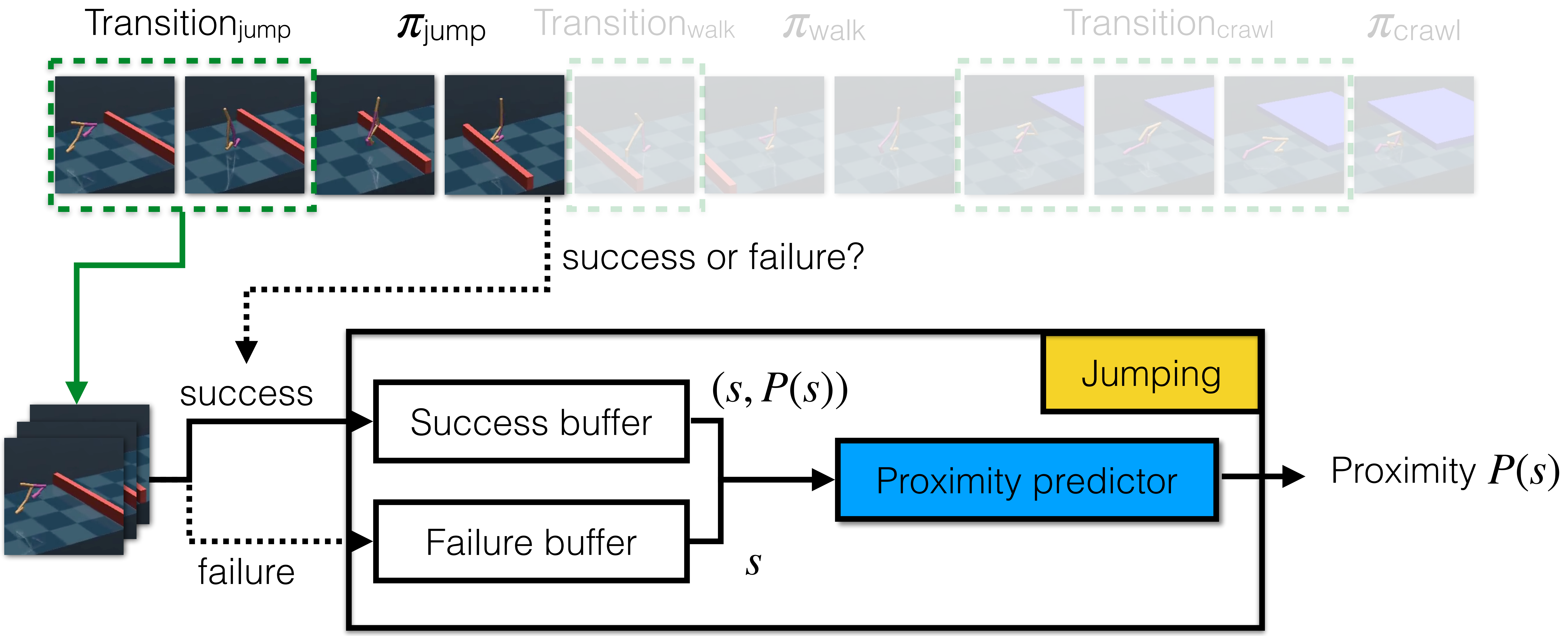
Bad initial states for π_{walk} **Good** initial states for π_{walk}

We define ***proximity*** as: $P(s) = \delta^{step} \quad \delta \in (0,1)$

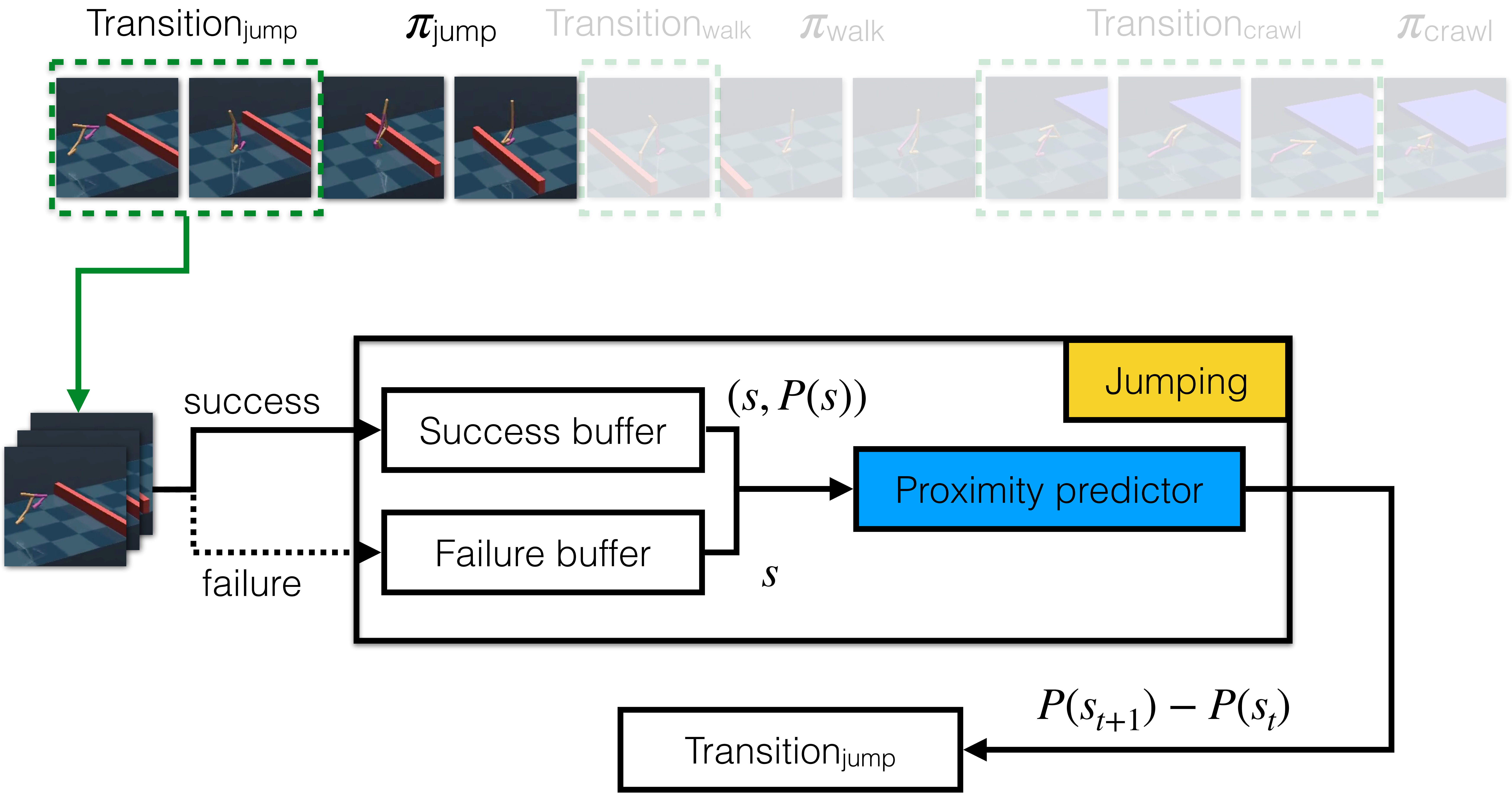
and provide ***proximity reward*** every step: $P(s_{t+1}) - P(s_t)$

Lee et al. *"Composing Complex Skills by Learning Transition Policies"* ICLR 2019

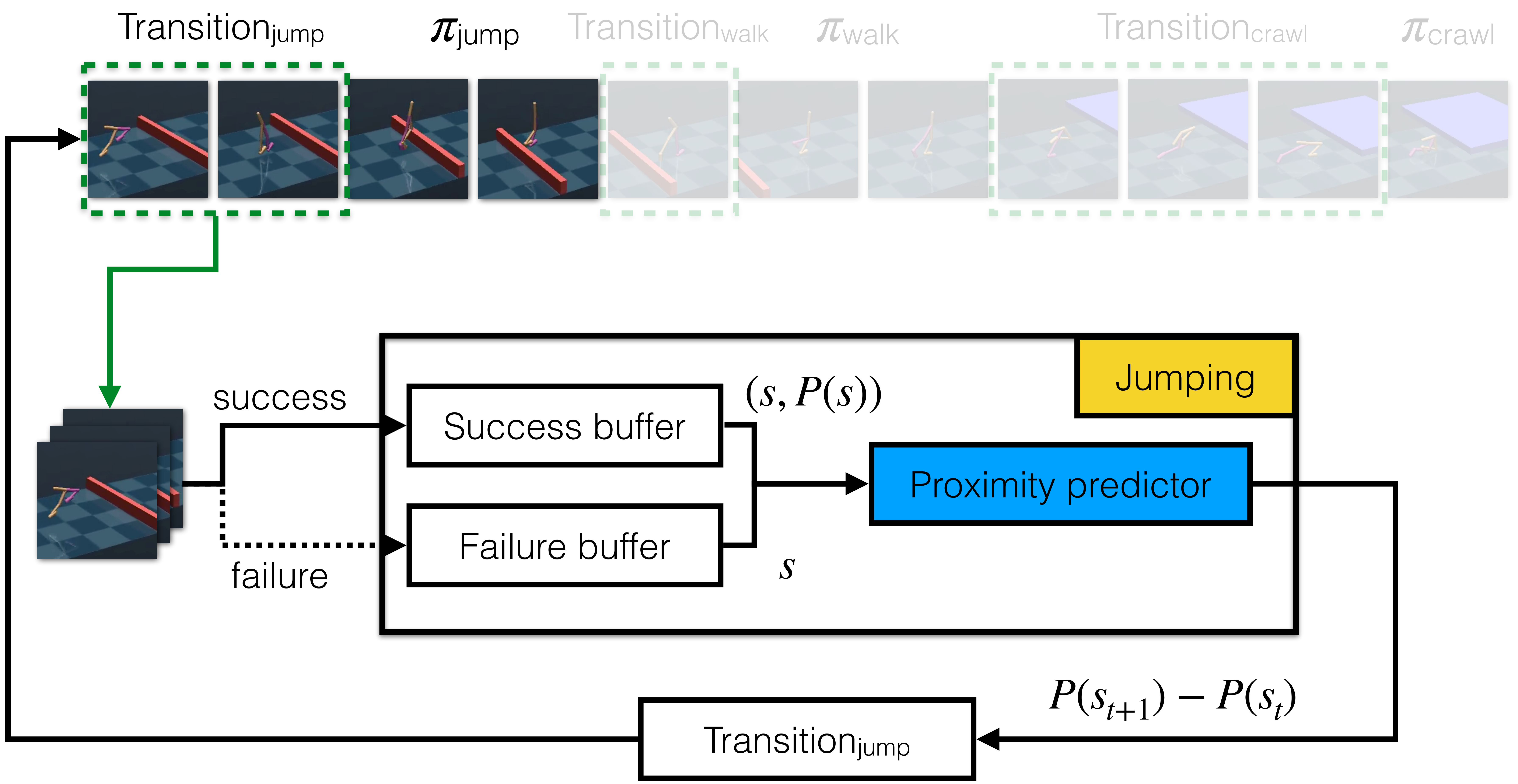
Learning proximity predictor



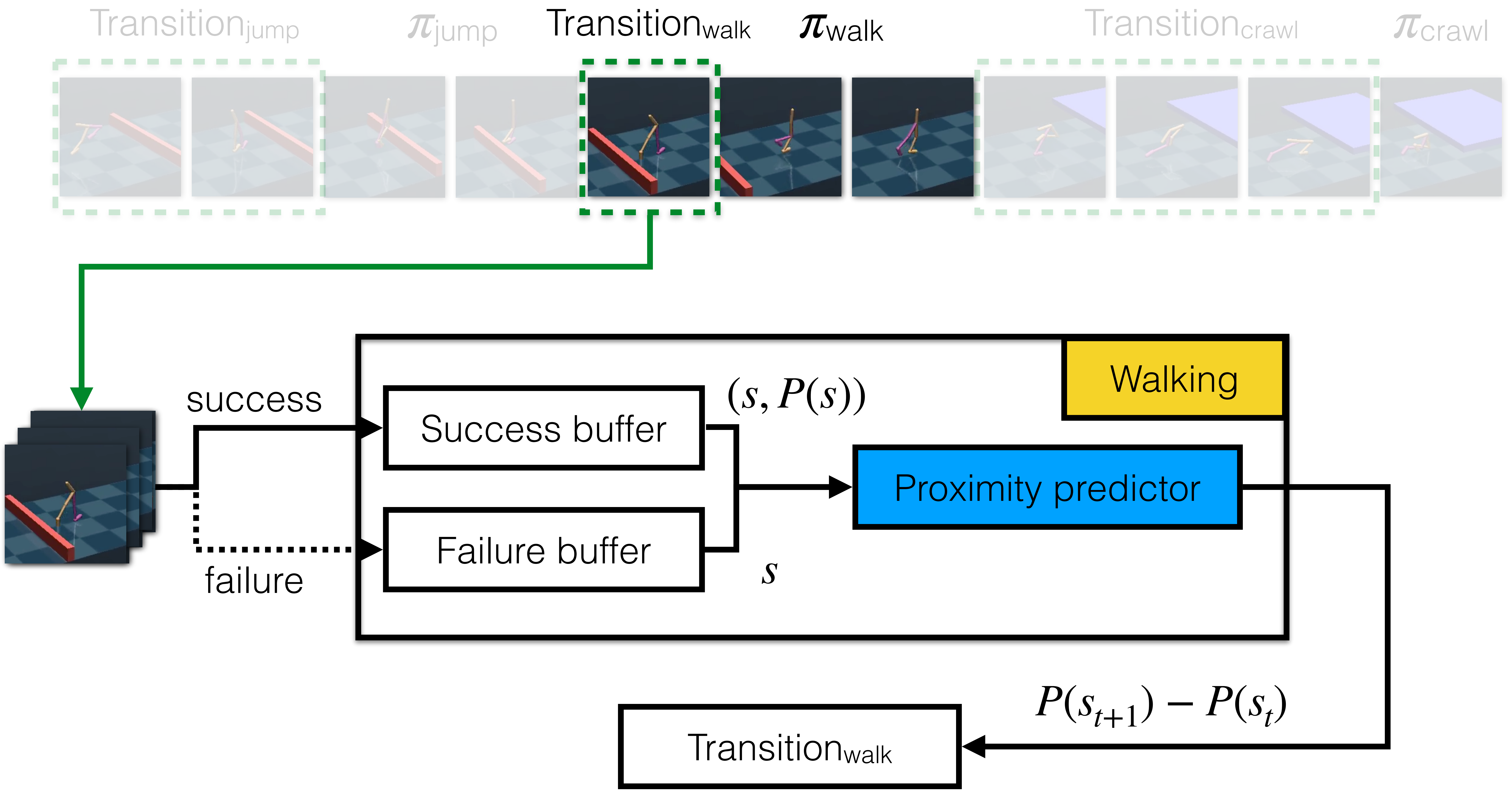
Learning proximity predictor



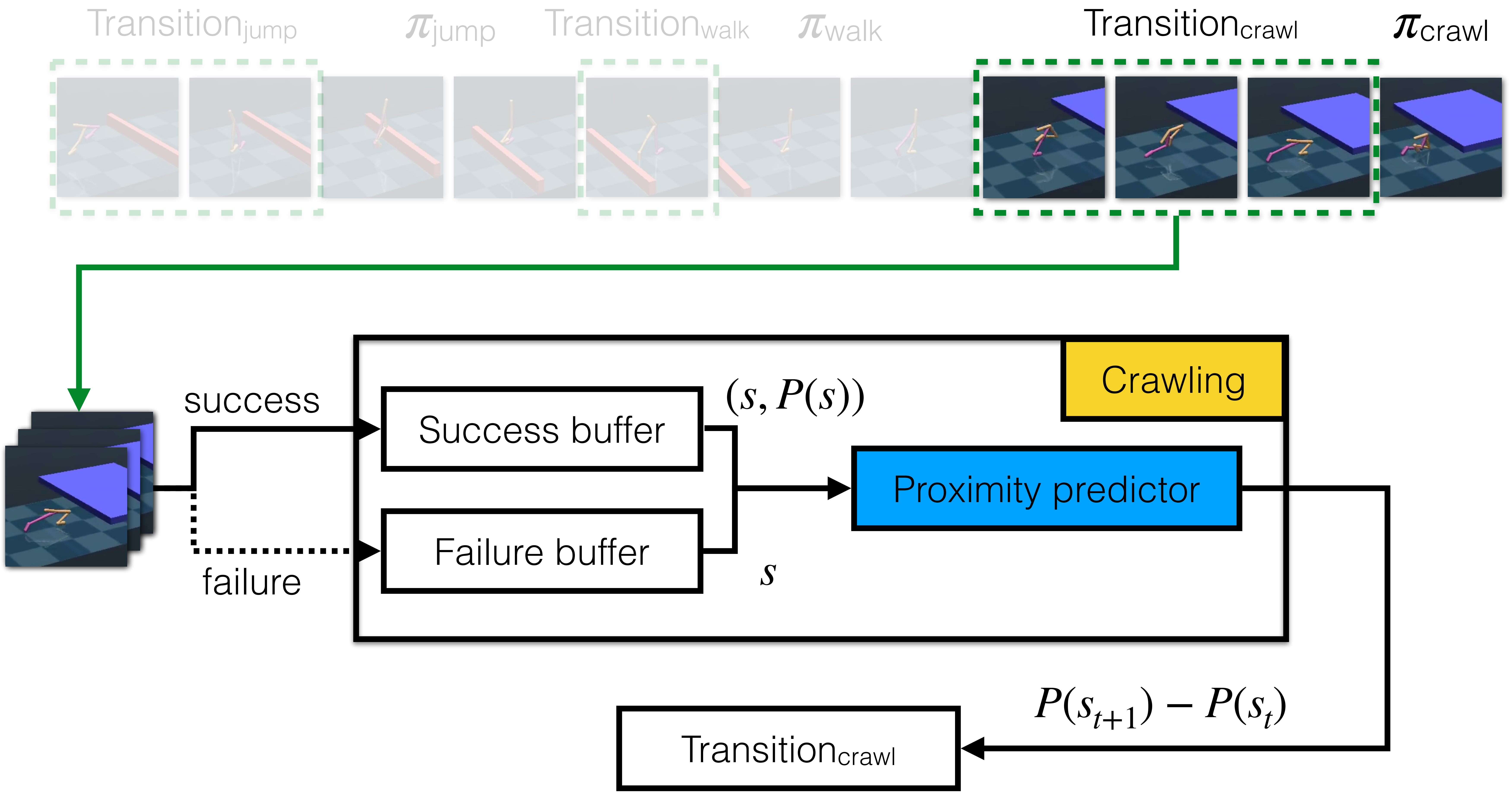
Learning proximity predictor



Learning proximity predictor



Learning proximity predictor

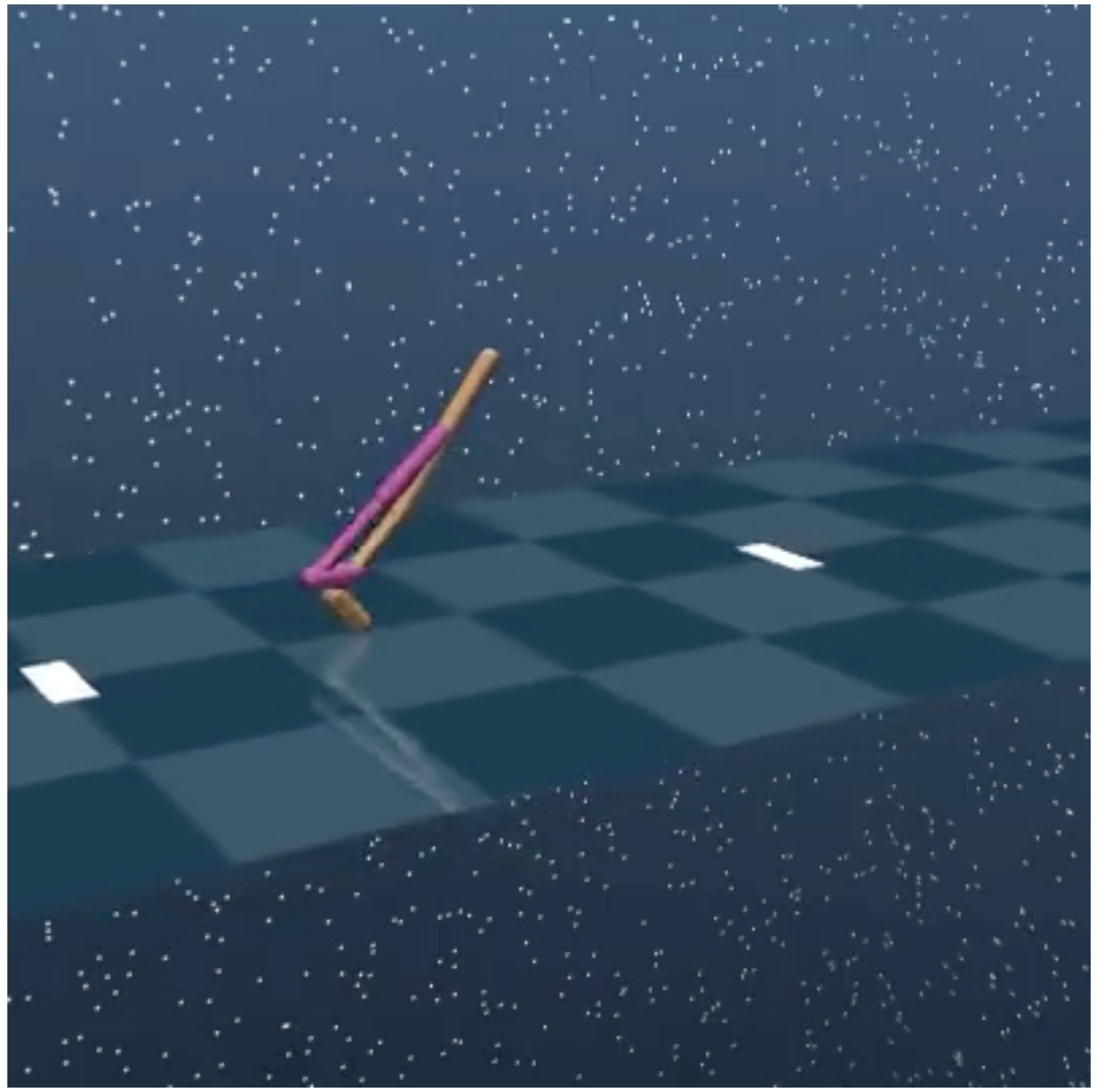


Patrol

Walk Forward

Transition

Walk Backward

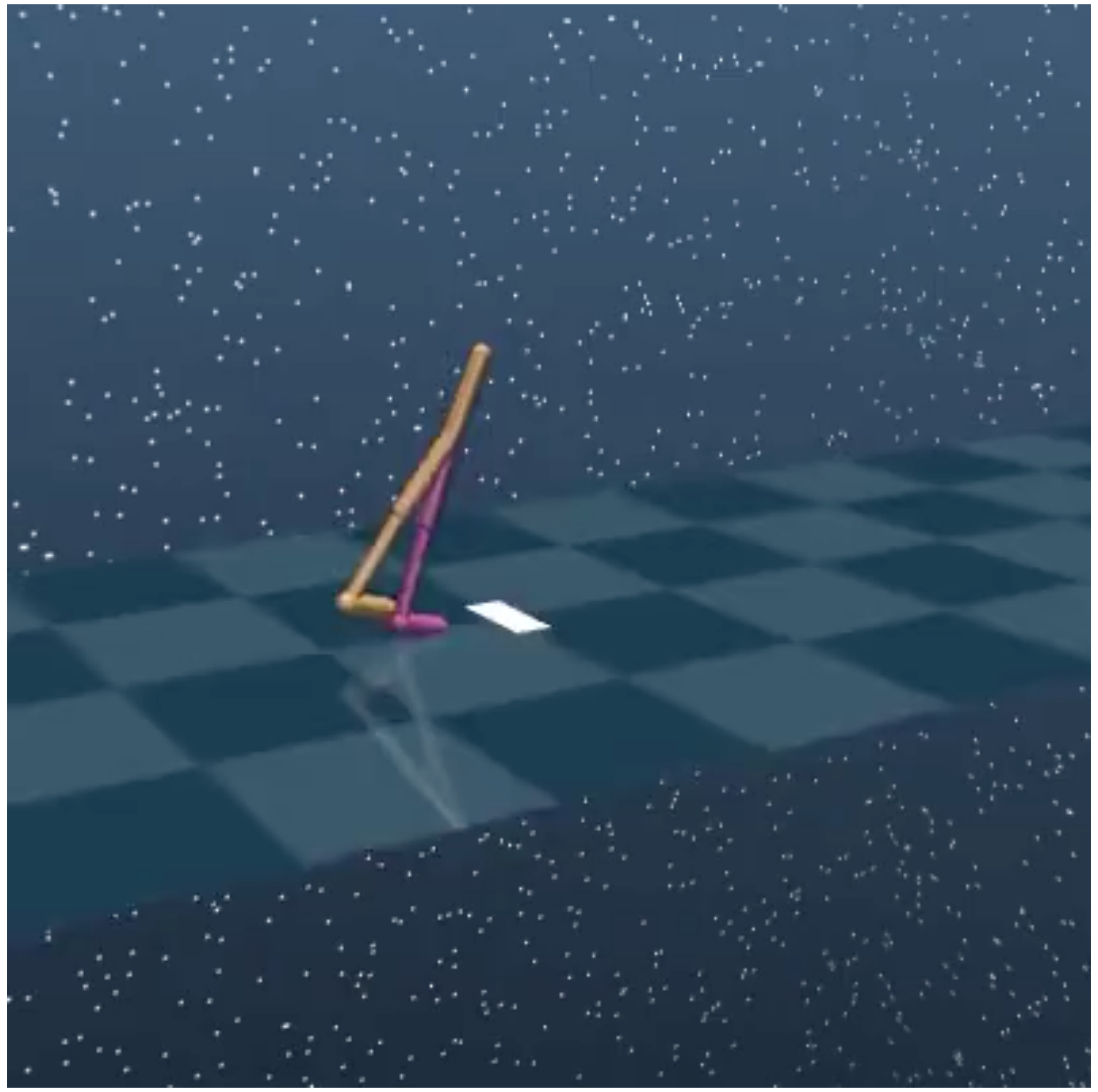


Patrol

Walk Forward

Transition

Walk Backward

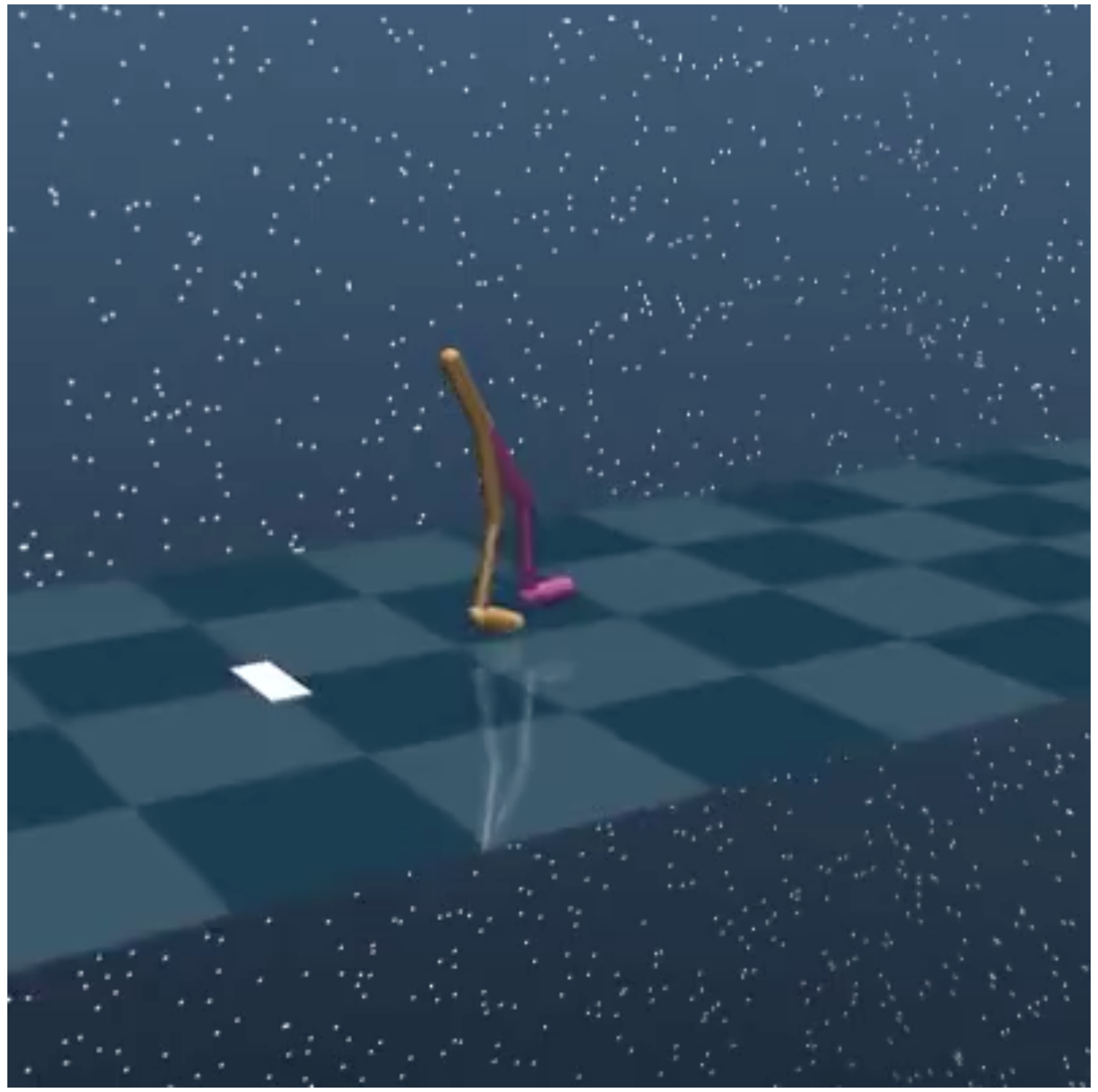


Patrol

Walk Forward

Transition

Walk Backward

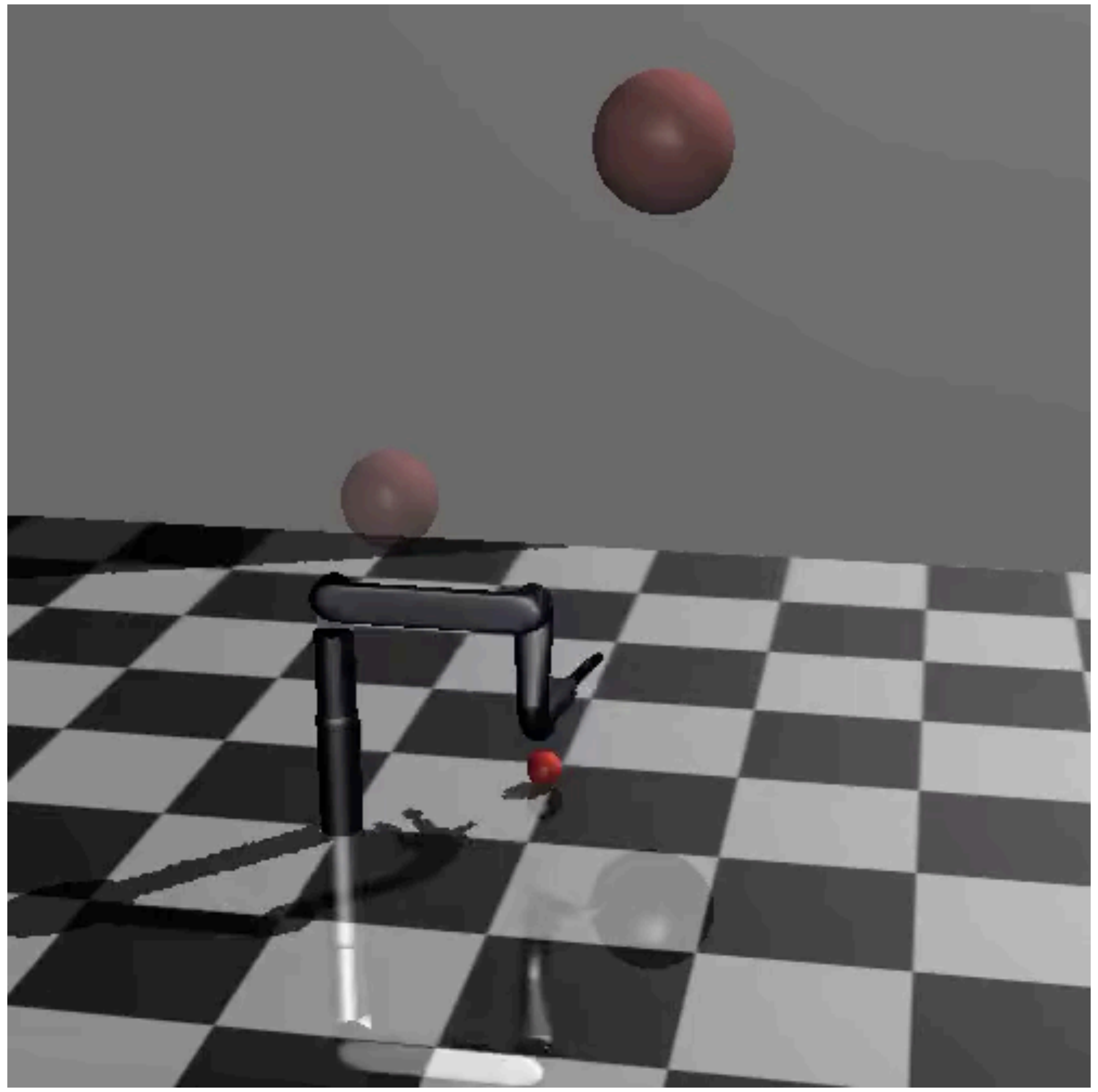


Serve (Toss & Hit)

Toss

Transition

Hit



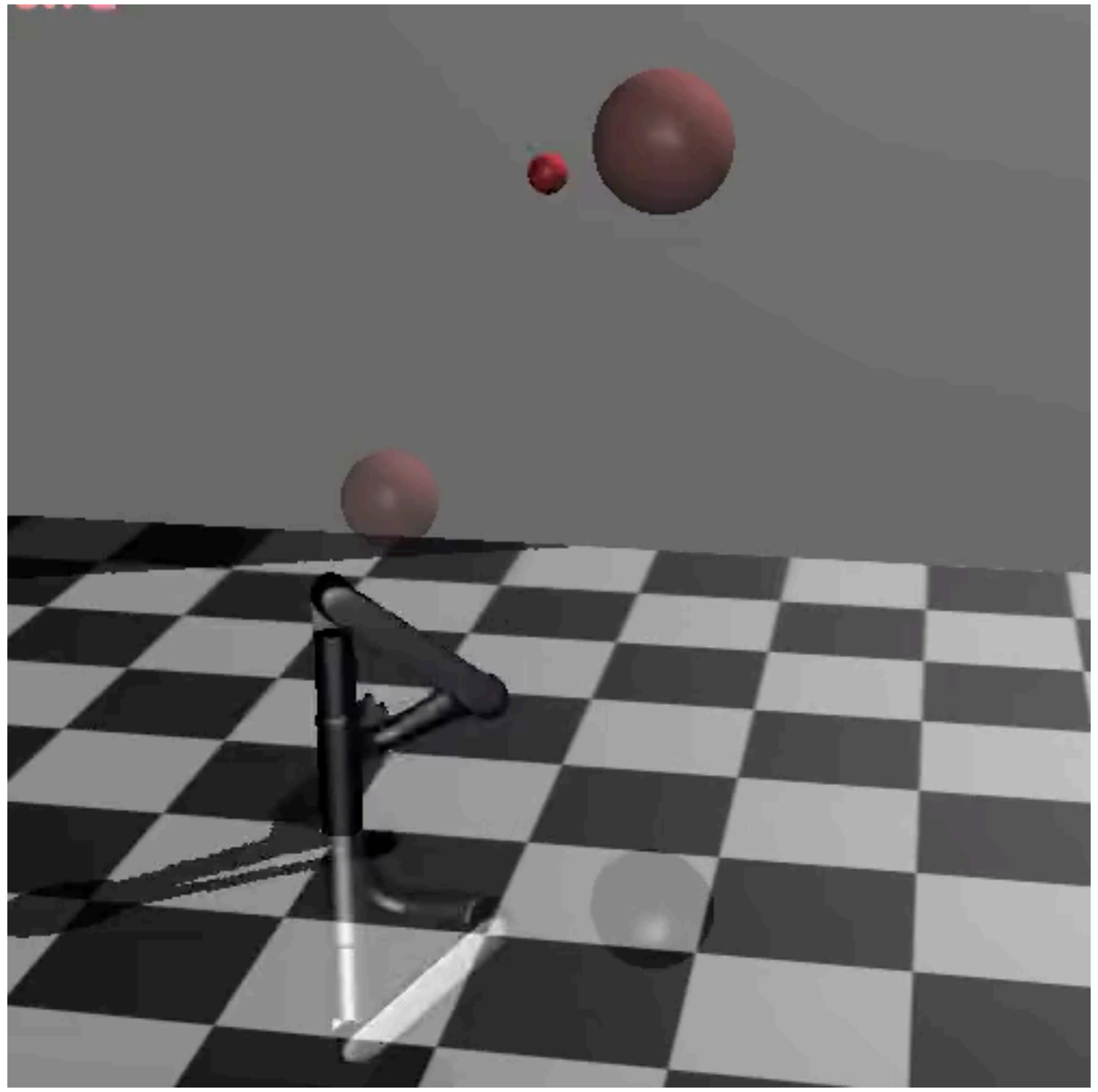
Lee et al. *“Composing Complex Skills by Learning Transition Policies”* ICLR 2019

Serve (Toss & Hit)

Toss

Transition

Hit

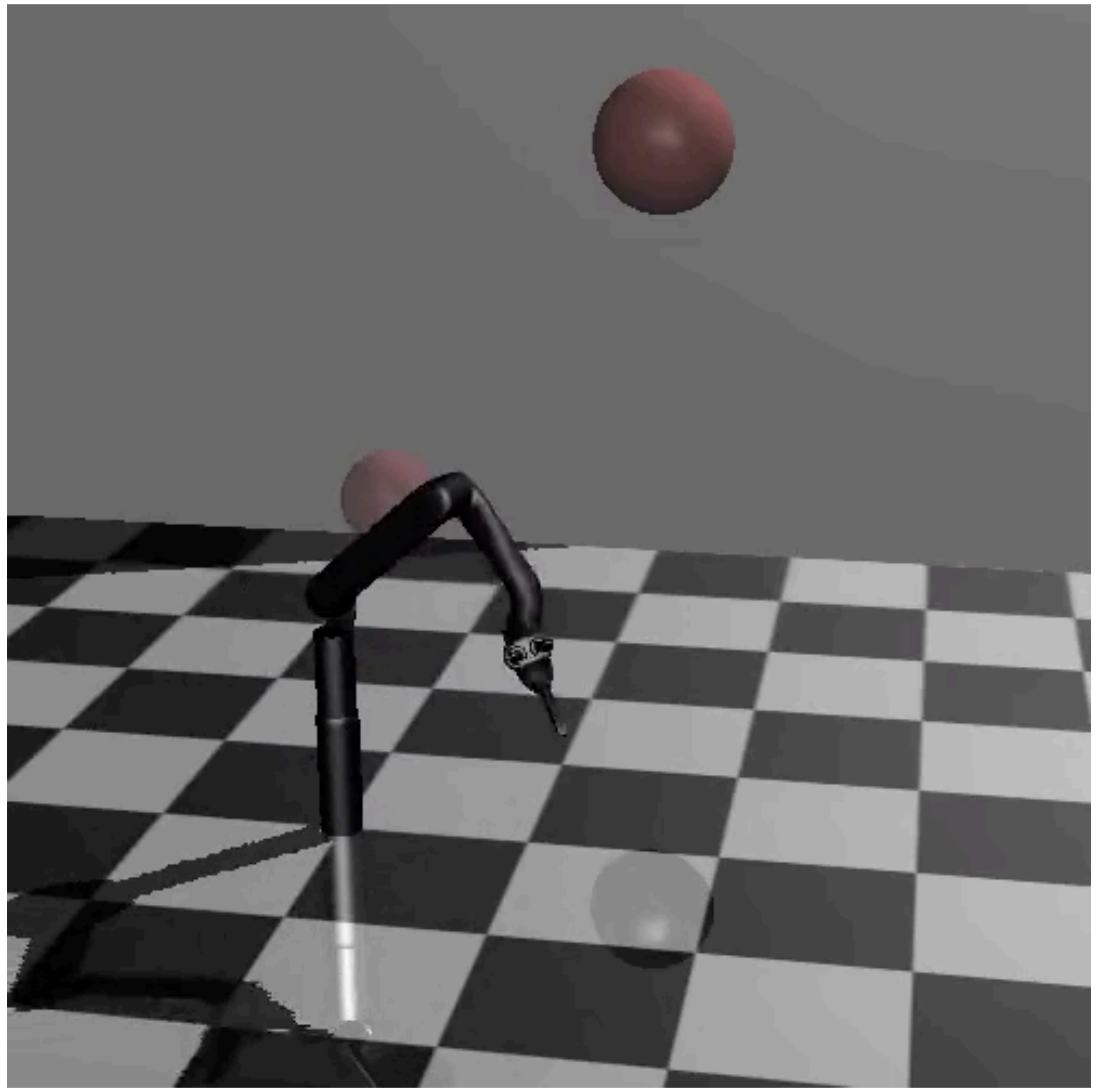


Serve (Toss & Hit)

Toss

Transition

Hit



Summary — Transition policy

We propose to **reuse skills** to compose **complex, long-horizon tasks**.

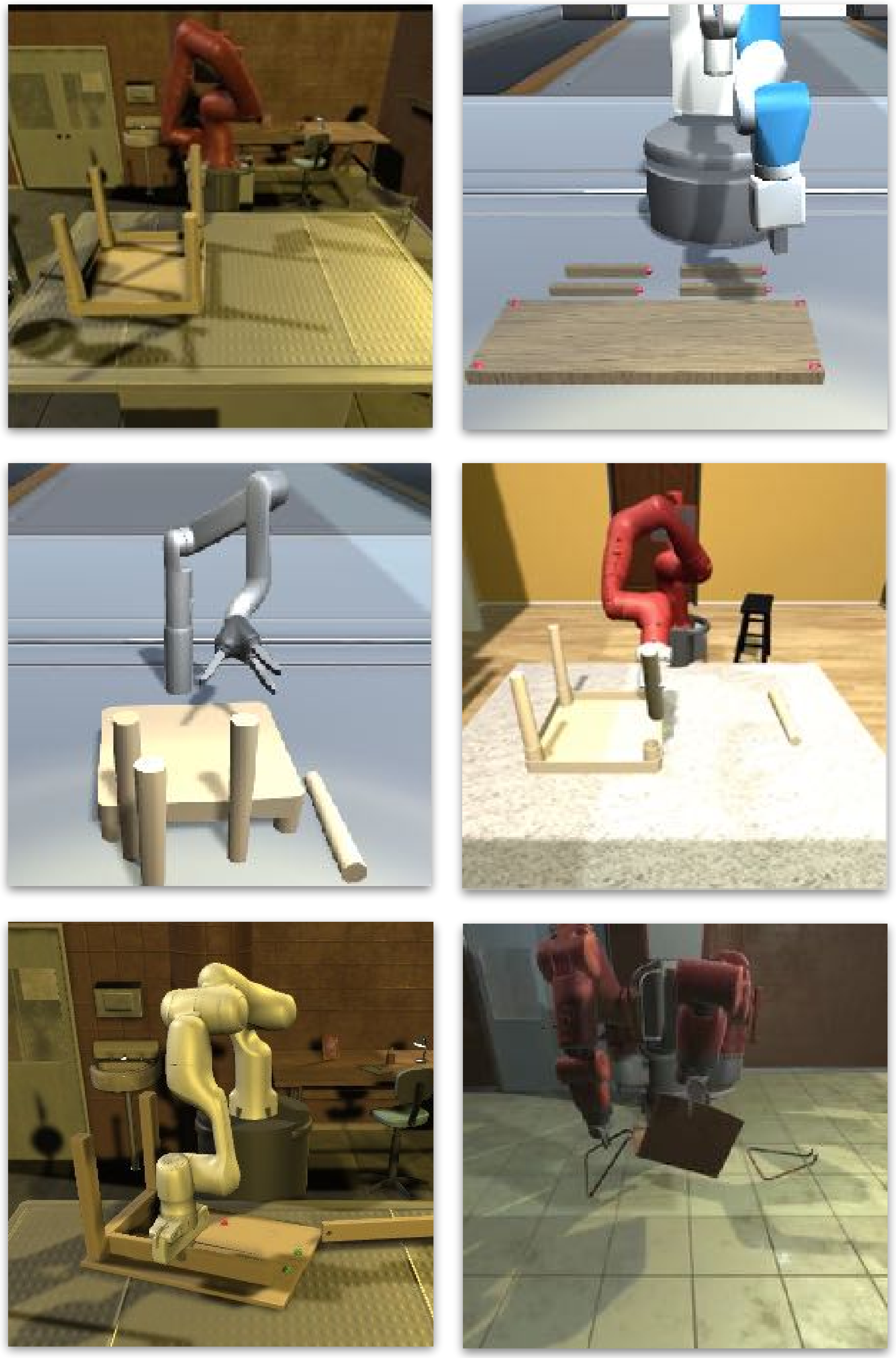
Naive execution of skills fails since the skills never learned to connect.

Transition policies learn to smoothly connect skills.

Proximity predictors provide dense reward for efficient training of transition policies.

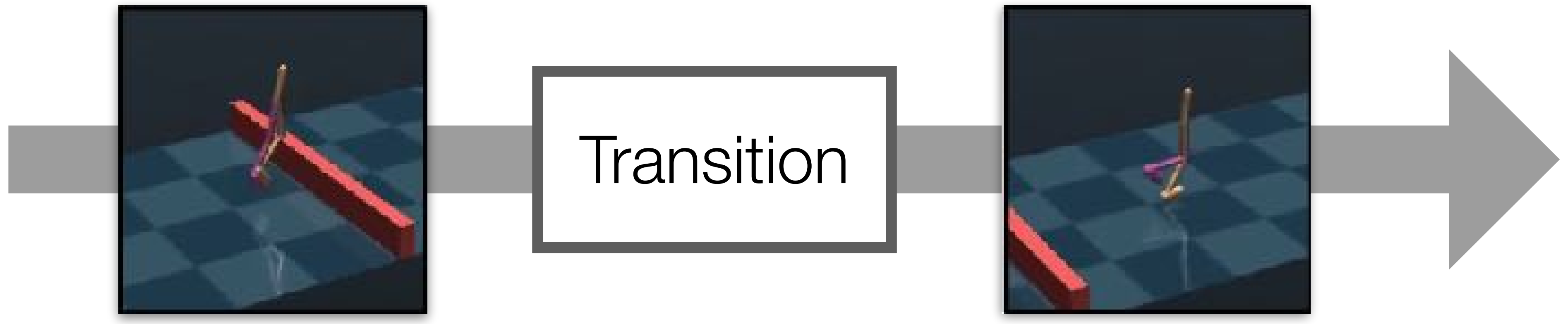
Lee et al. *"Composing Complex Skills by Learning Transition Policies"* ICLR 2019

Scaling robot learning to long-horizon tasks

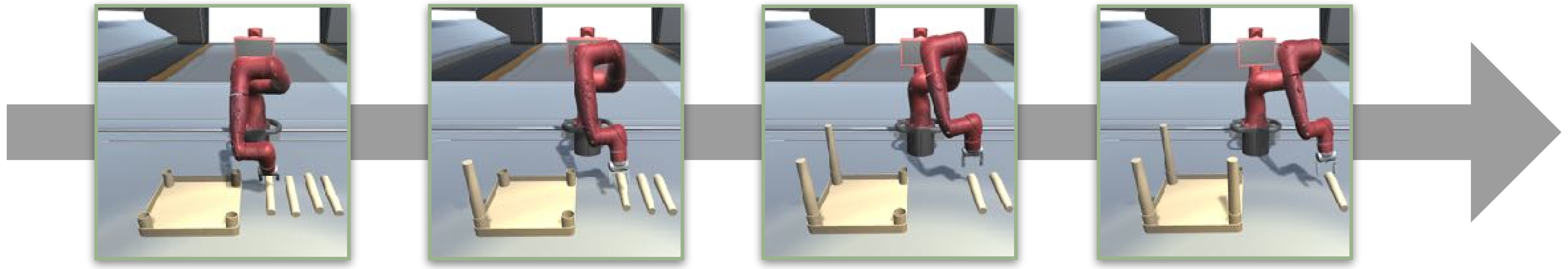


IKEA furniture assembly [ICRA'21]

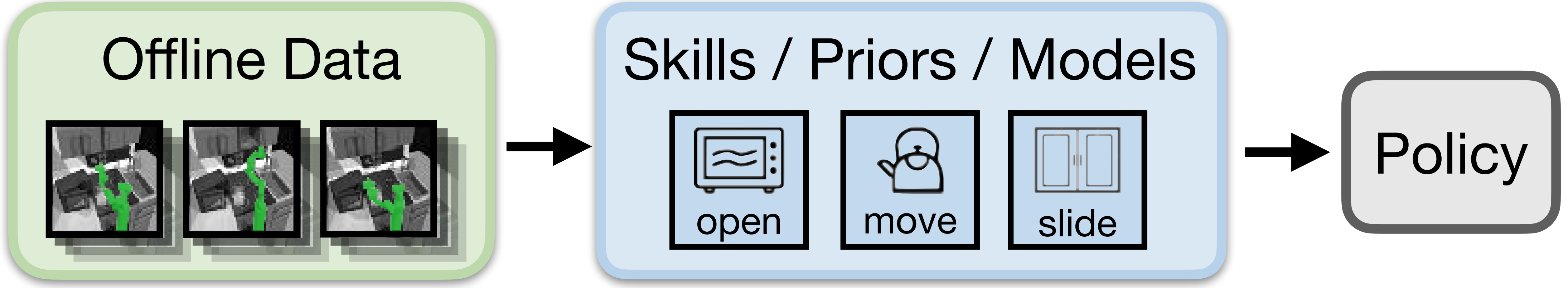
Benchmark



Transition policy [ICLR'19]



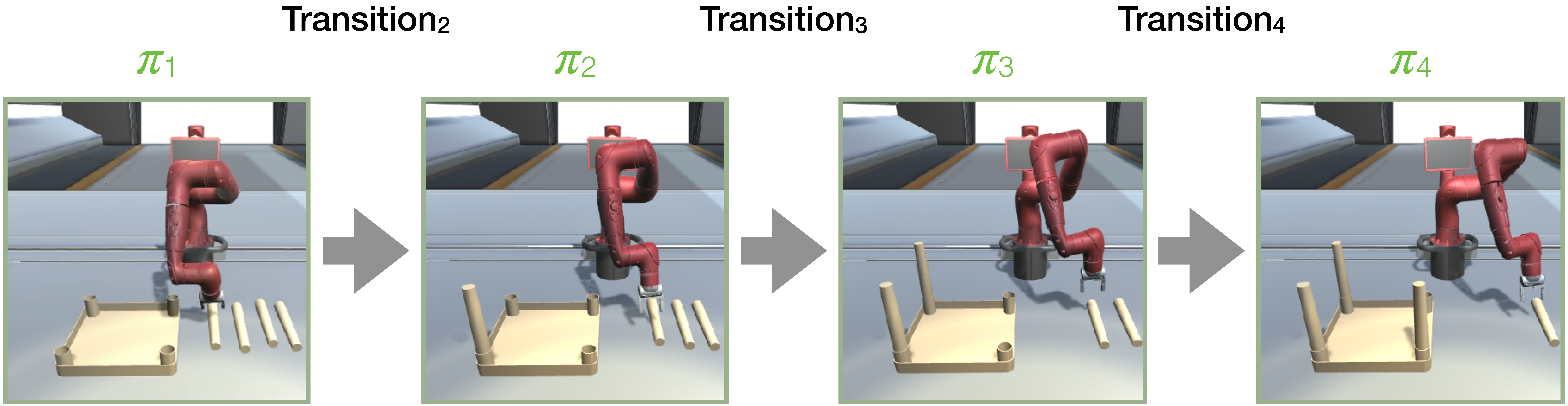
Skill chaining via fine-tuning [CoRL'21]



Skill-based RL [CoRL'20, CoRL'21, CoRL'22]

Algorithm

Let's try furniture assembly!

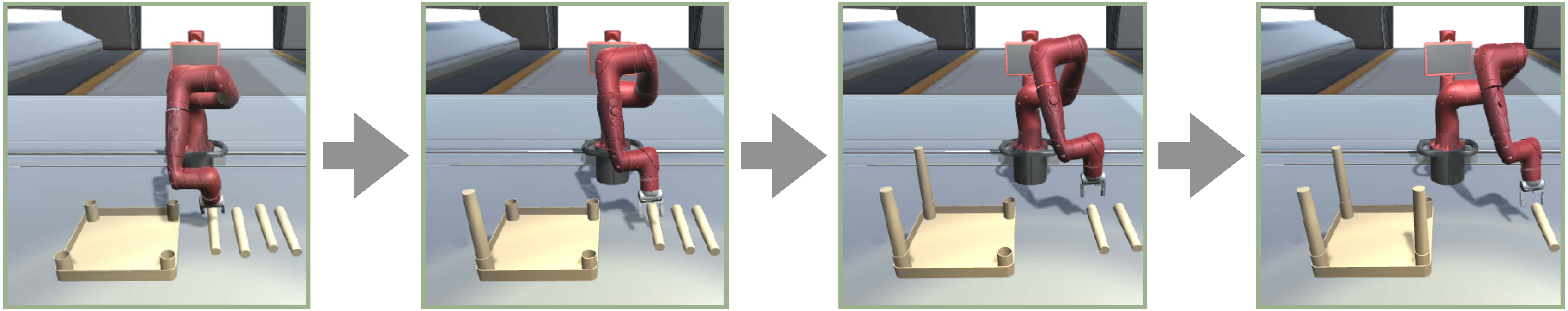


Lee et al. *"Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization"* CoRL 2021

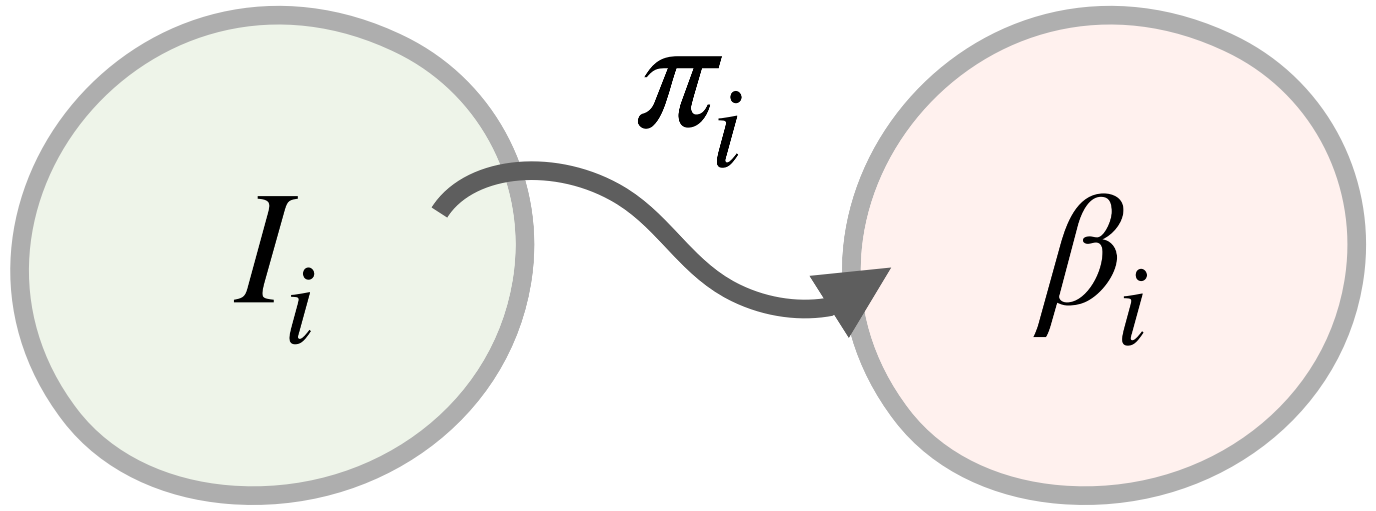
Let’s try furniture assembly!

~~Transition$_2$~~ ~~Transition$_3$~~ ~~Transition$_4$~~

π_1 π_2 π_3 π_4



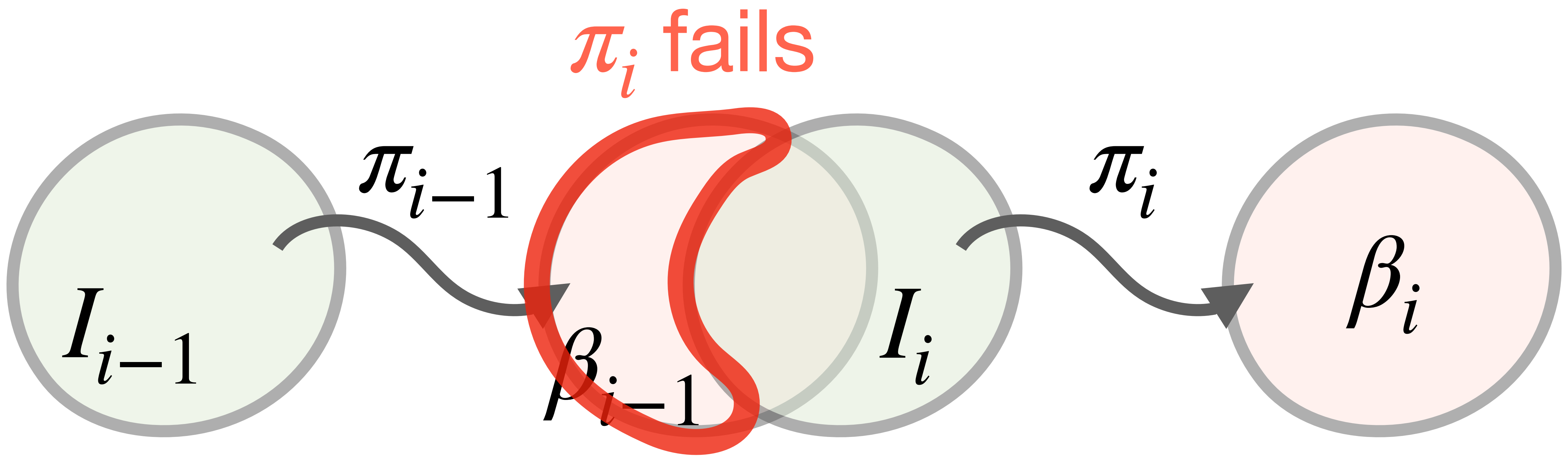
Lee et al. *“Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization”* CoRL 2021



Initiation set I_i: all successful **initial** states of π_i

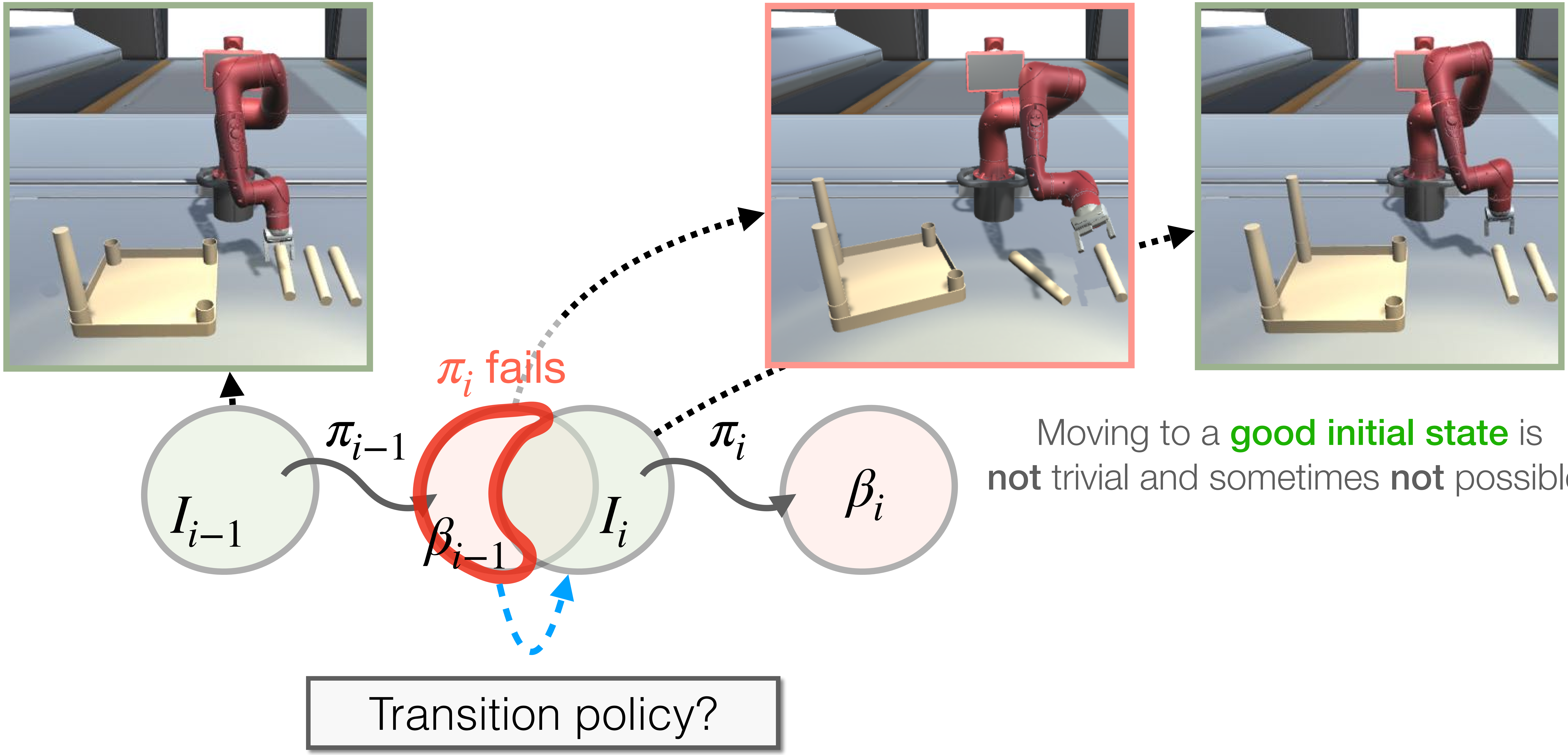
Termination set β_i: successful **terminal** states of π_i

Lee et al. *"Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization"* CoRL 2021



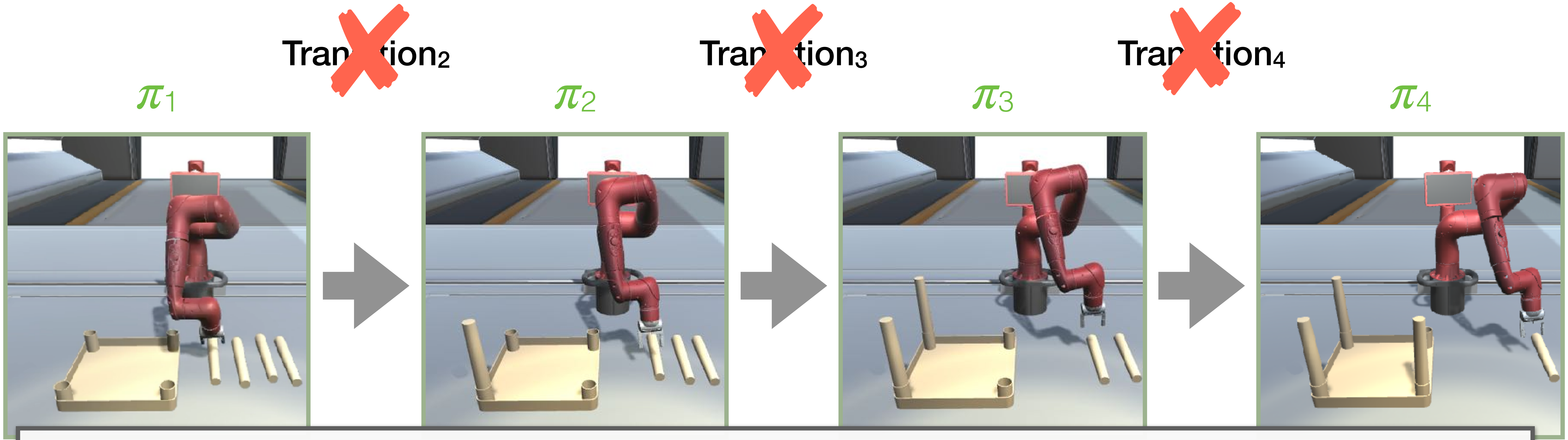
Initiation set I_i: all successful **initial** states of π_i

Termination set β_i: successful **terminal** states of π_i



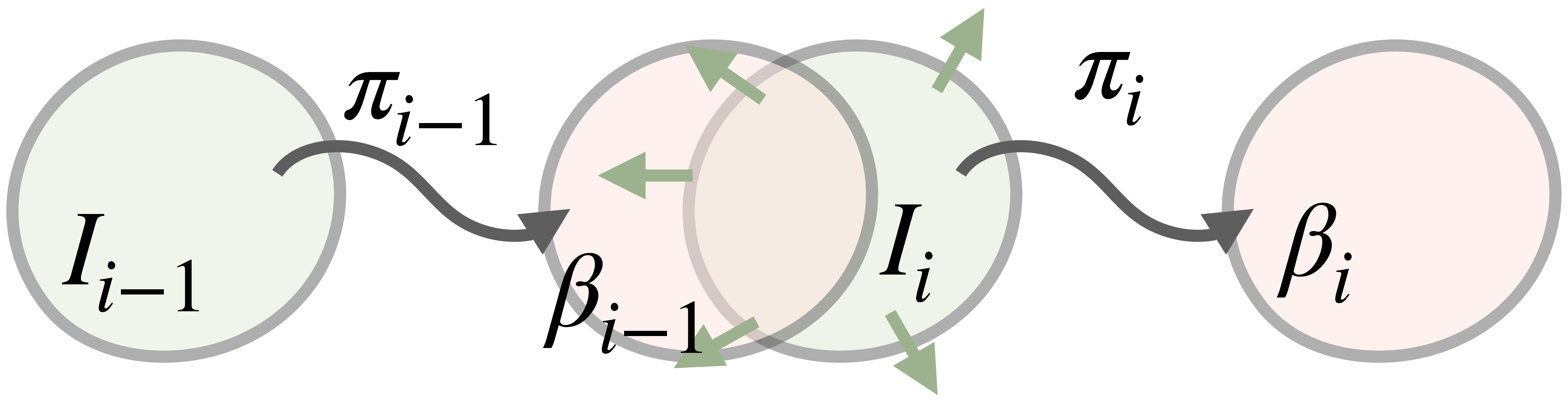
Moving to a **good initial state** is **not** trivial and sometimes **not** possible!

Let's try furniture assembly!



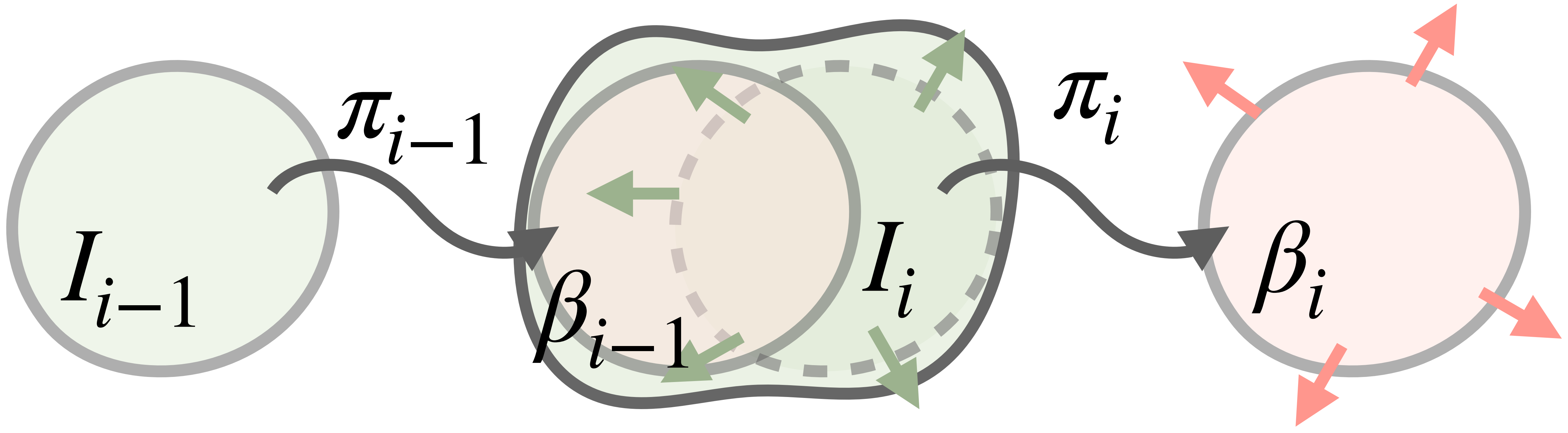
Transition policy cannot work with **irreversible** or **uncontrollable** states
(e.g. object inside gripper, attached objects)

Chaining skills



Increase I_i

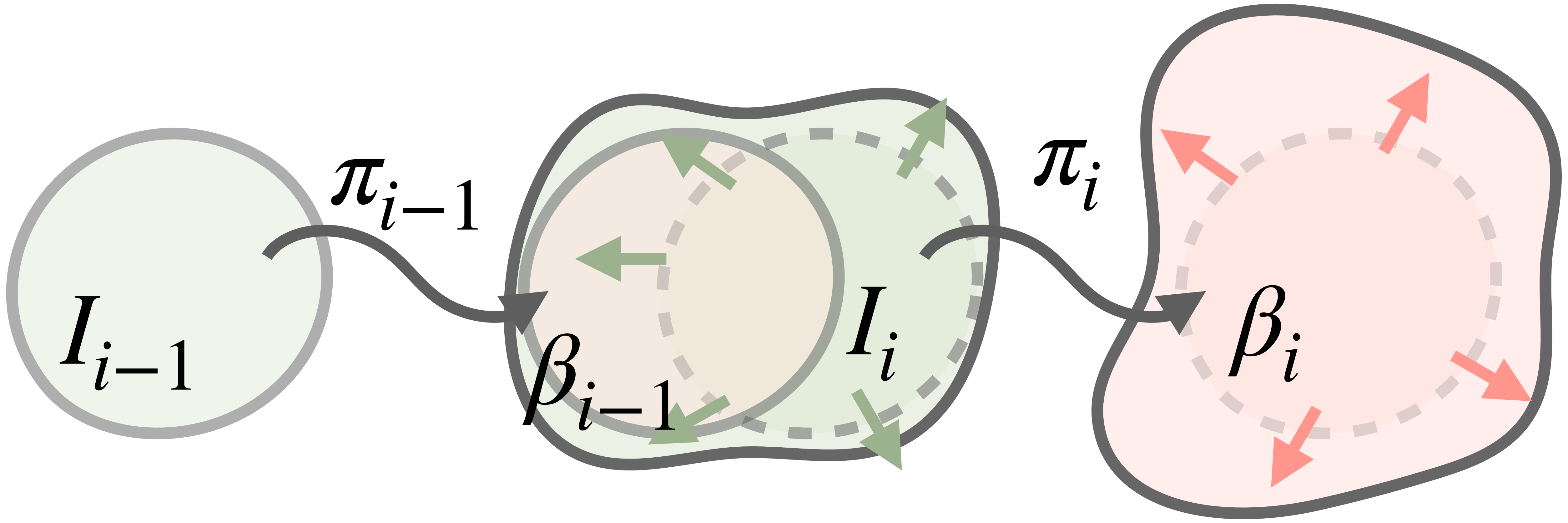
Chaining skills



Increase I_i

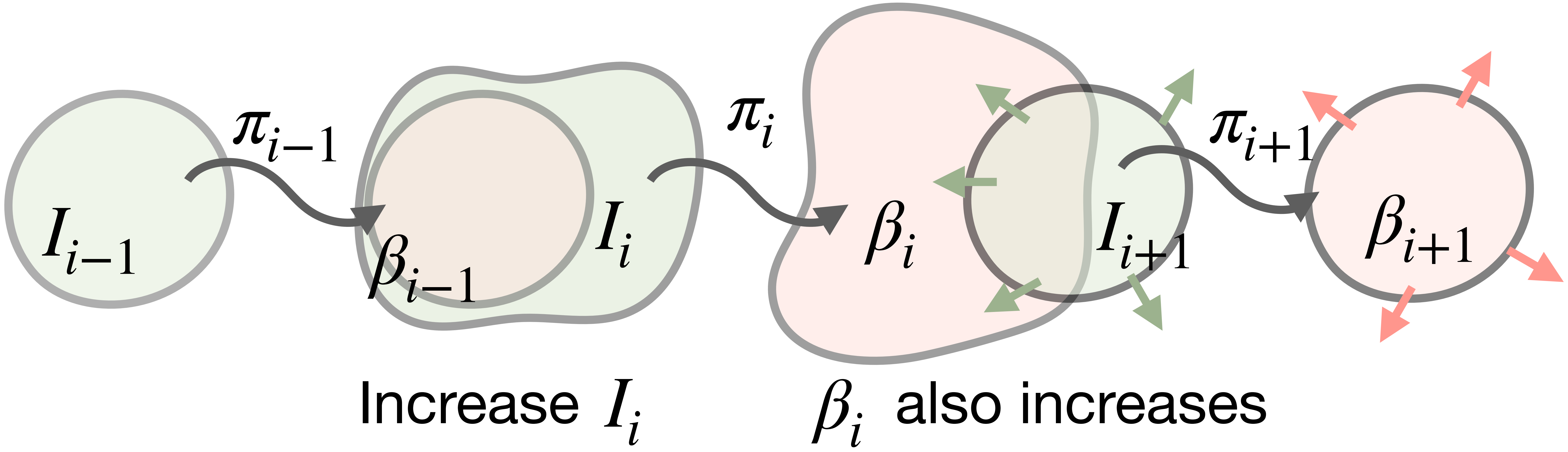
β_i also increases

Chaining skills



Increase I_i β_i also increases

Chaining more skills…



Chaining more skills…

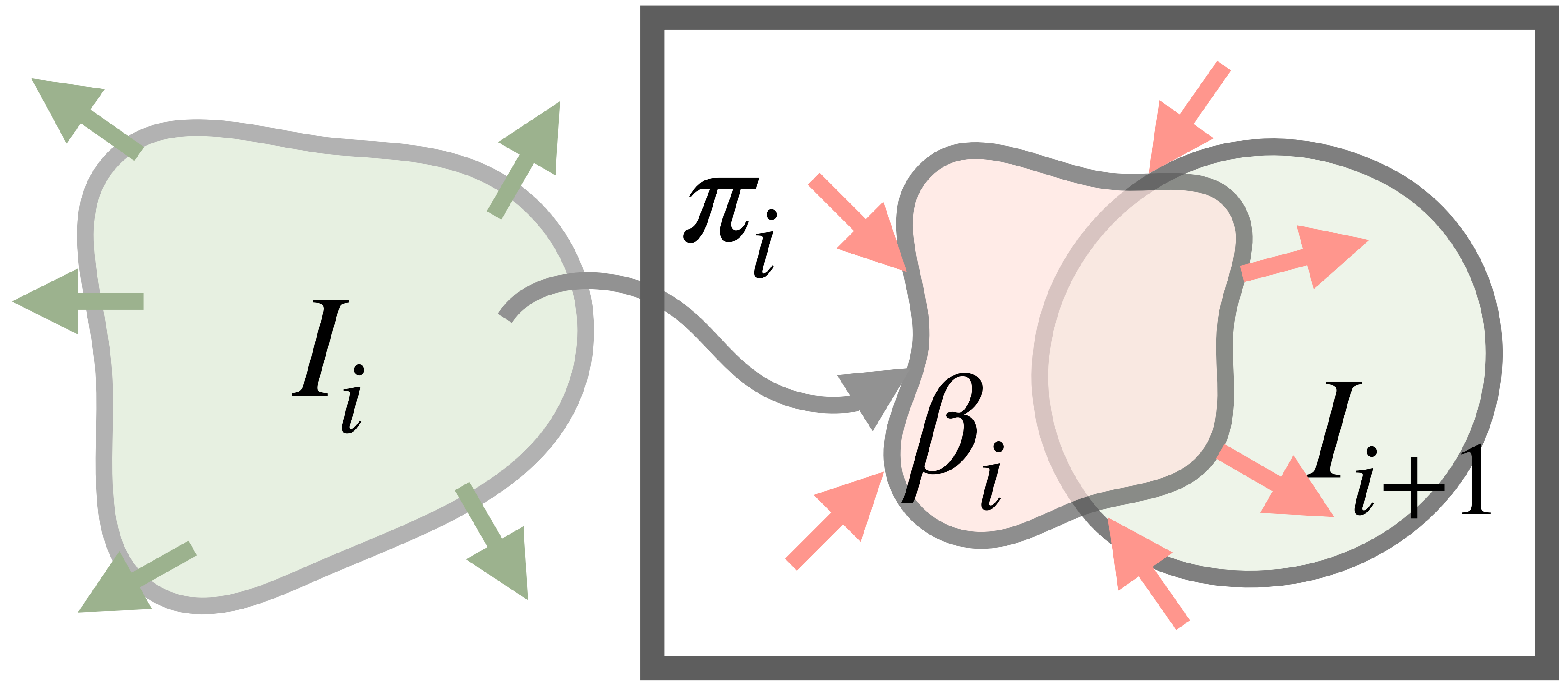
To chain more skills:

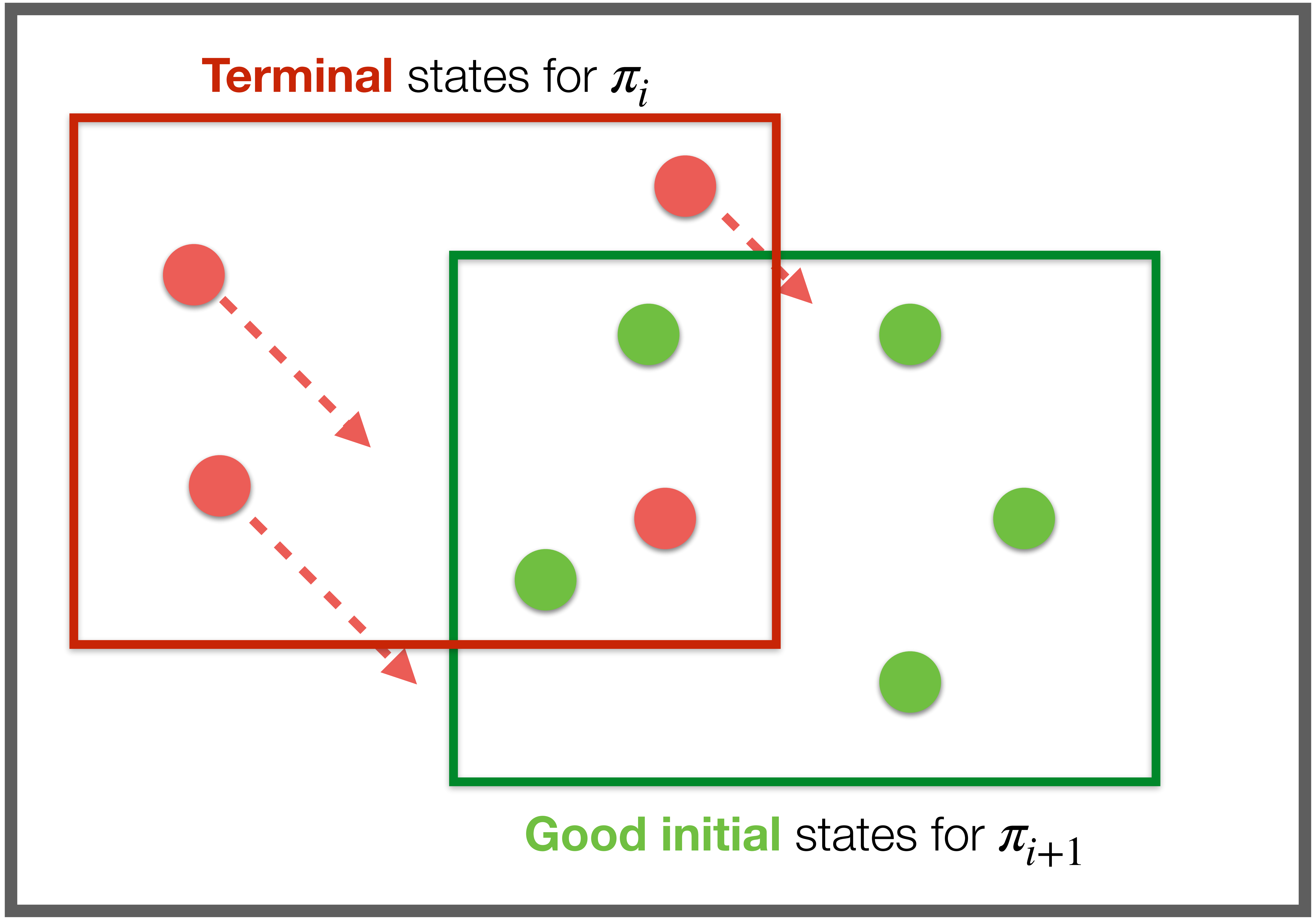
(1) **increase** initiation set, while

(2) keep termination set **small**

Terminal STAte Regularization (T-STAR)

Fine-tune π_i to cover larger I_i

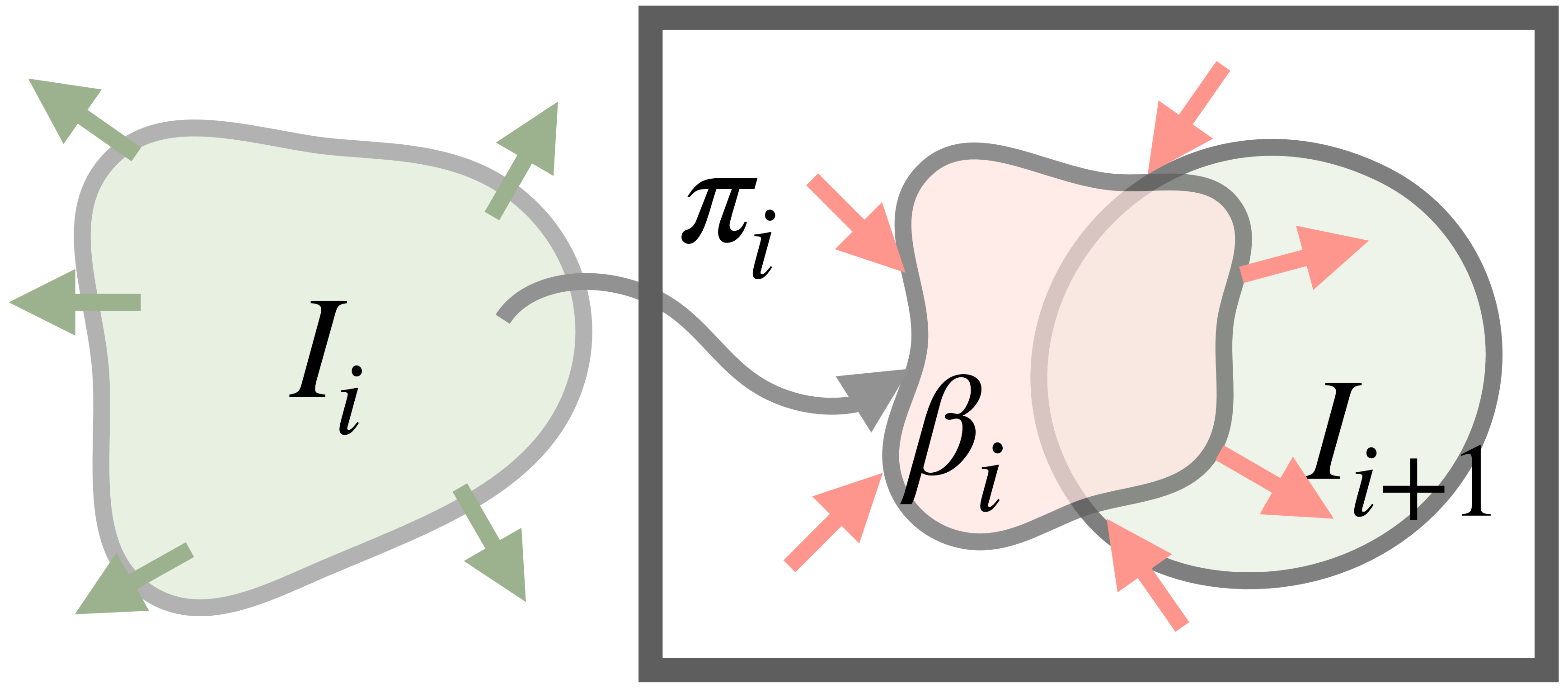




Lee et al. *"Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization"* CoRL 2021

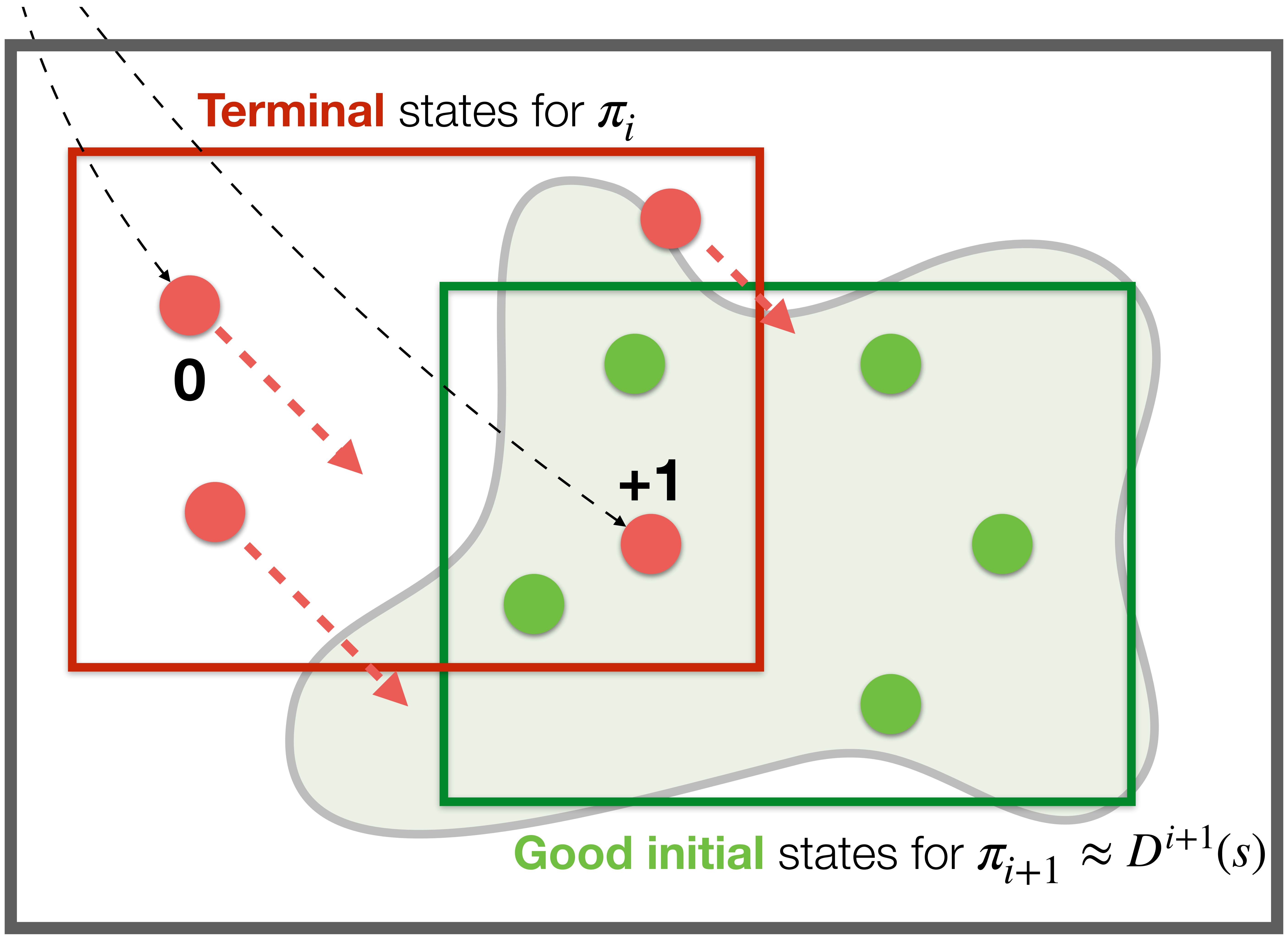
Terminal STAte Regularization (T-STAR)

Fine-tune π_i to cover larger I_i

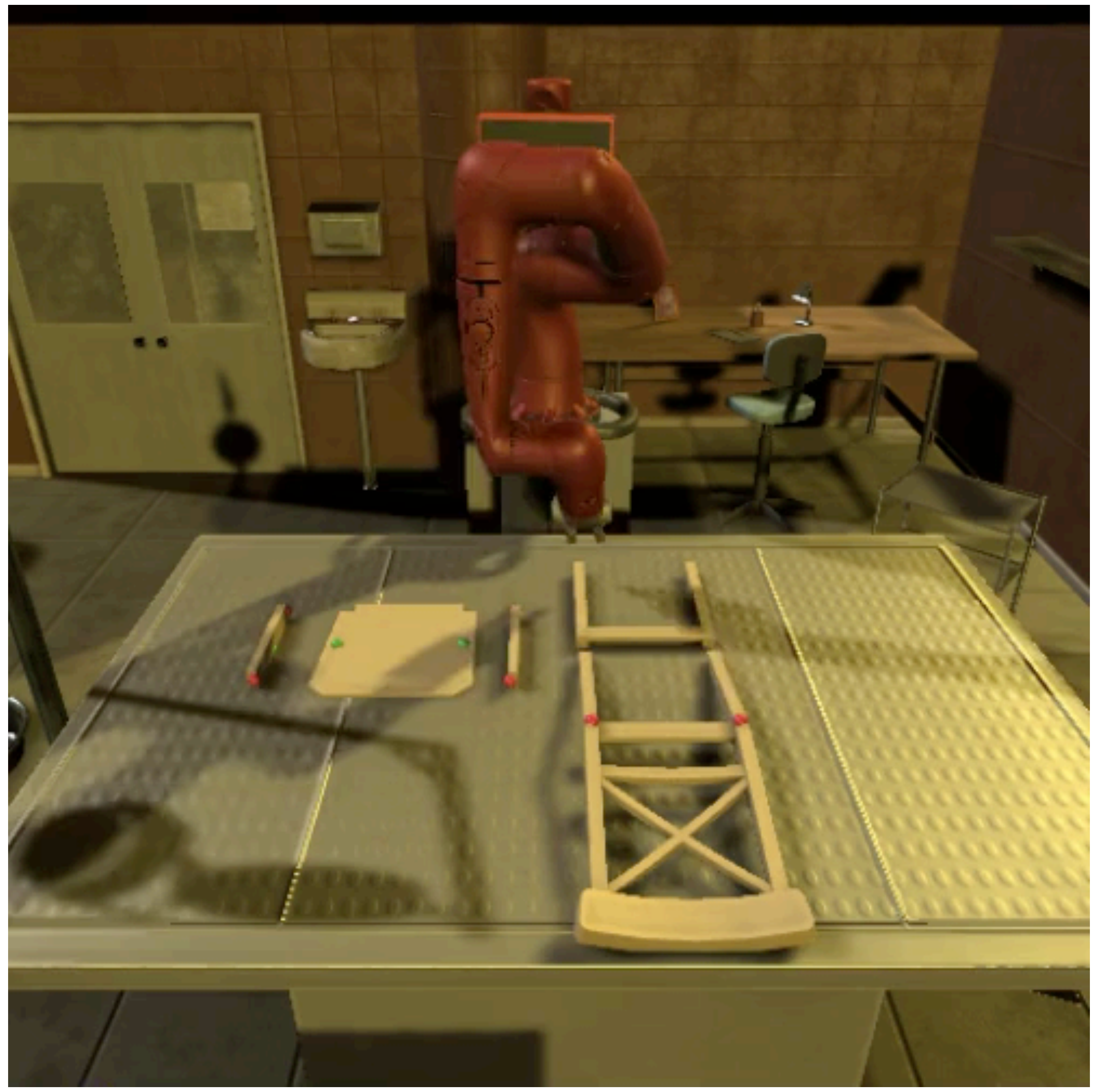


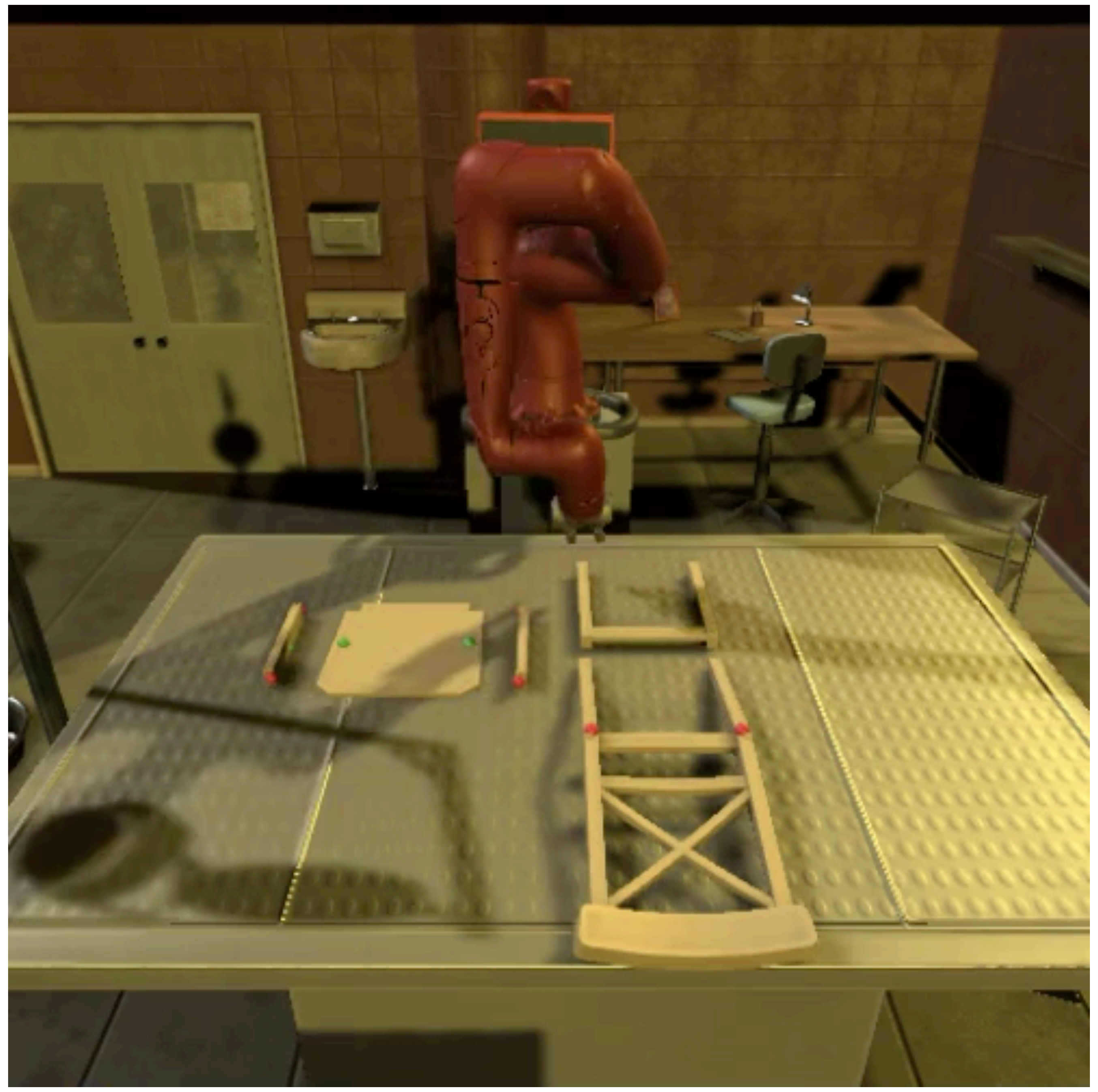
Reward π_i if a **terminal state** is close to I_{i+1}

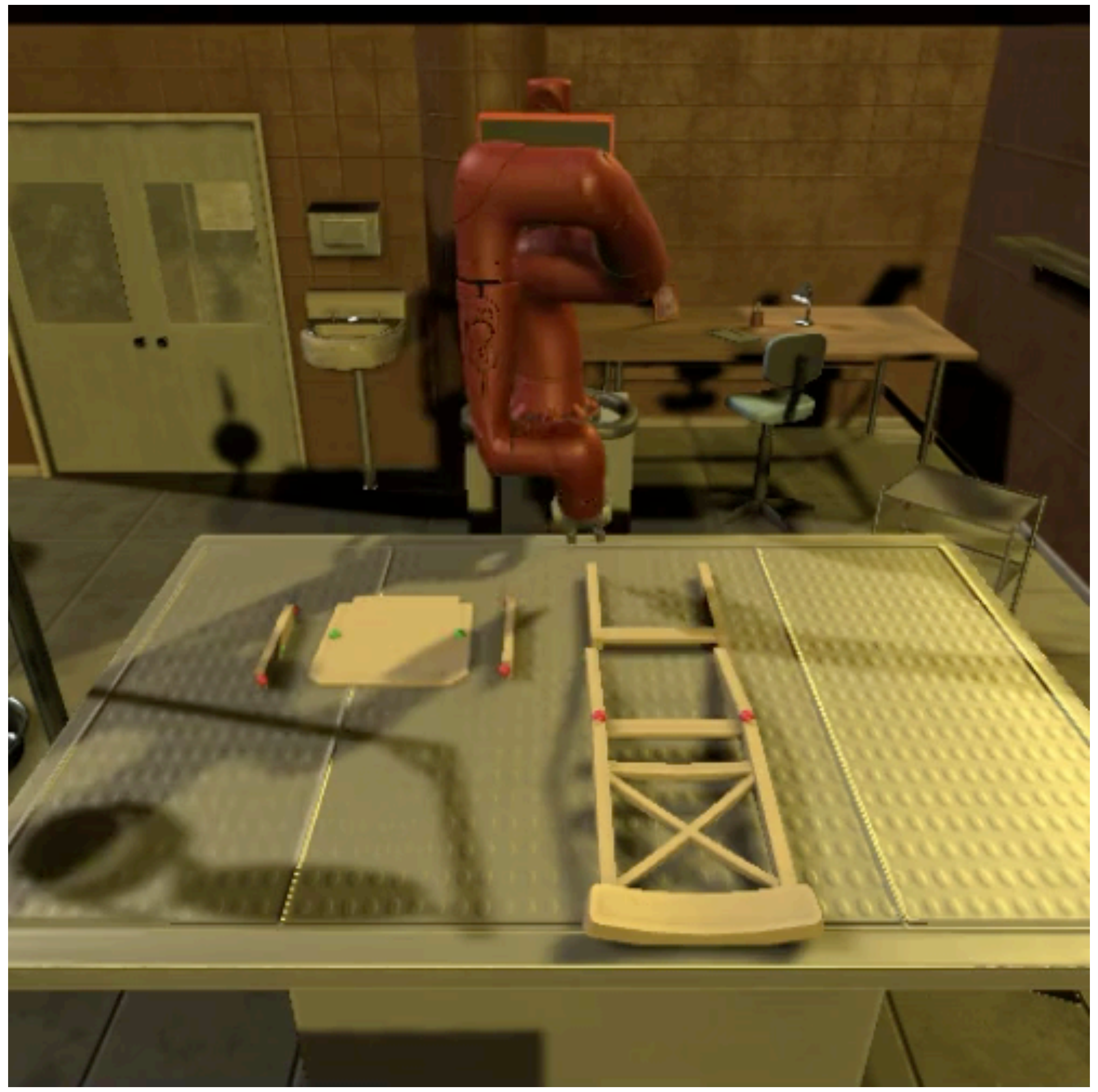
$$R^i_{TSR}(s) = 1_{s \in \beta_i} D^{i+1}(s)$$



Initiation set discriminator D^{i+1} distinguishes β_i and I_{i+1}







GAIL+PPO

Learning from scratch

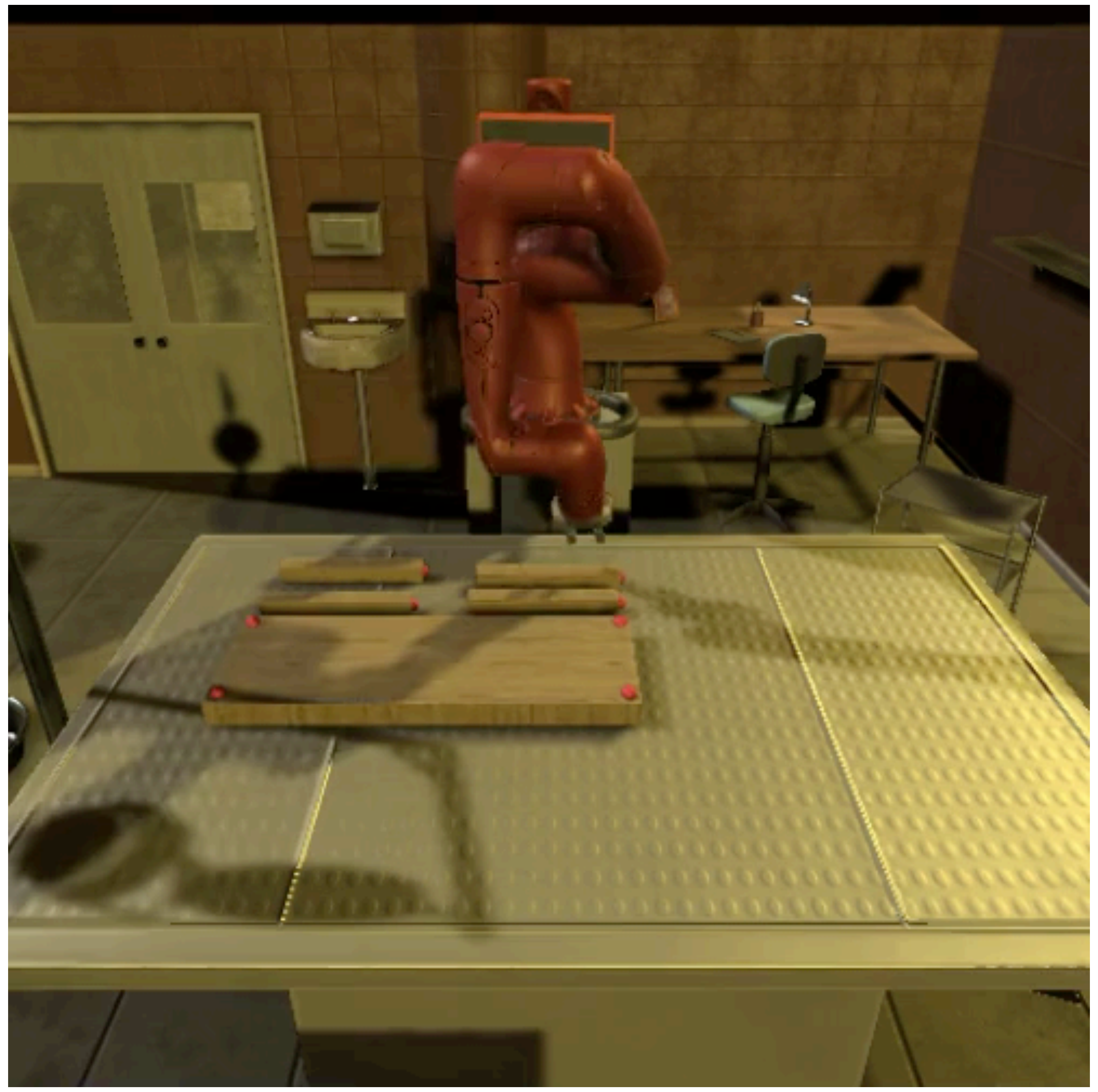
Clegg et al. 2018

Fine-tune skills **w/o** terminal state regularization

Ours

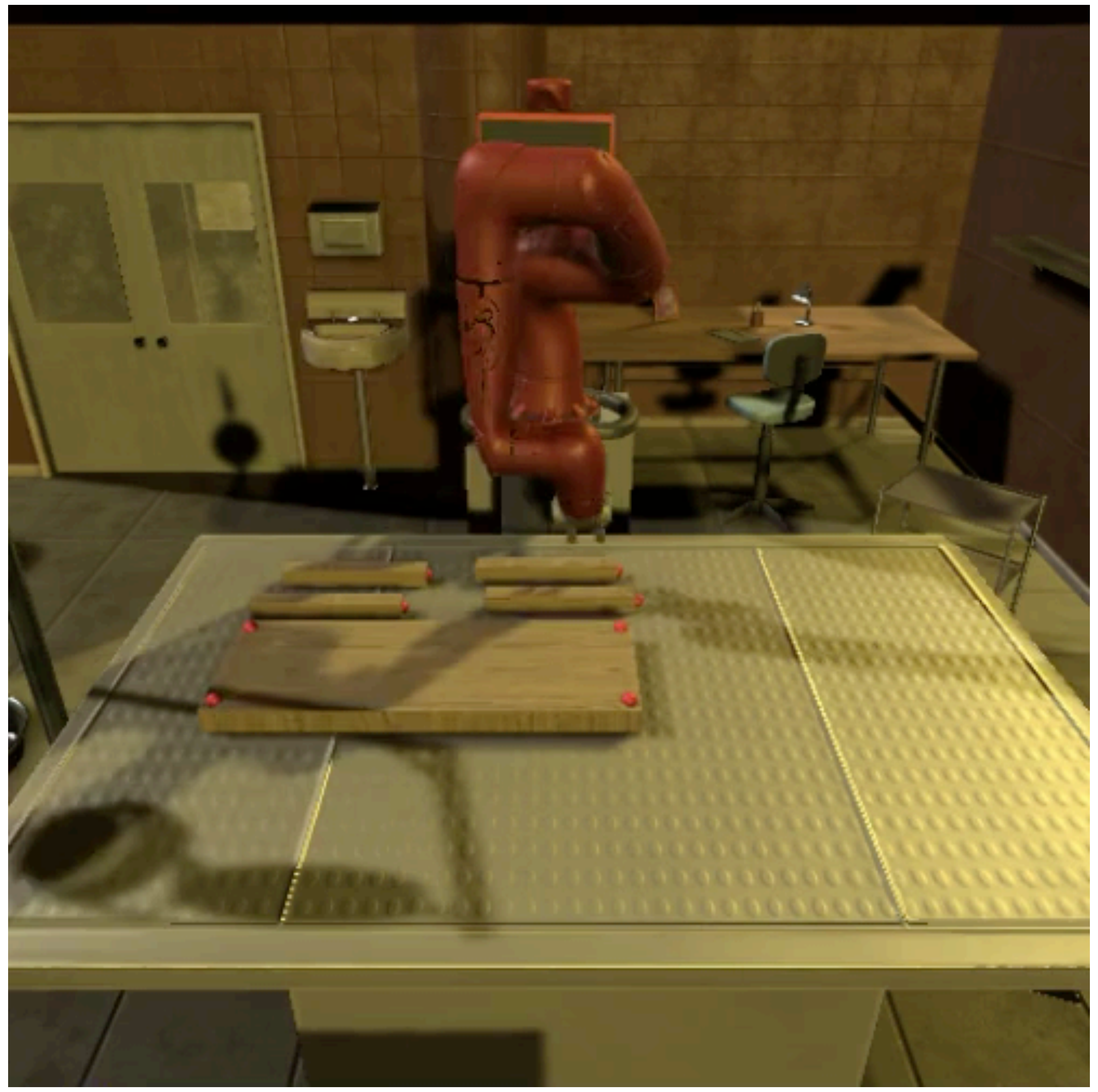
Fine-tune skills **with** terminal state regularization

Lee et al. *"Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization"* CoRL 2021



GAIL+PPO

Learning from scratch



Clegg et al. 2018

Fine-tune skills **w/o** terminal state regularization

Ours

Fine-tune skills **with** terminal state regularization

Lee et al. *"Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization"* CoRL 2021

Summary — T-STAR

We propose to **fine-tune skills** to compose **complex, long-horizon tasks**.

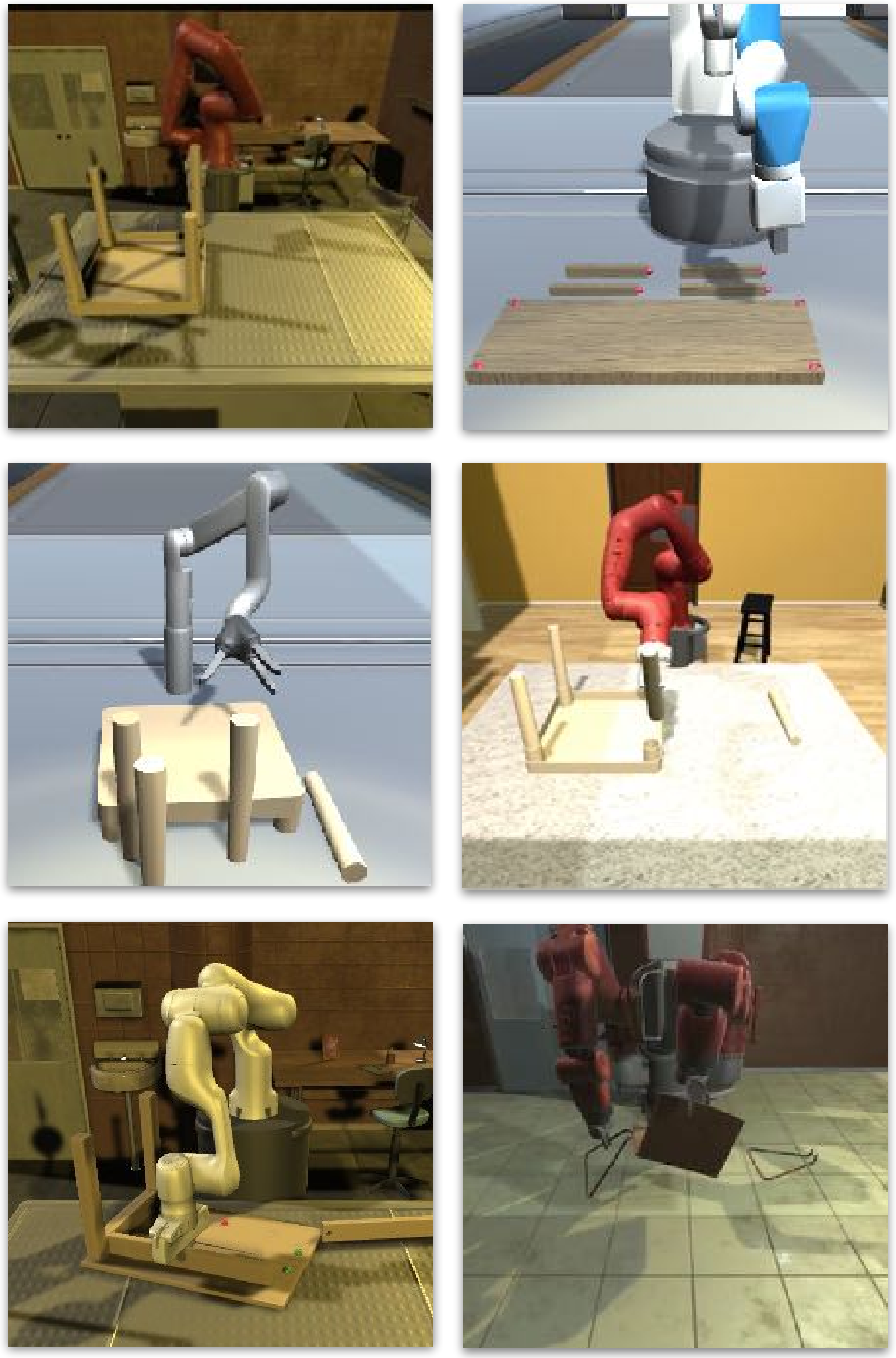
Naive fine-tuning of skills fails since the skills end up with very large terminal state distributions.

Terminal state regularization effectively encourage bounded terminal states.

Our approach solves the furniture assembly tasks.

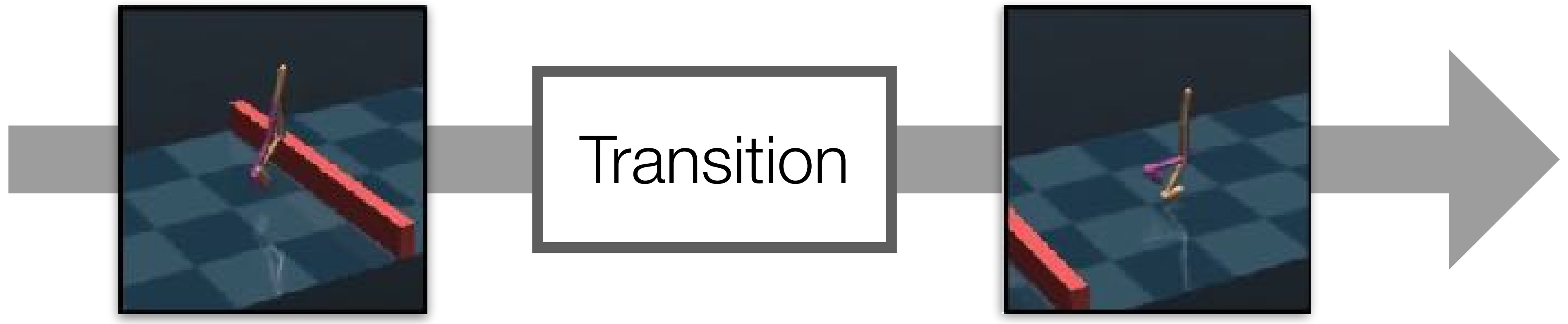
Lee et al. *"Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization"* CoRL 2021

Scaling robot learning to long-horizon tasks

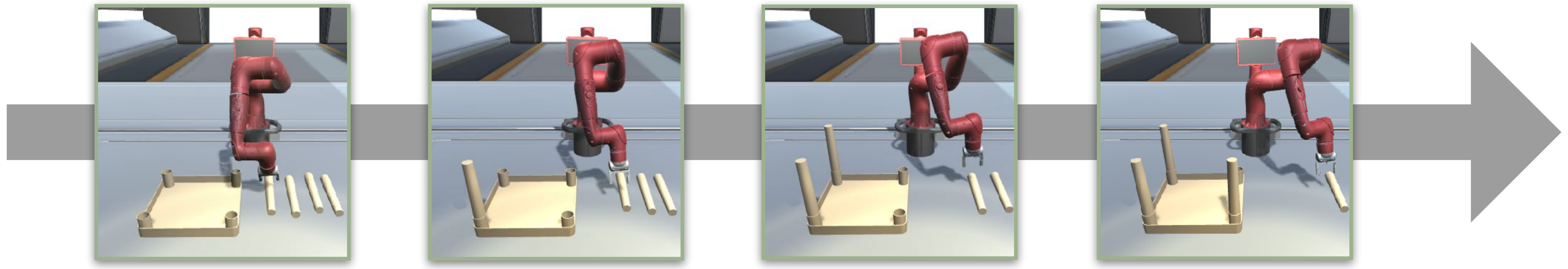


IKEA furniture assembly [ICRA’21]

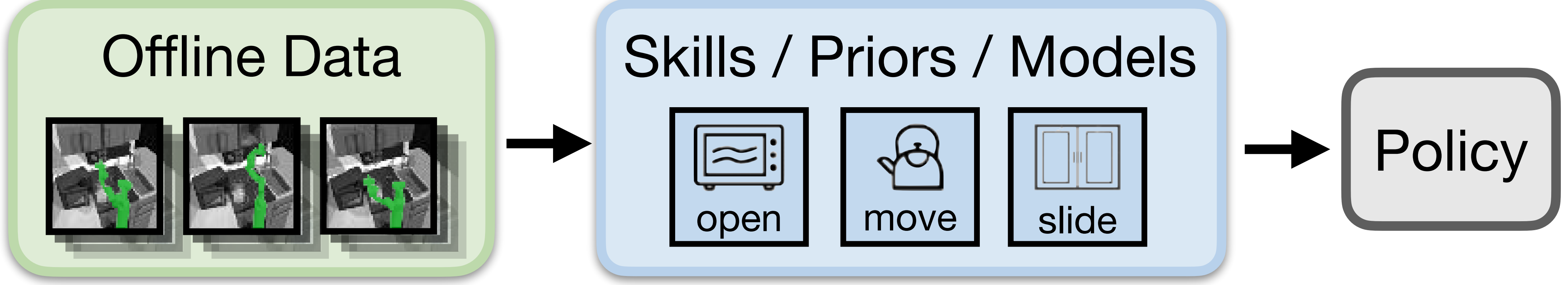
Benchmark



Transition policy [ICLR’19]



Skill chaining via fine-tuning [CoRL’21]



Skill-based RL [CoRL’20, CoRL’21, CoRL’22]

Algorithm

Skill composition with pre-defined skills

Jumping Walking Crawling

– Requires **manual definition** of skill repertoire

– Requires **skill execution order**

Can we learn skills from data?

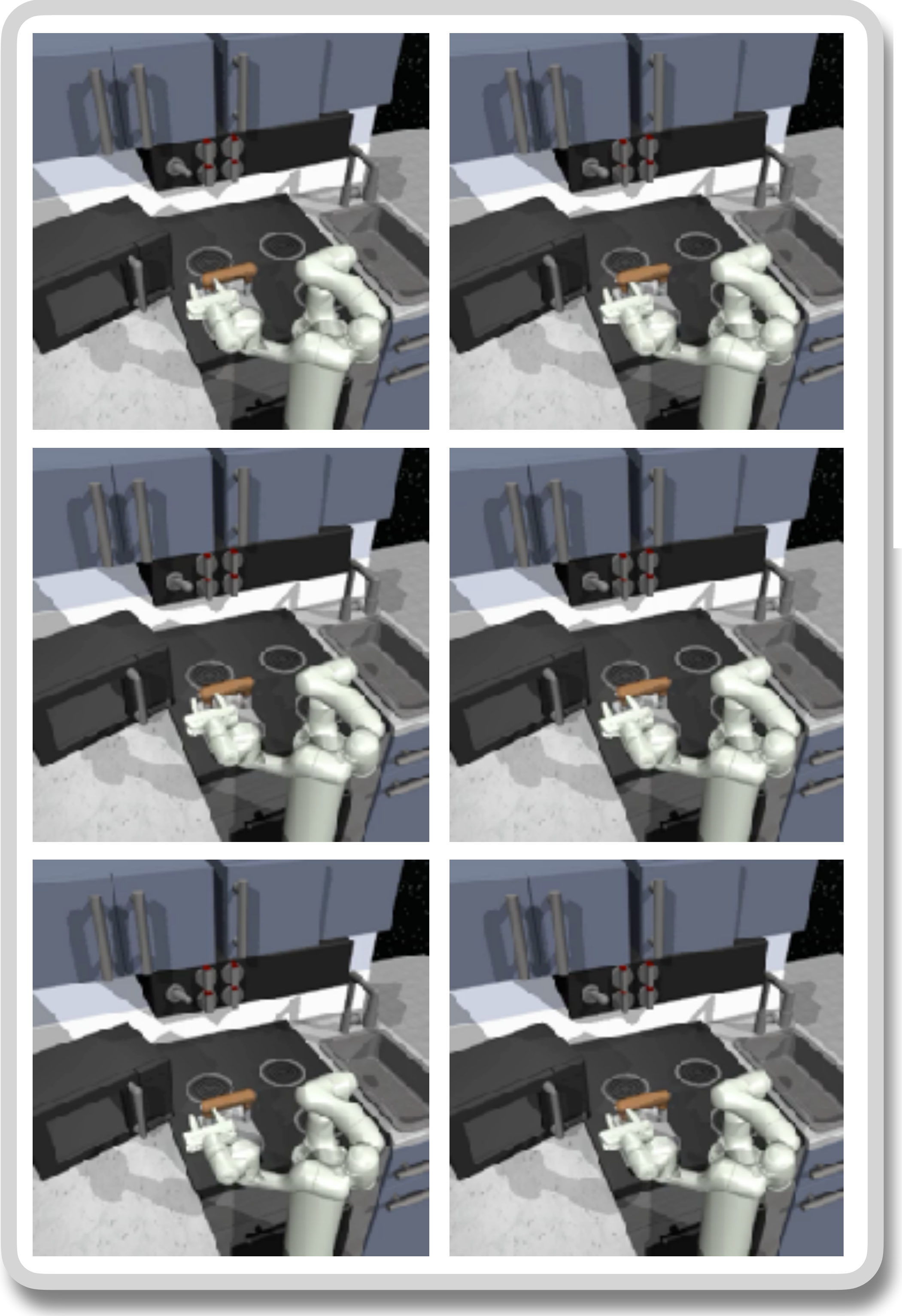
Schaal *"Dynamic Movement Primitives"* 2006

Lee et al. *"Composing Complex Skills by learning Transition Policies"* ICLR 2019

Lee et al. *"Adversarial Skill Chaining for Long-Horizon Robot Manipulation via Terminal State Regularization"* CoRL 2021

Skill Prior RL (SPiRL)

Large Offline Dataset



Pertsch, **Lee** et al. CoRL 2020, CoRL 2021

Skill-based Model-based RL (SkiMo)

Shi, Lim, **Lee**. *"Skill-based Model-based RL"* CoRL 2022

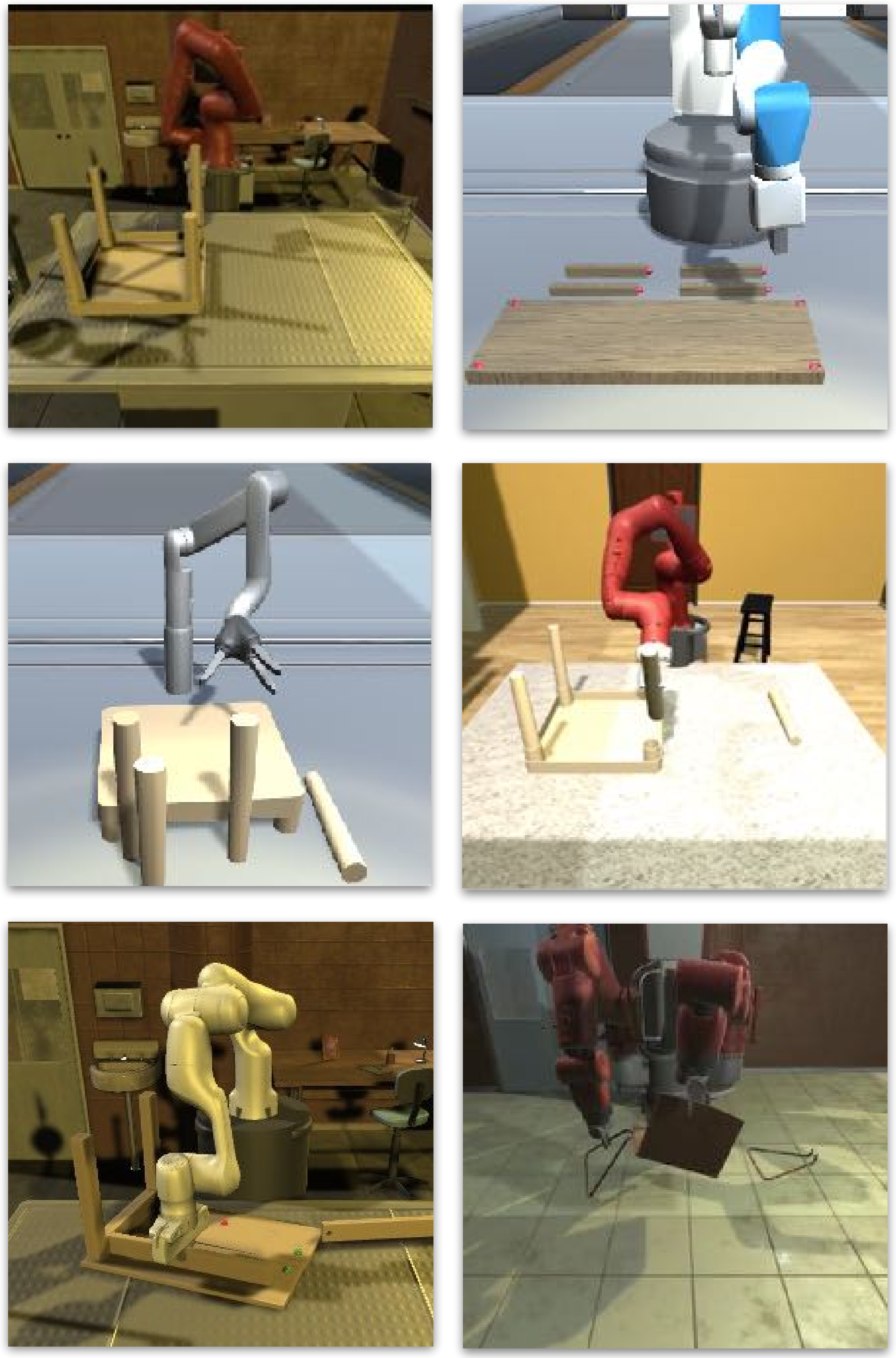
Results

Maze   Kitchen (MKBL)   Kitchen (MLSH)   CALVIN

Dreamer   SPiRL   SPiRL+Dreamer   SkiMo (Ours)

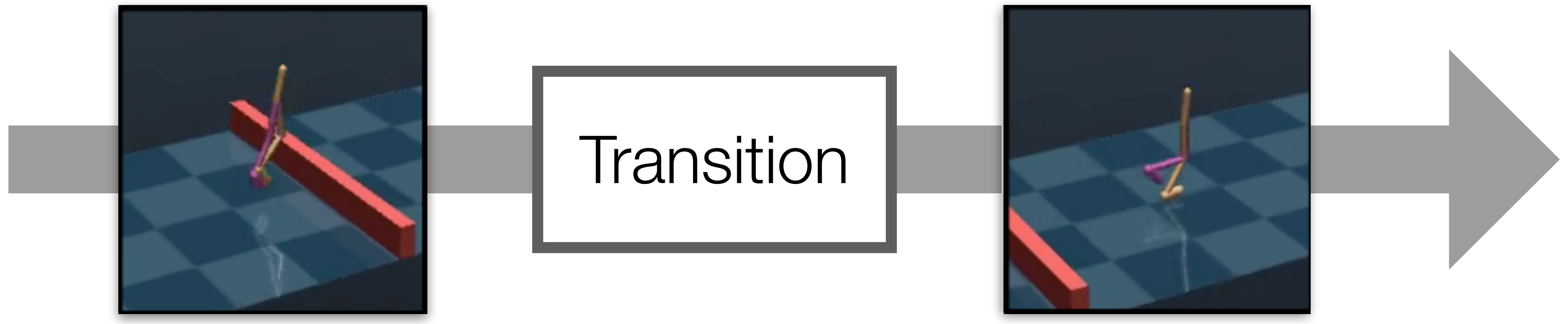
Shi, Lim, **Lee**. *"Skill-based Model-based Reinforcement Learning"* CoRL 2022

Scaling robot learning to long-horizon tasks

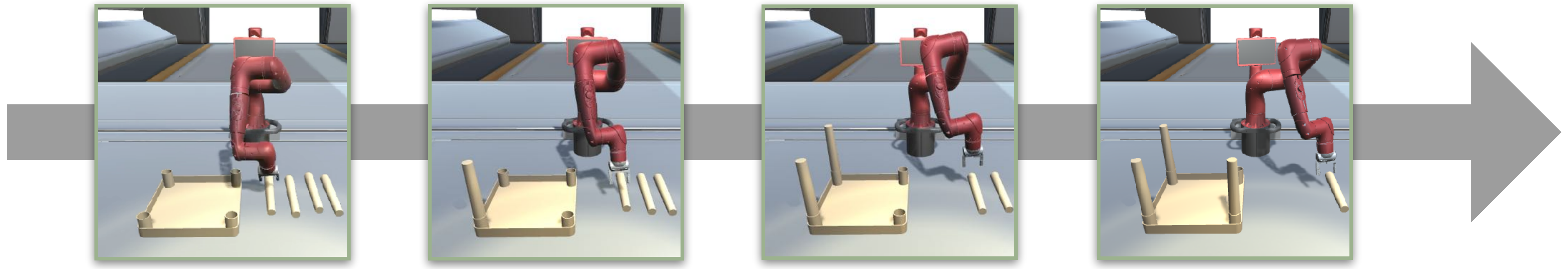


IKEA furniture assembly [ICRA'21]

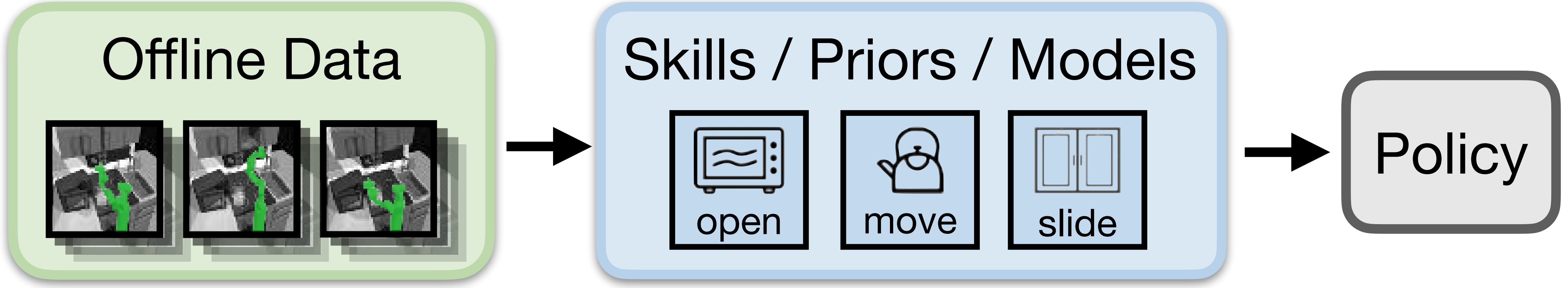
Benchmark



Transition policy [ICLR'19]



Skill chaining via fine-tuning [CoRL'21]



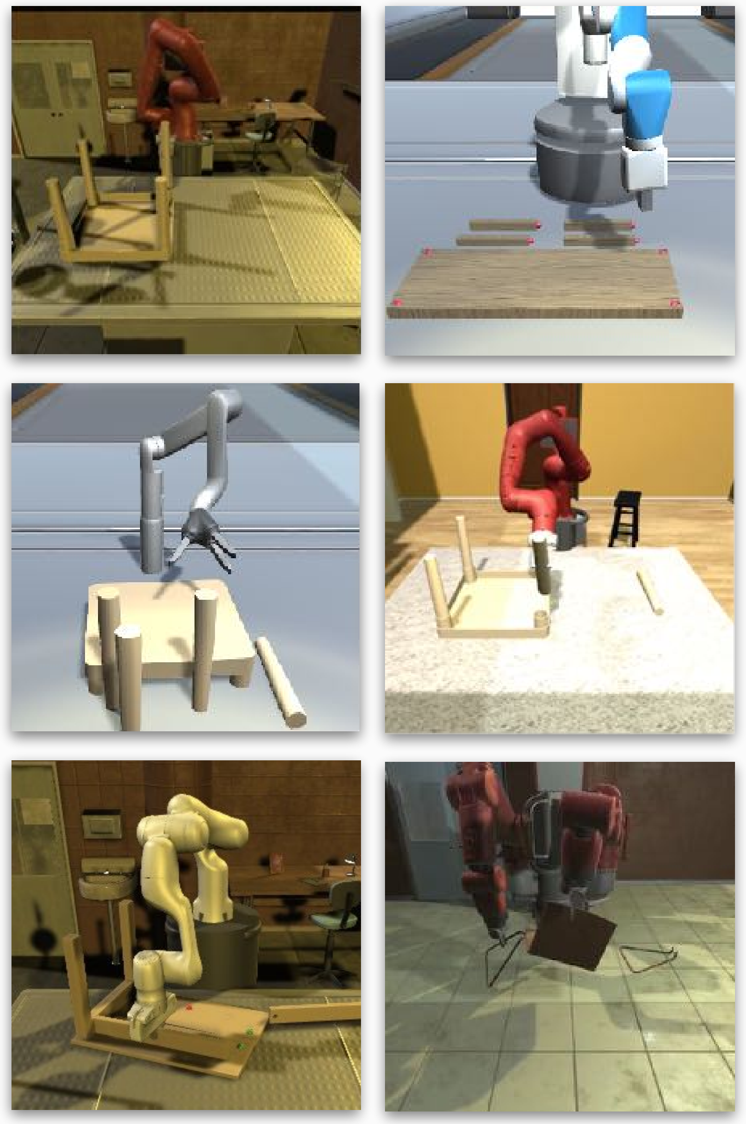
Skill-based RL [CoRL'20, CoRL'21, CoRL'22]

Algorithm

My research goal

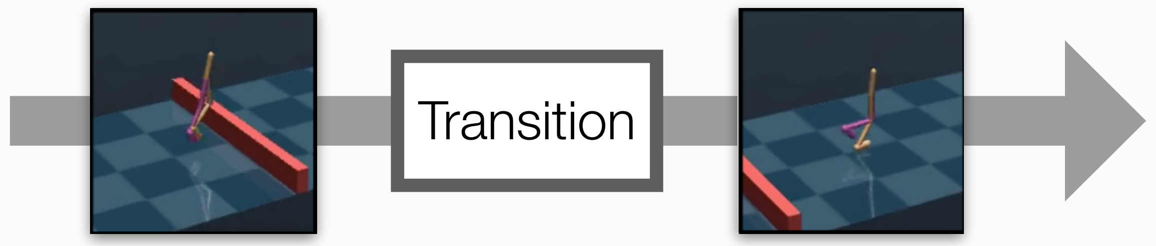
Scaling robot learning to real-world tasks

Complex, long-horizon

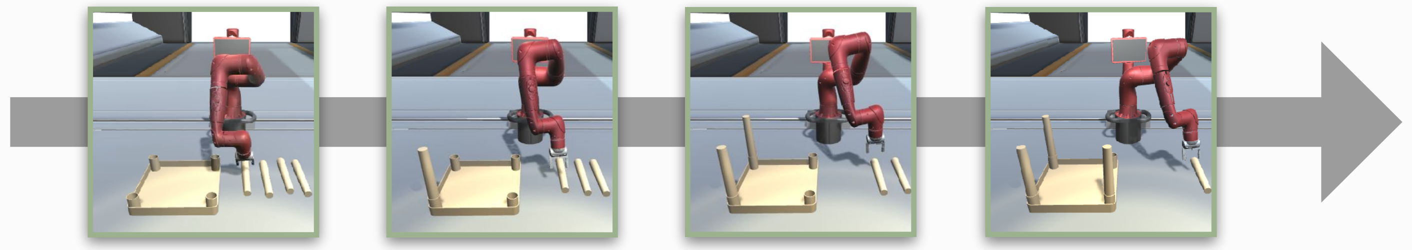


IKEA furniture assembly [ICRA’21]

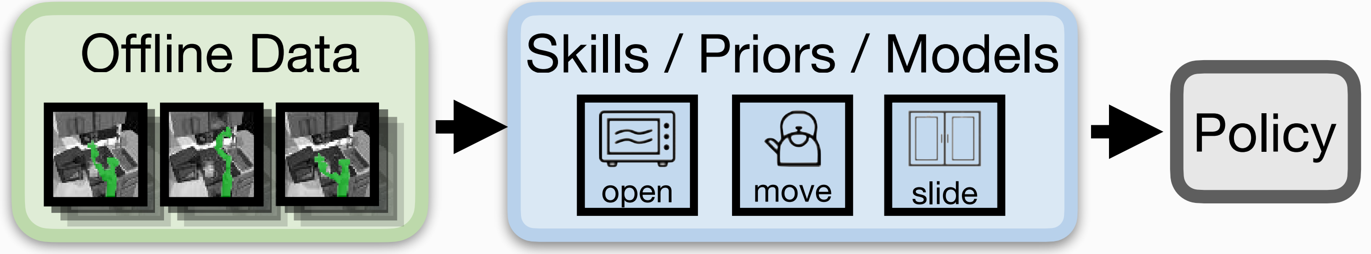
Benchmark



Transition policy [ICLR’19, ICLR’20]



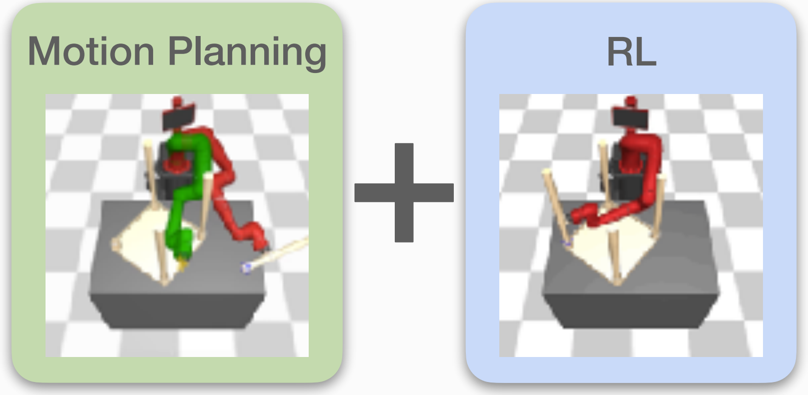
Skill chaining via fine-tuning [CoRL’21]



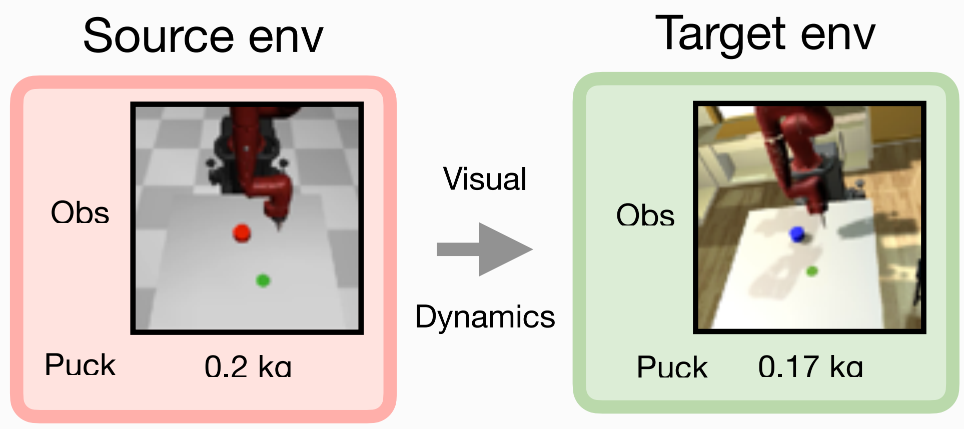
Skill prior [CoRL’20, CoRL’21, CoRL’22]

Algorithm

In the real world

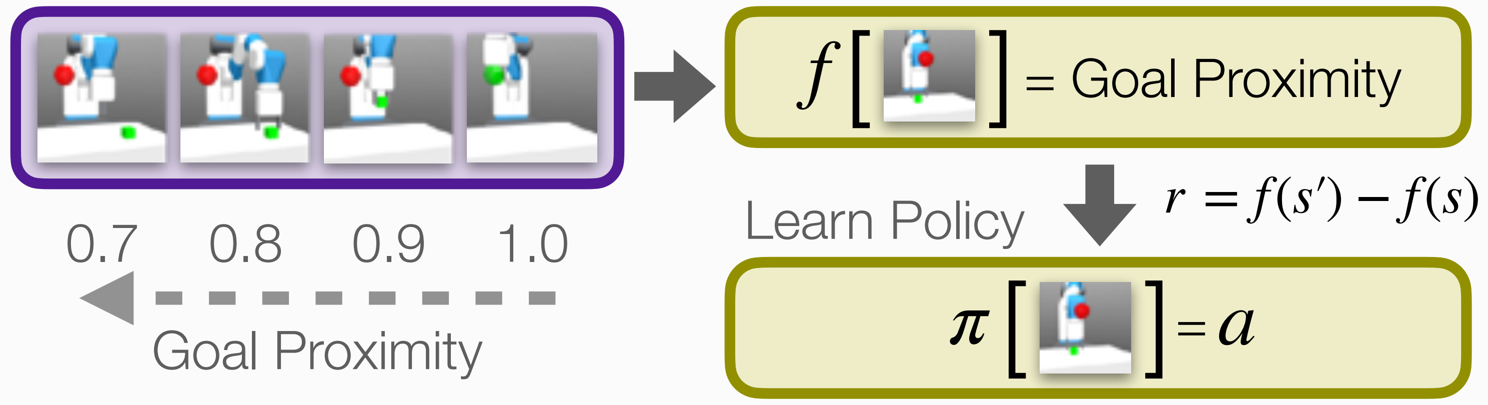


Combining motion planner [CoRL’20, CoRL’21]



Policy transfer [RSS’21]

Transfer learning



Learning from Observation [CoRL’19, NeurIPS’21]

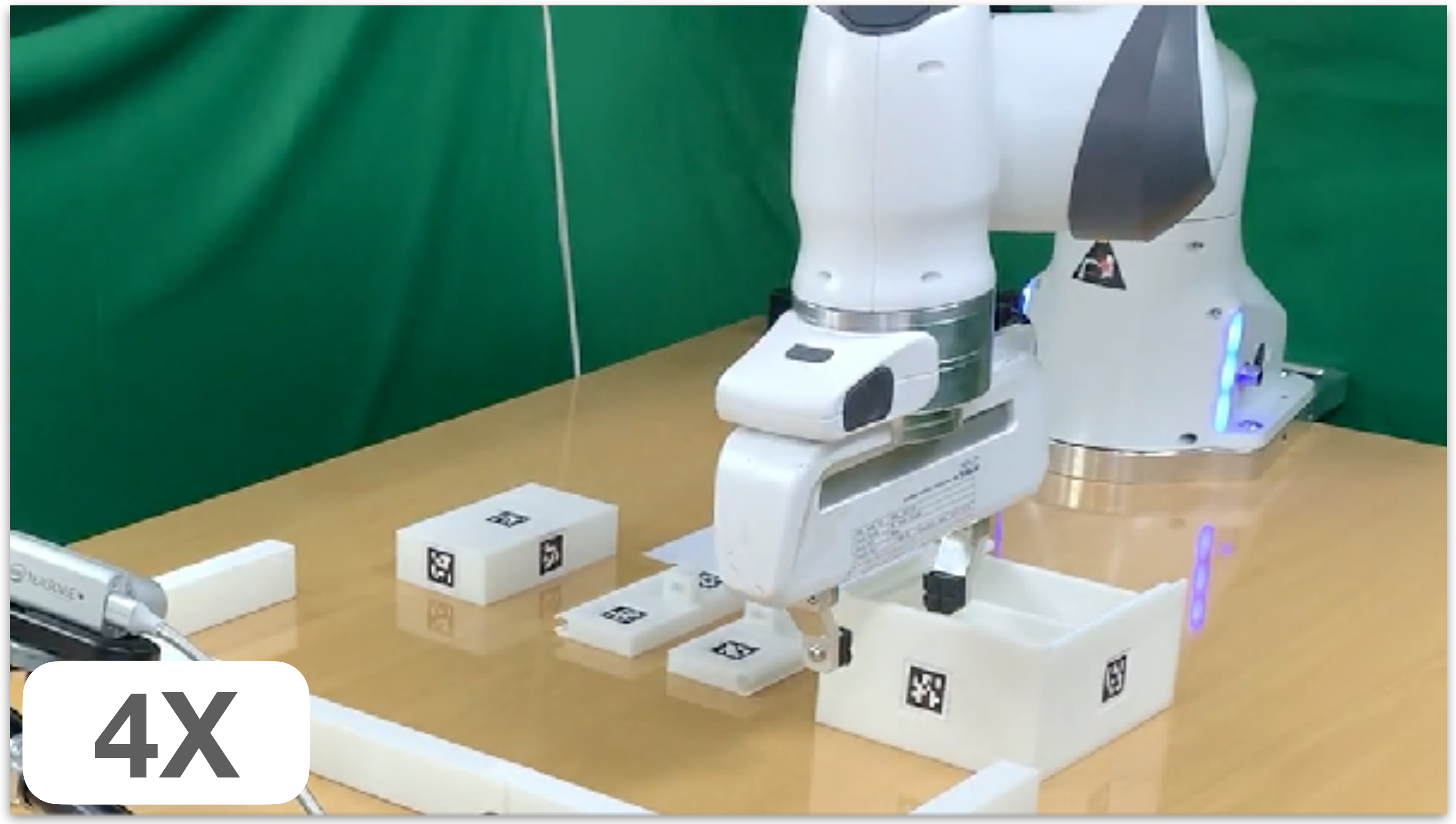
Imitation learning

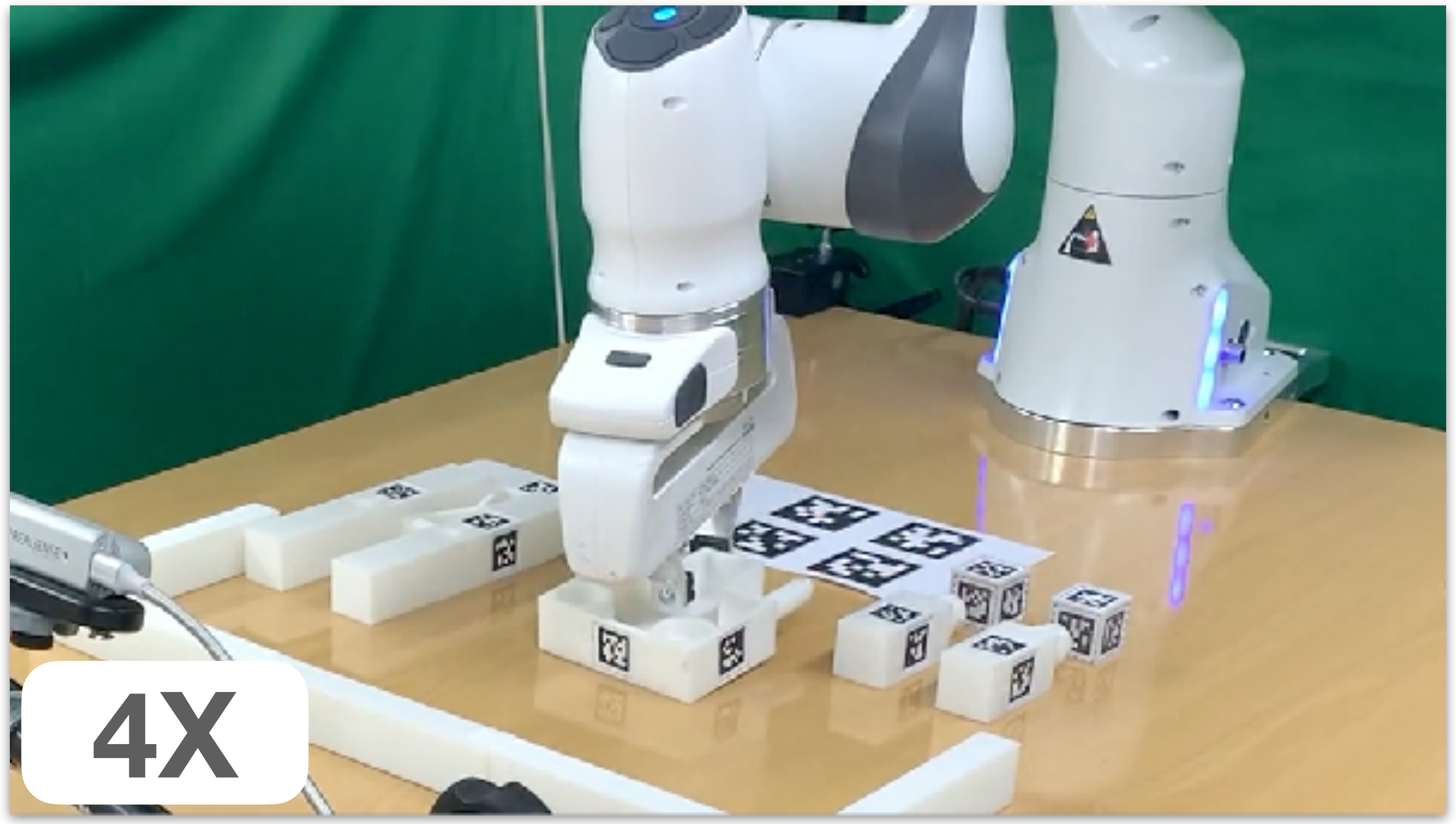
My research goal:

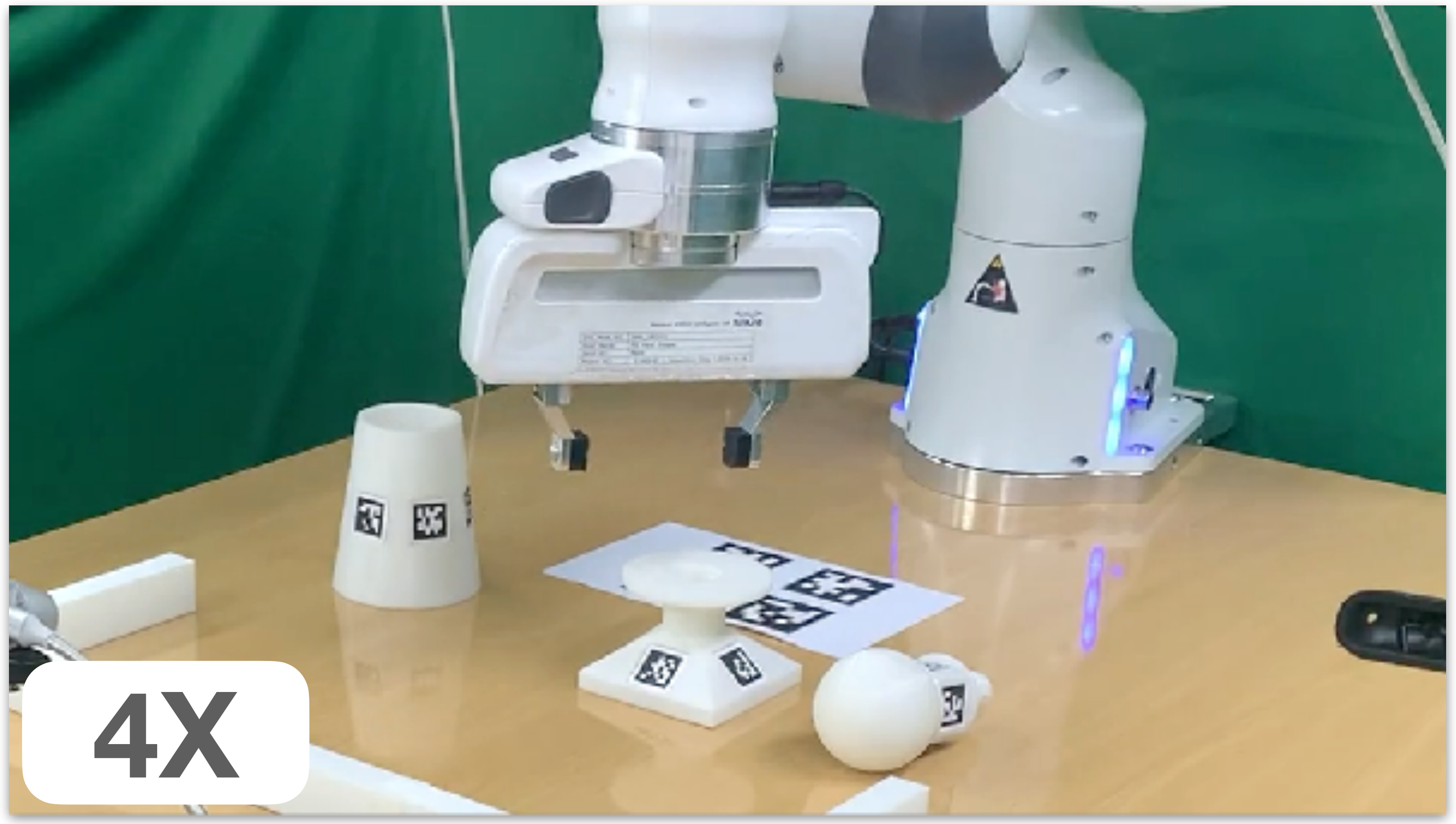
Scaling robot learning to real-world tasks

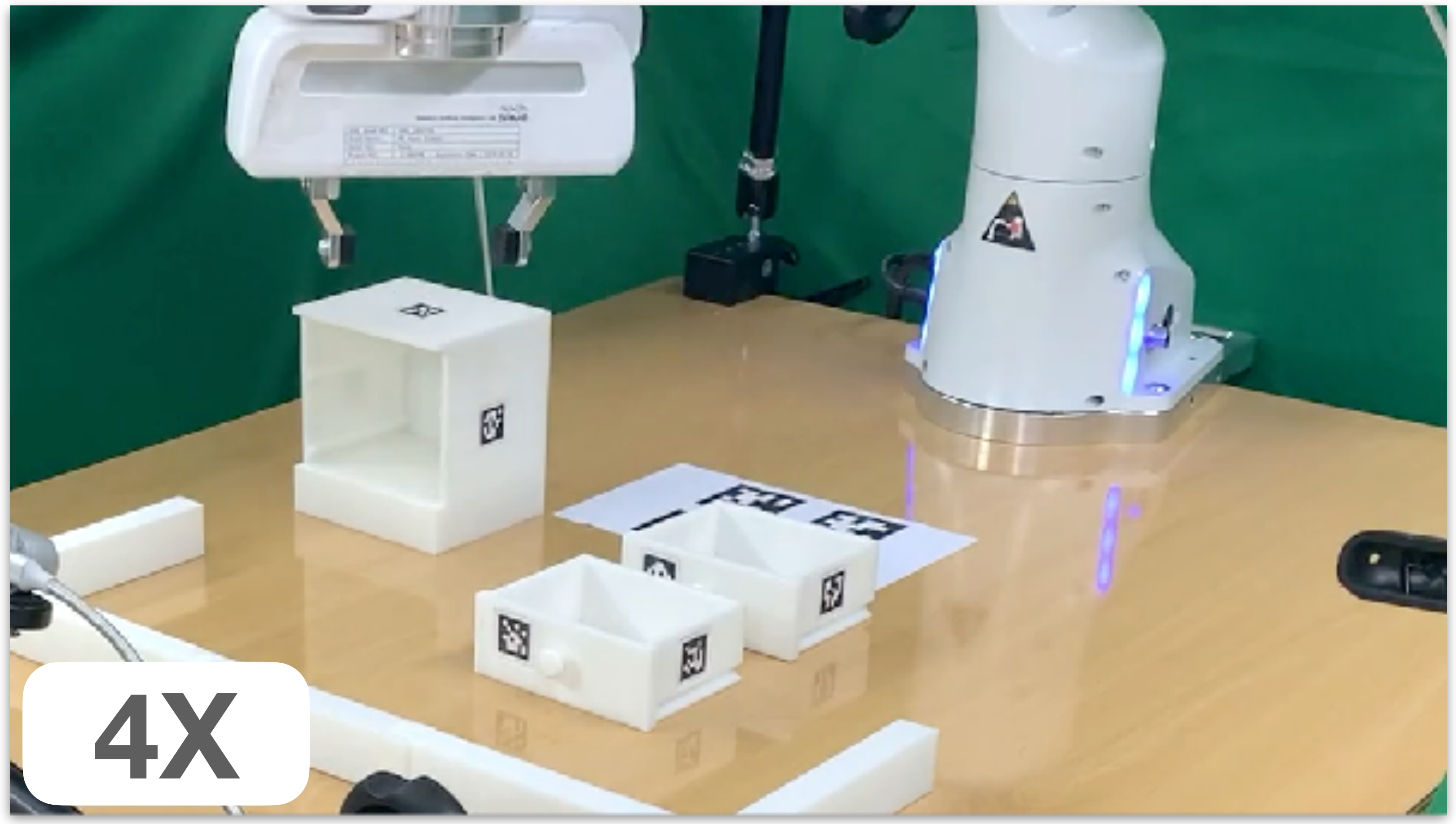

REALSENSE
2X

Learning furniture assembly in the real world







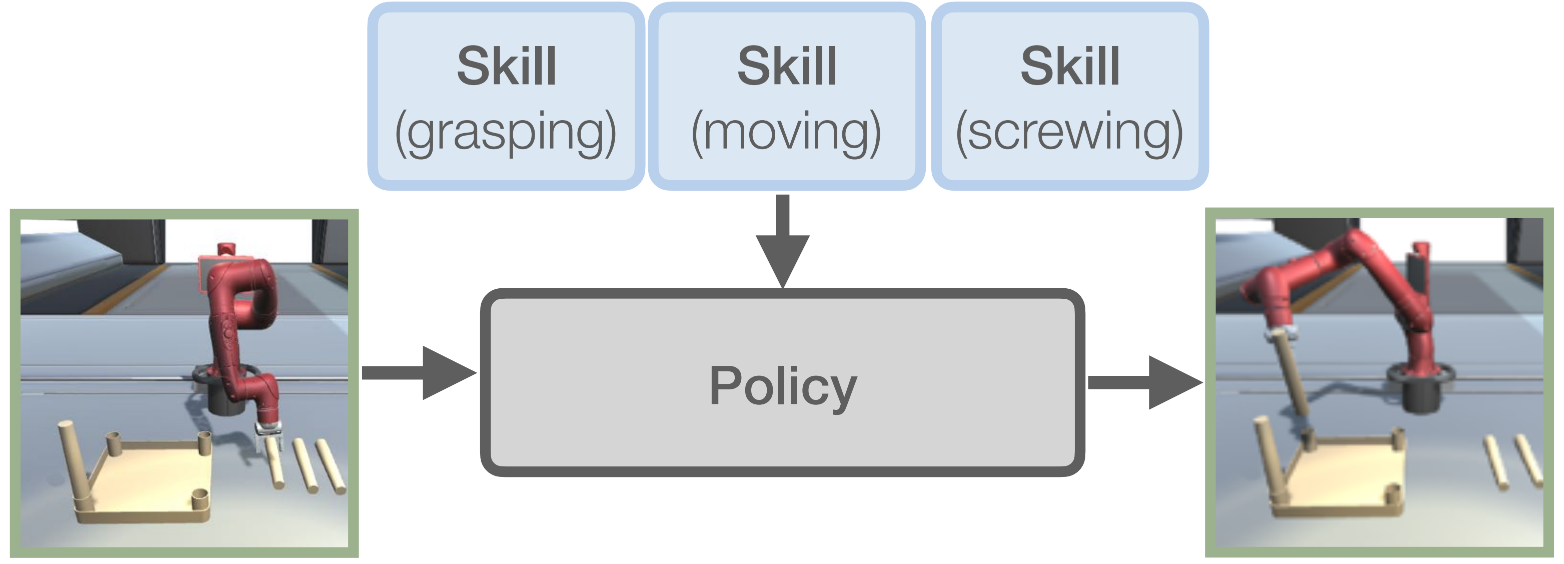


Robot learning in the real world

Real-world benchmark

Diverse data

Efficient learning method



Robot learning in the real world

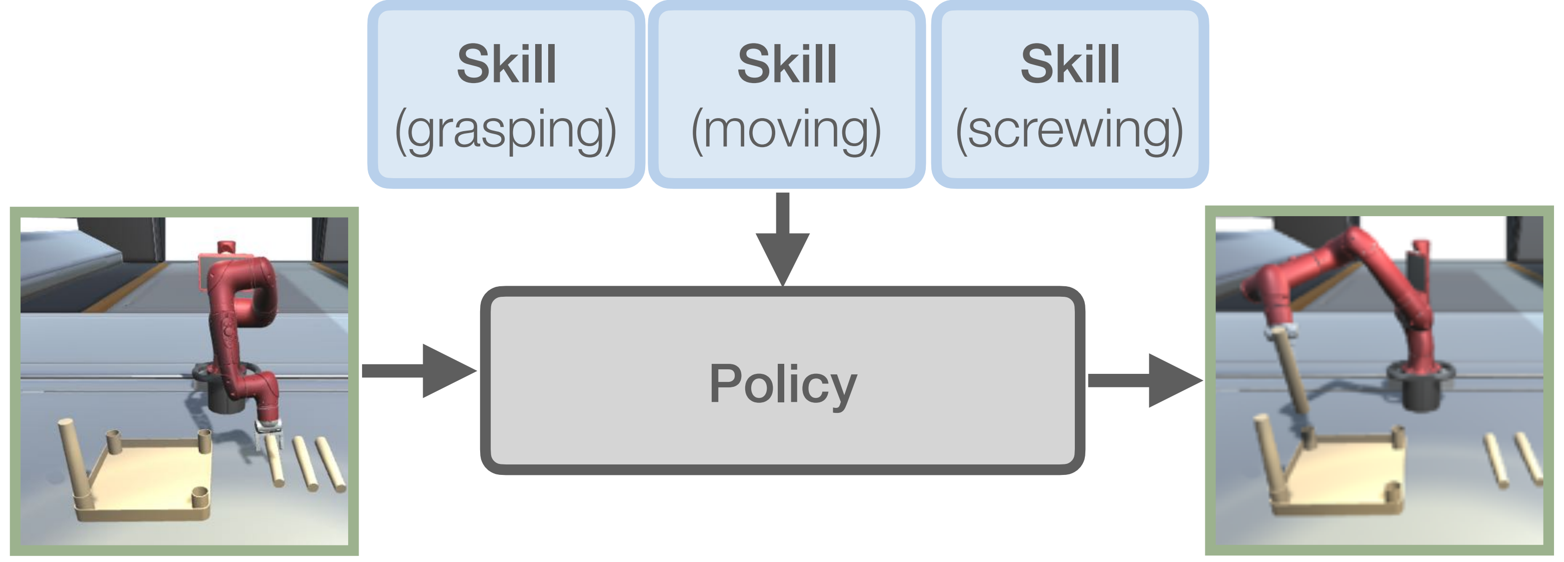
Real-world benchmark

+

Diverse data

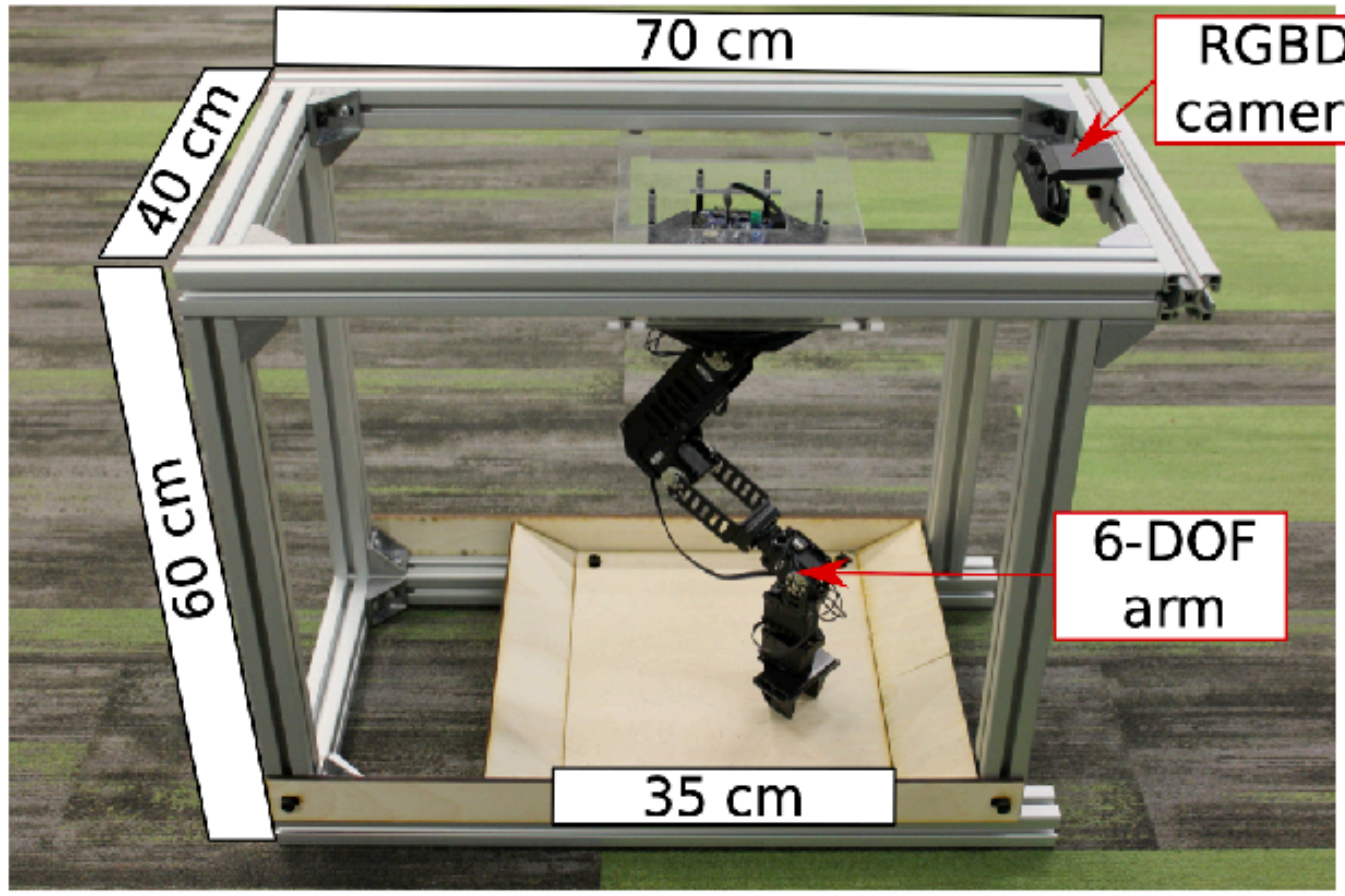
+

Efficient learning method

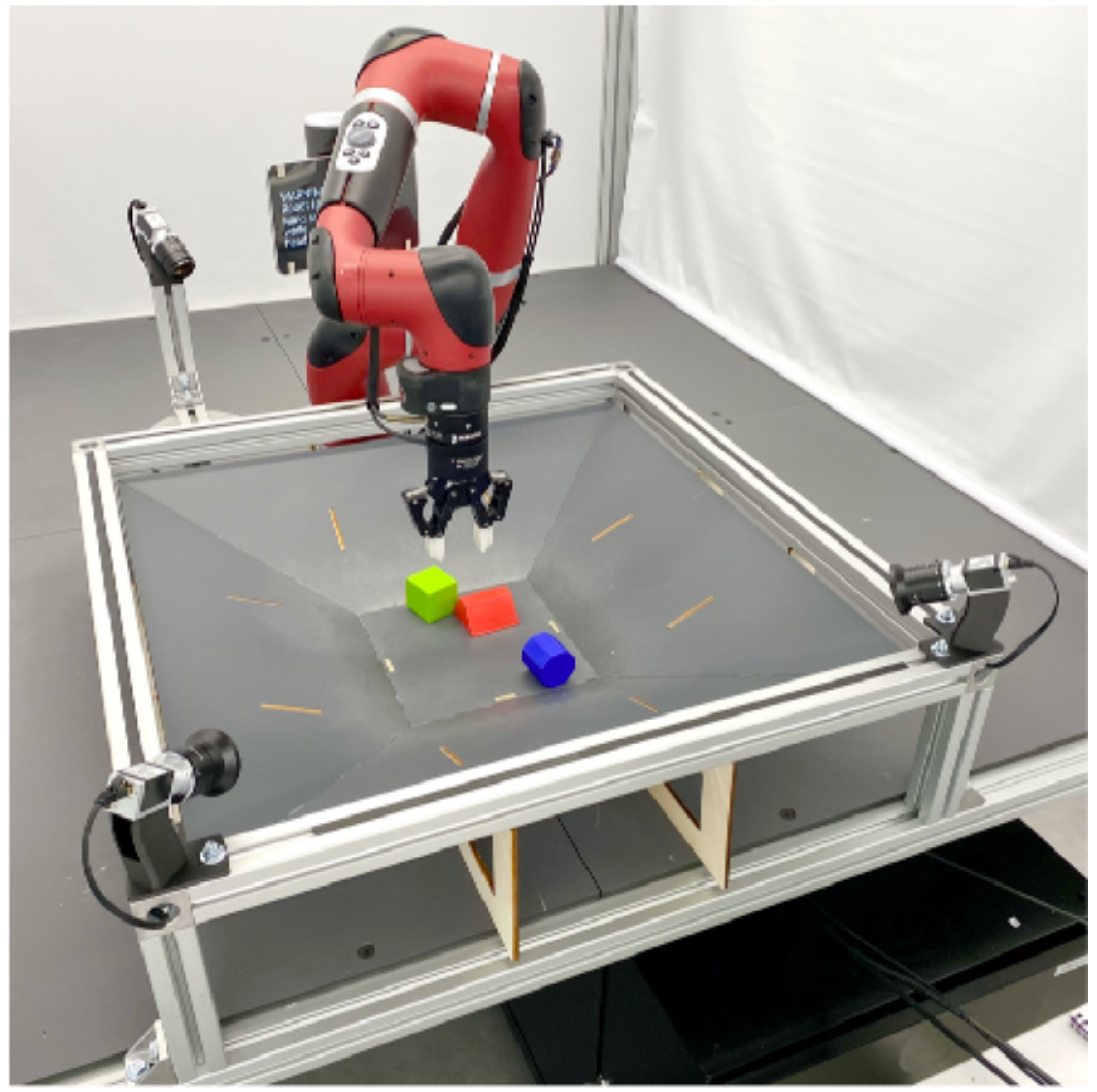


Prior real-world benchmarks

YCB



REPLAB



RGB-Stack

Callie et al. *"Yale-CMU-Berkeley Dataset for Robotic Manipulation Research"* IJRR 2017
Yang et al. *"REPLAB: A Reproducible Low-Cost Arm Benchmark Platform for Robotic Learning"* ICRA 2019
Lee et al. *"Beyond Pick-and-Place: Tackling Robotic Stacking of Diverse Shapes"* CoRL 2021

Furniture assembly in the real world

Reproducible "real-world" benchmark

No sim-to-real gap

Long-horizon tasks

Consists of 3 - 6 parts to be assembled

Dexterous skills

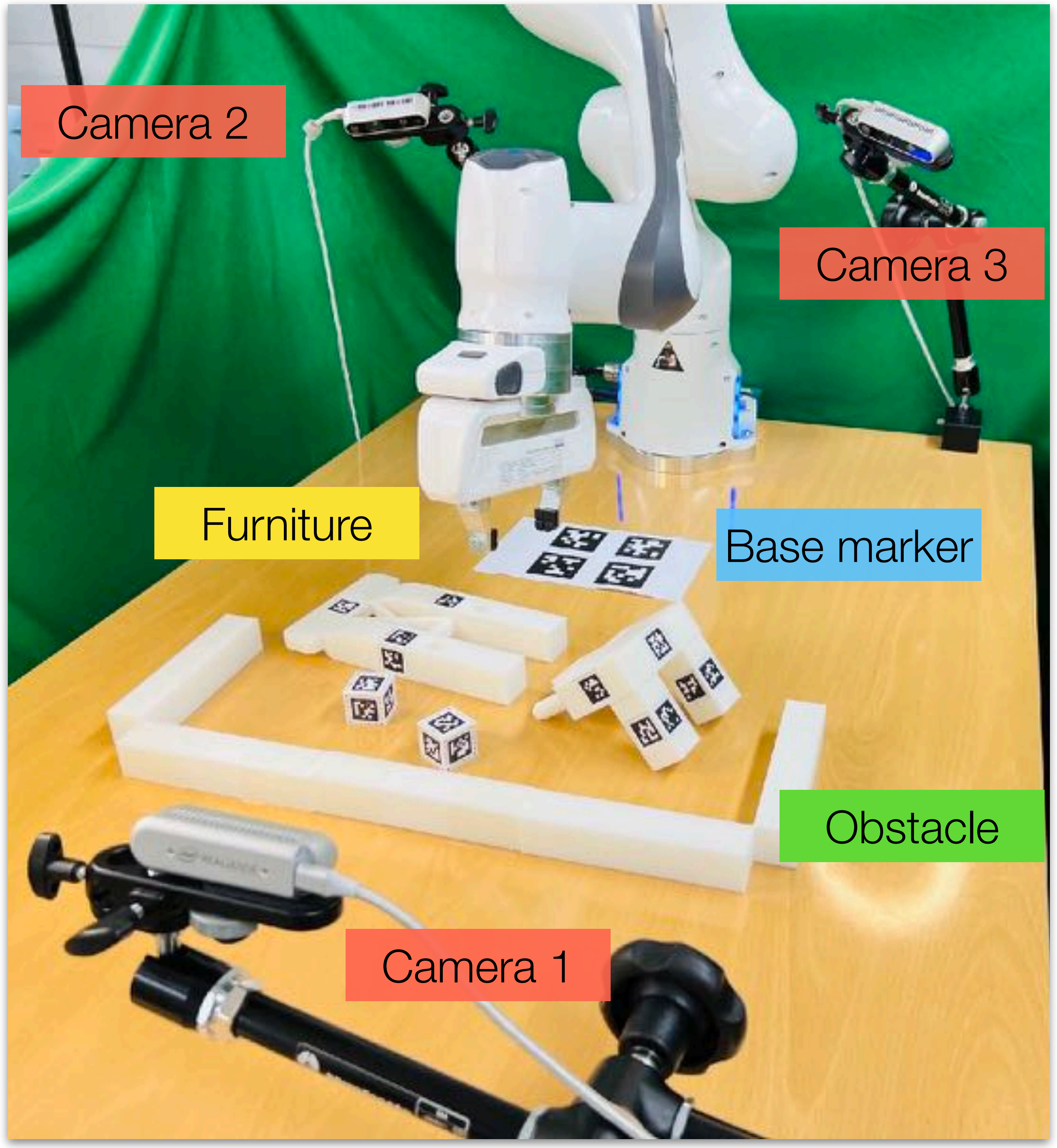
Screwing, inserting, flipping — interactions between objects

A variety of task instances

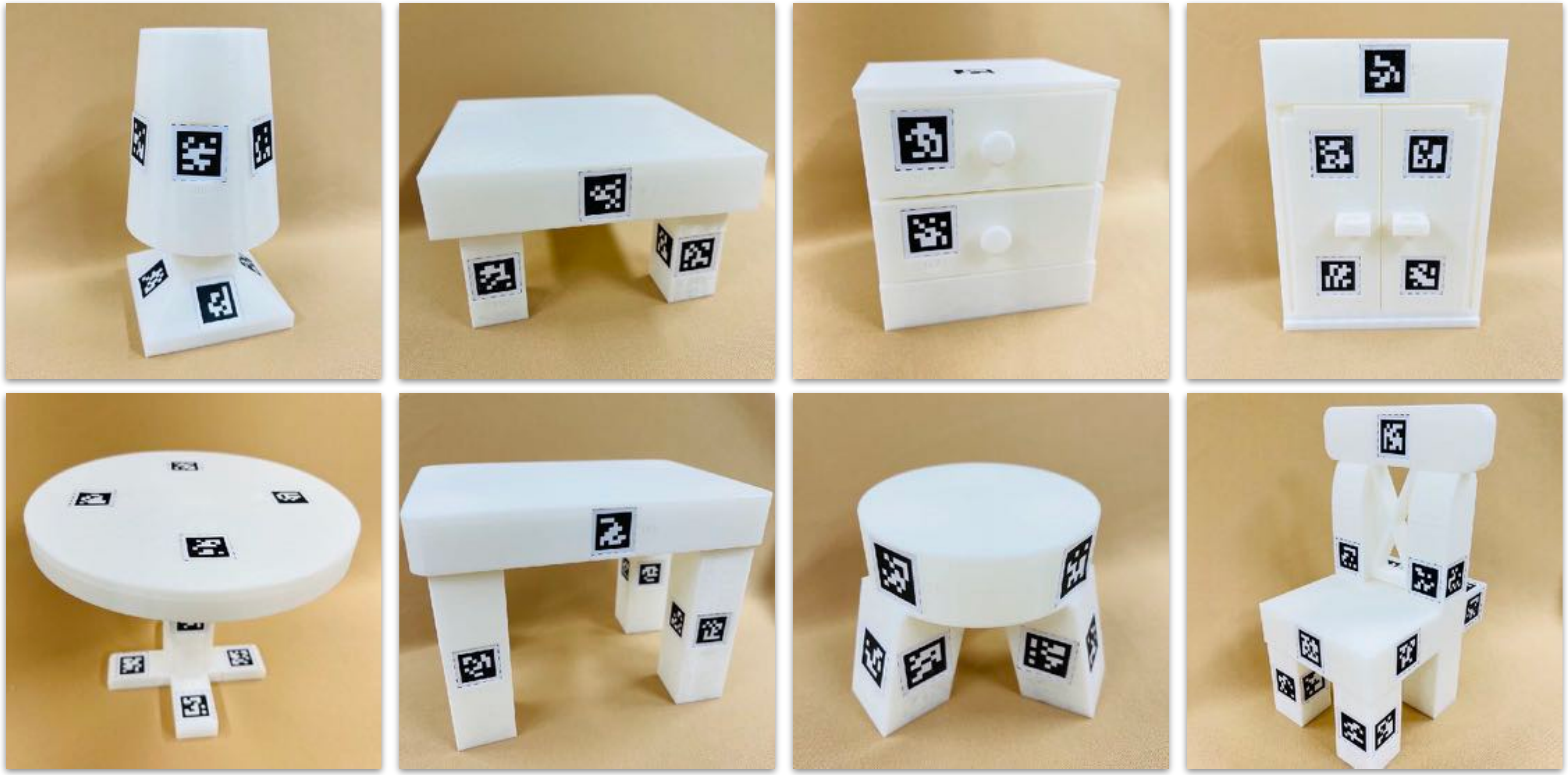
8 tasks with different difficulties and challenges

Lee et al. *"Real-World Furniture Assembly Benchmark"* In progress

Furniture assembly in the real world



Environment

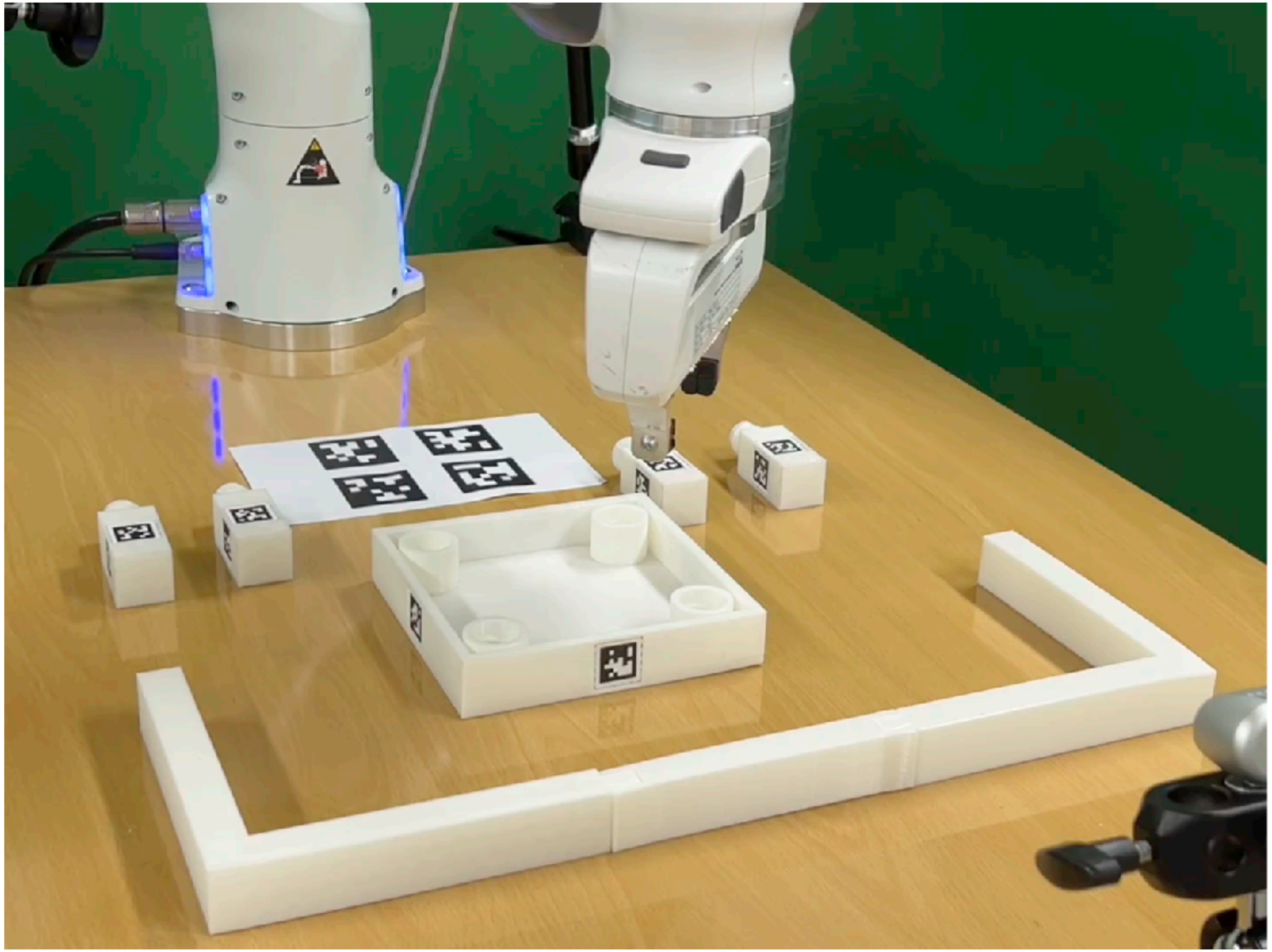


Furniture models

Furniture assembly in the real world

Lee et al. *"Real-World Furniture Assembly Benchmark"* In progress

Furniture assembly in the real world



Offline RL on 900+ trajectories

Success rates: 10-20%

Robot learning in the real world

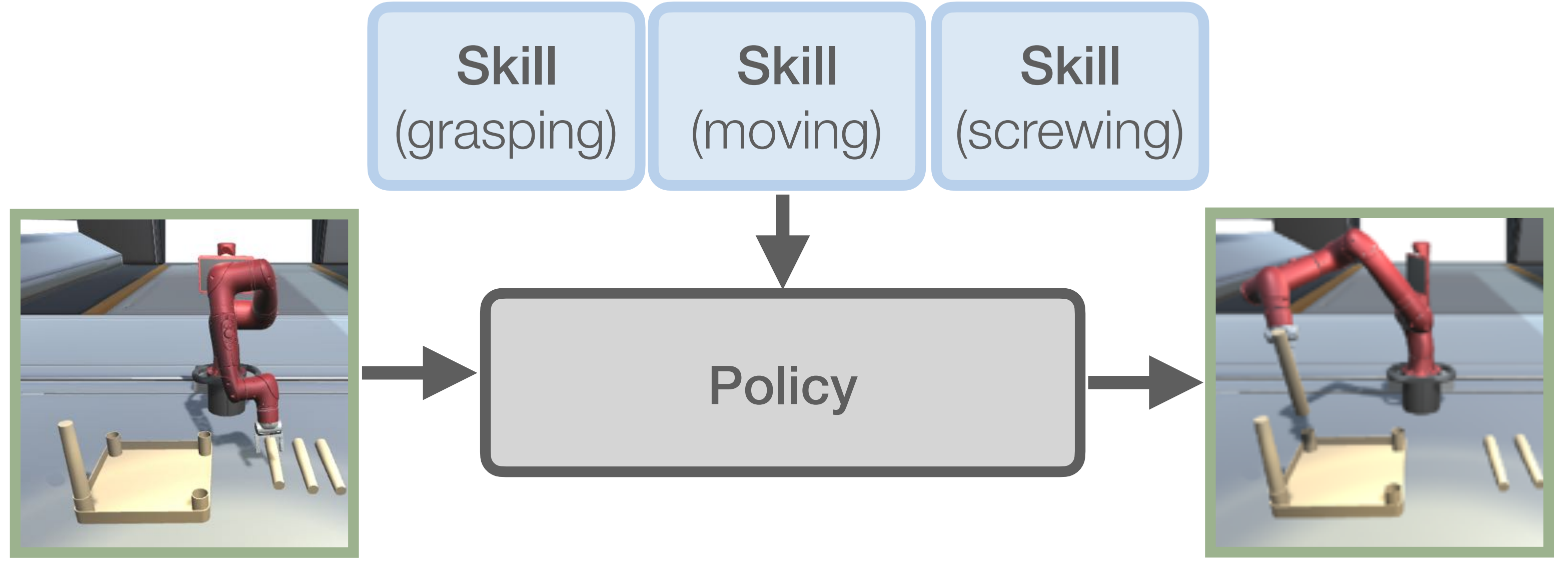
Real-world benchmark

\+

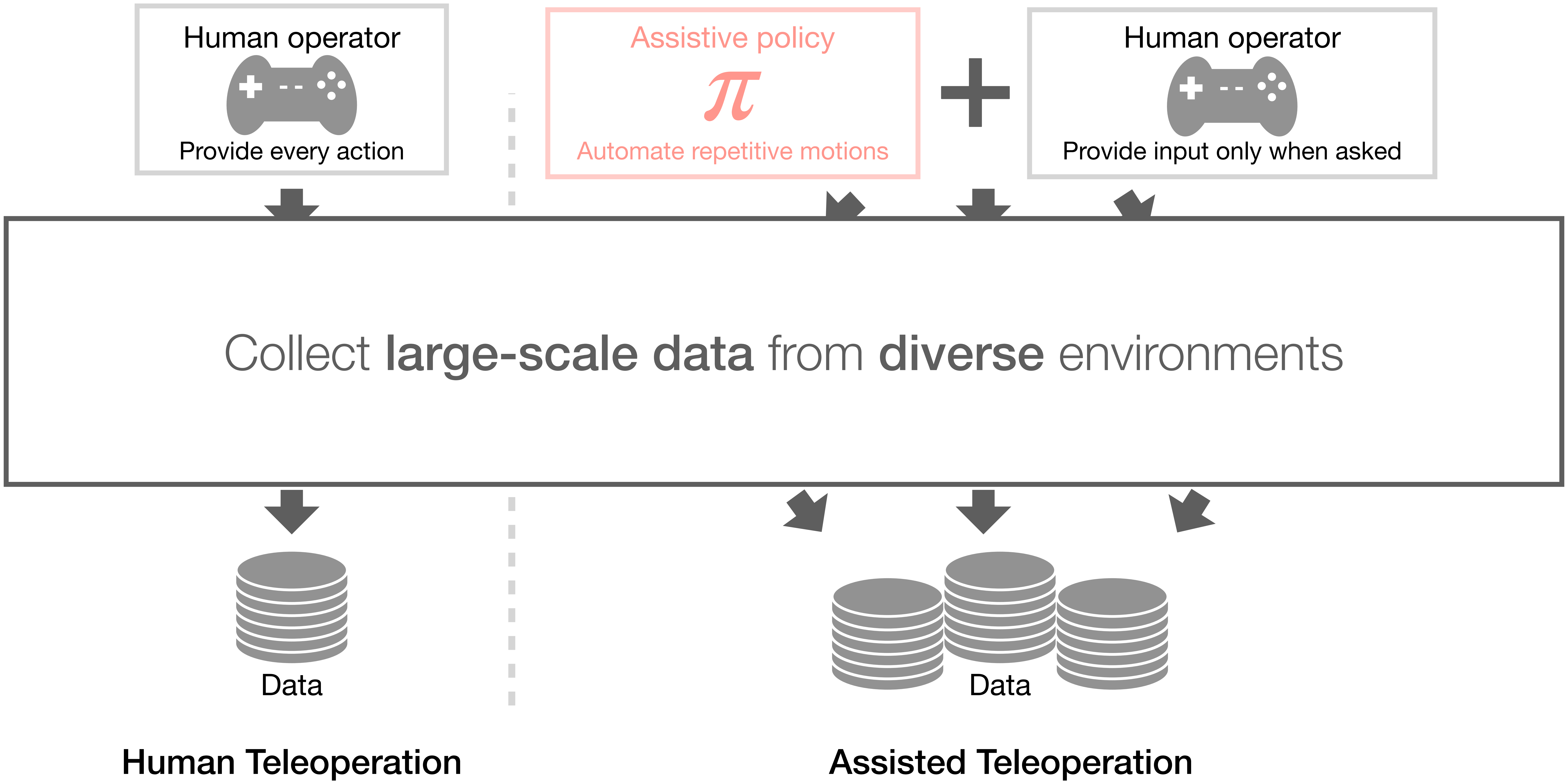
Diverse data

\+

Efficient learning method



Assisted teleoperation for easy data collection



Dass*, Pertsch*, **Lee** et al. "Assisted Teleoperation for Scalable Robot Data Collection" In submission

Large-scale robot dataset

“Real”-world data

Diverse environments

Diverse tasks

Large-scale data

Robot learning in the real world

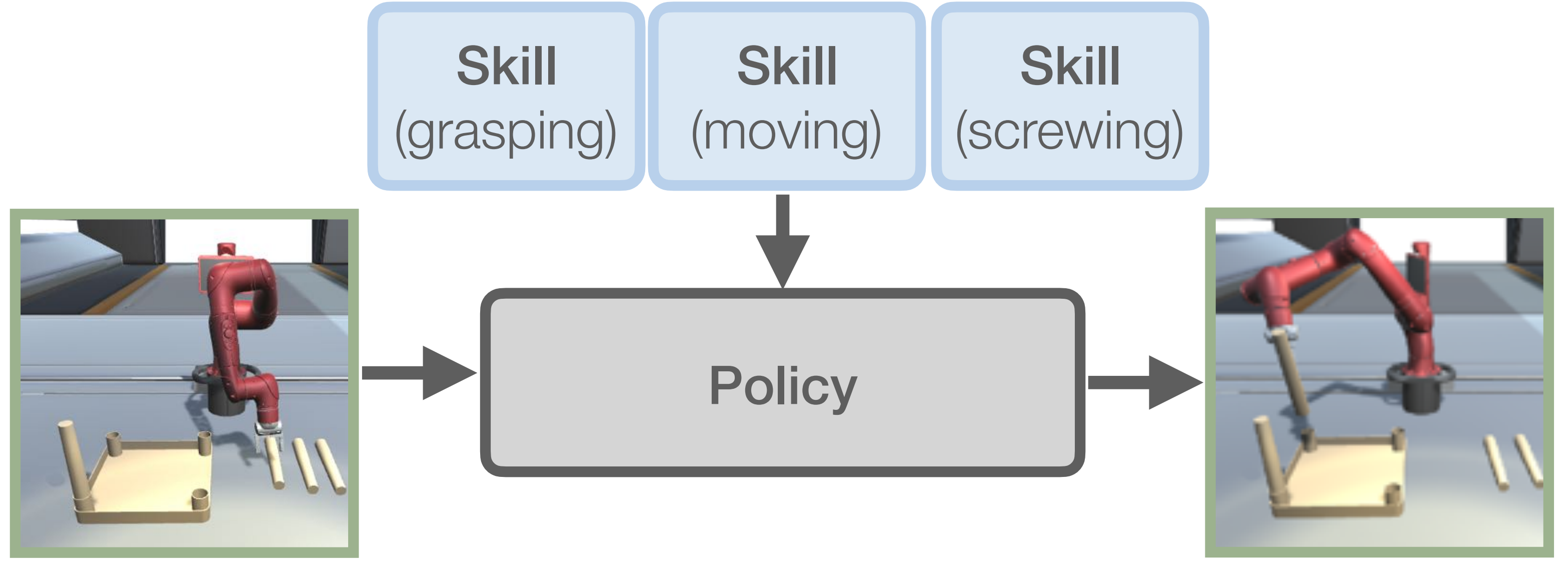
Real-world benchmark

\+

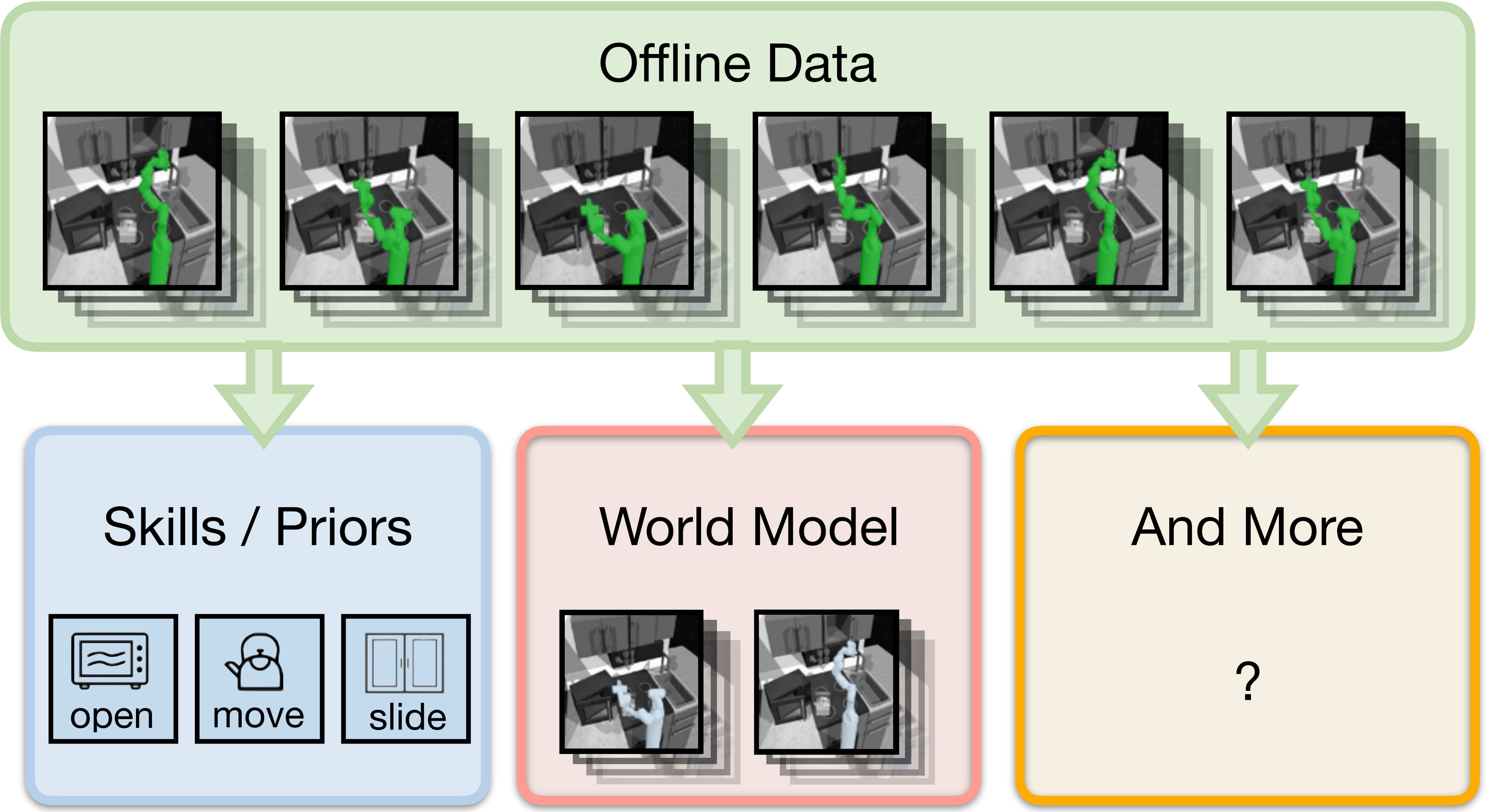
Diverse data

\+

Efficient learning method



Leverage offline data



Future directions

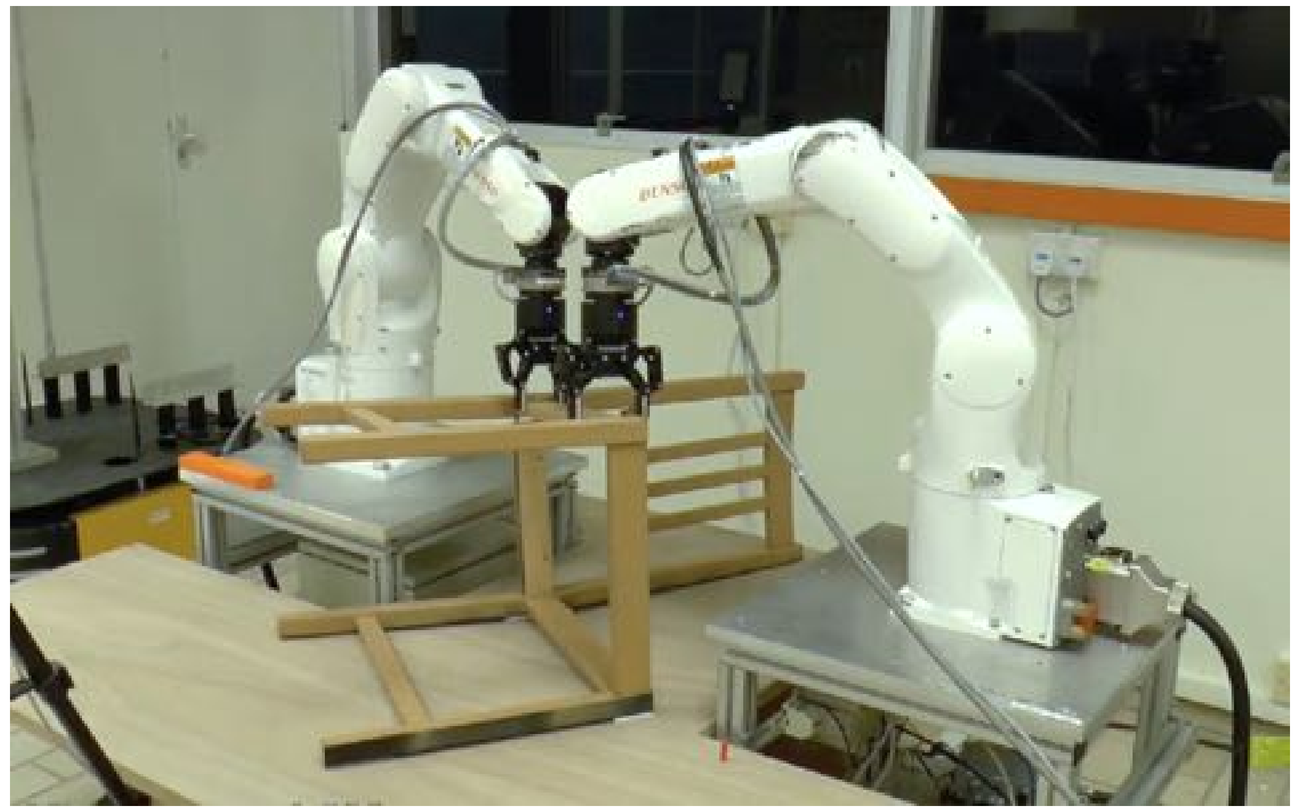
Scaling robot learning to real-world tasks

Complex long-horizon tasks in the real world

High-complexity (high-DoF) robot systems

Diversity of tasks

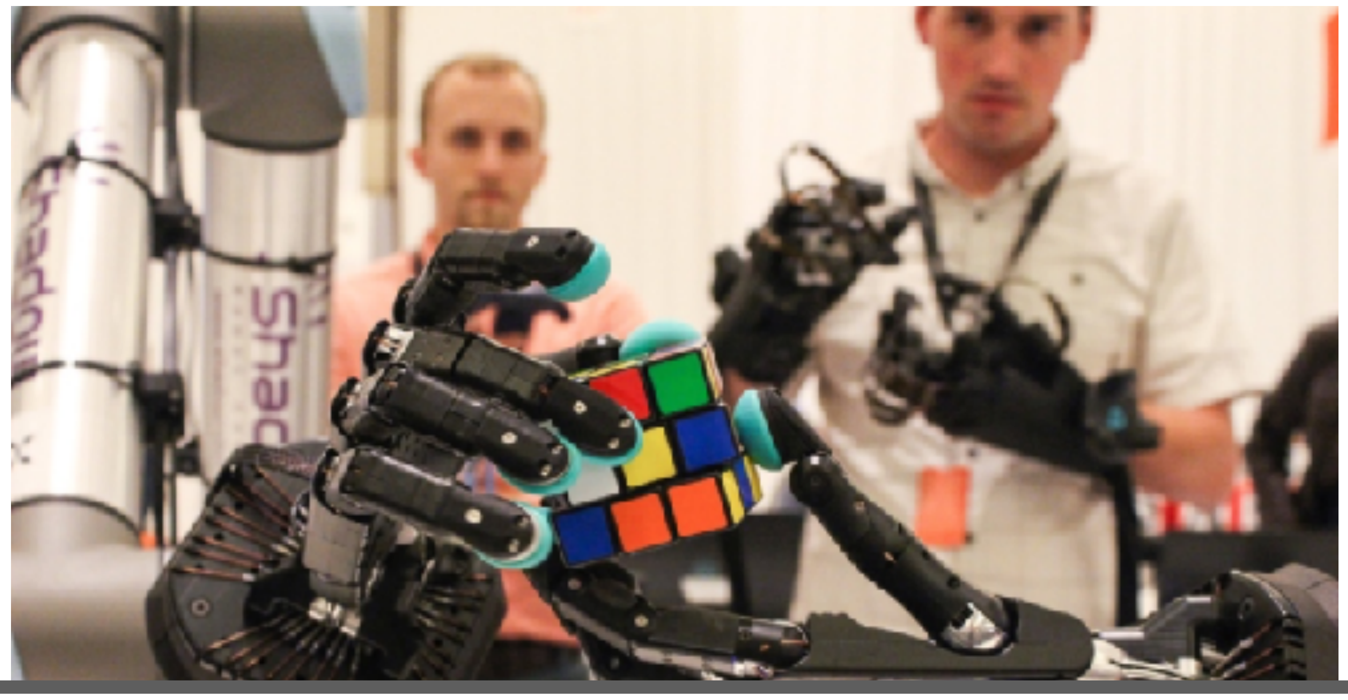
Future direction 1: Complex tasks



Prior knowledge (e.g. skills, world model, priors) from data
Unsupervised data collection
Multi-modal sensory representations

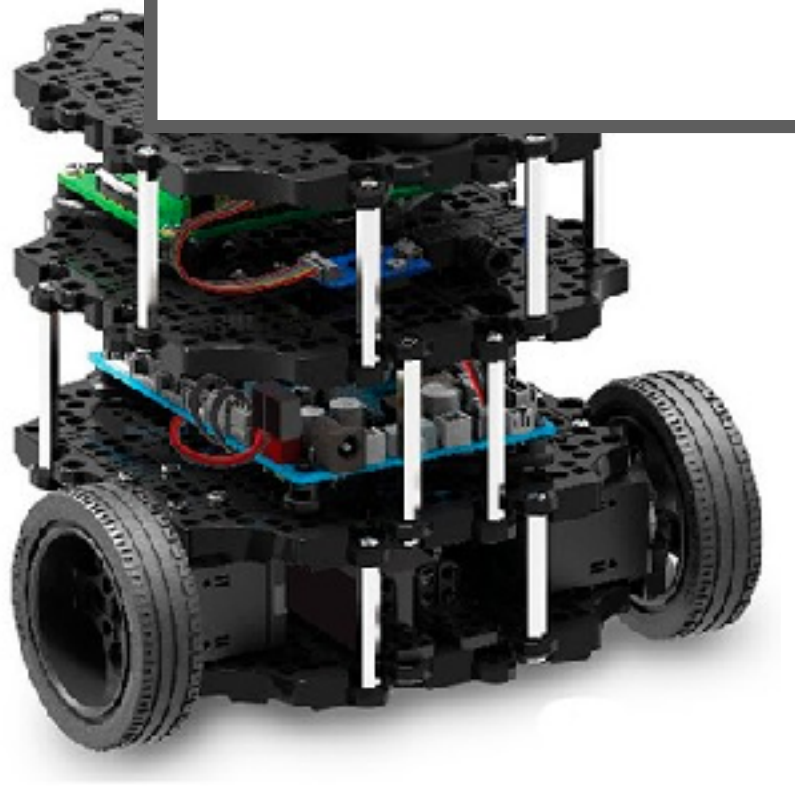
Future direction 2: Complex robot systems



M

Modular framework & Composable skills

Dexterous & mobile manipulation



Navigation

Mobile manipulation

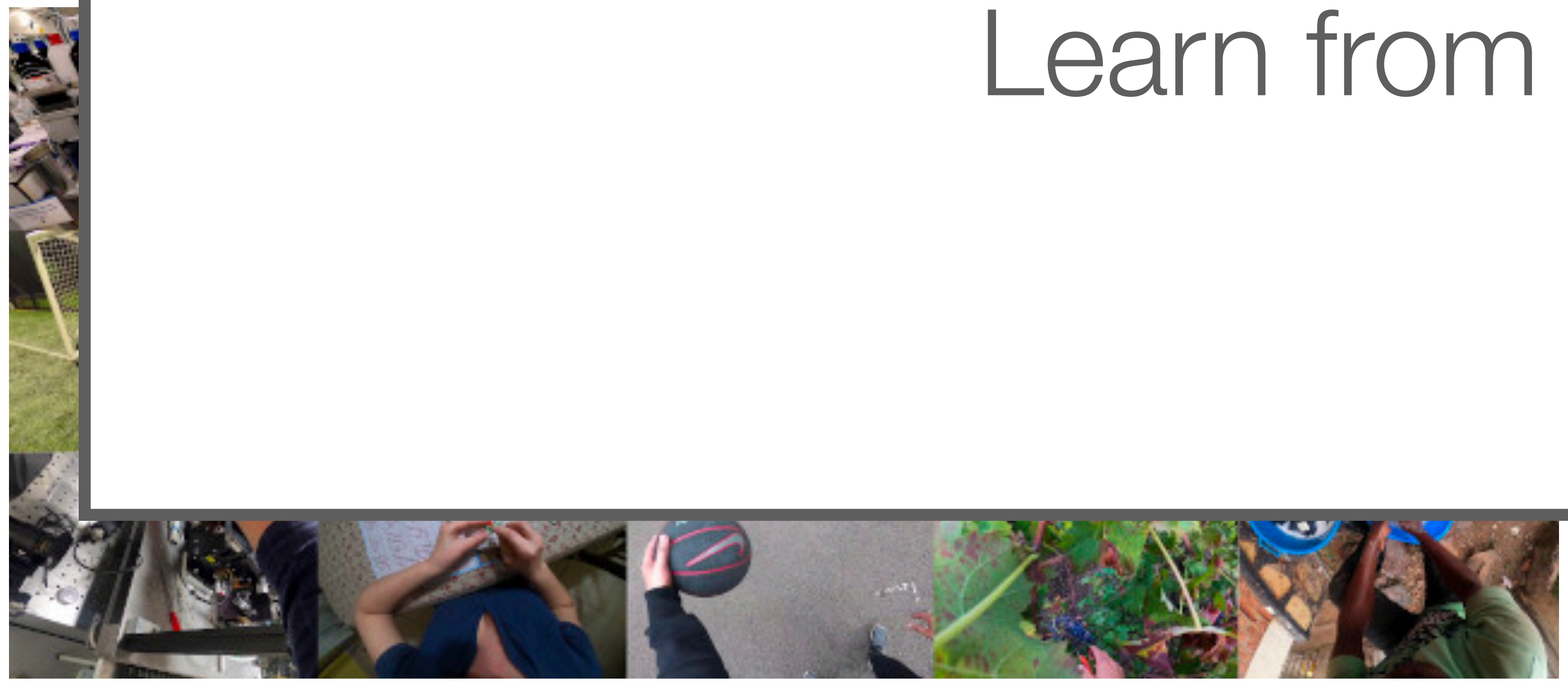
Future direction 3: Diverse tasks

General-purpose robots

Learn from human video and text data

&

Meta-learning

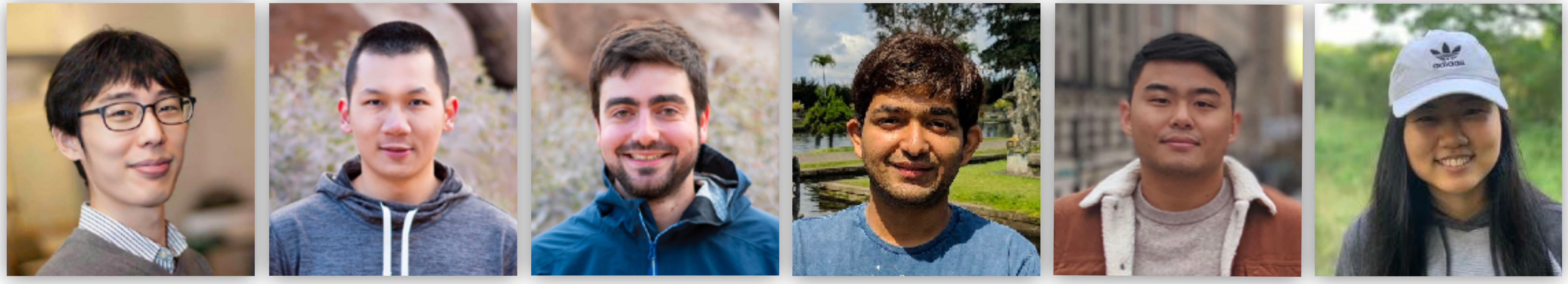


Human videos (Ego4D, Youtube)

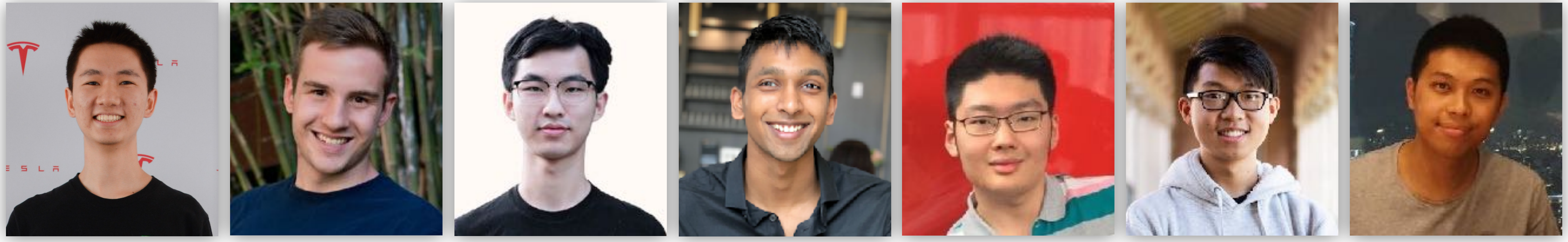
Text data and instructions

Acknowledgements

CLVR lab @ USC & KAIST



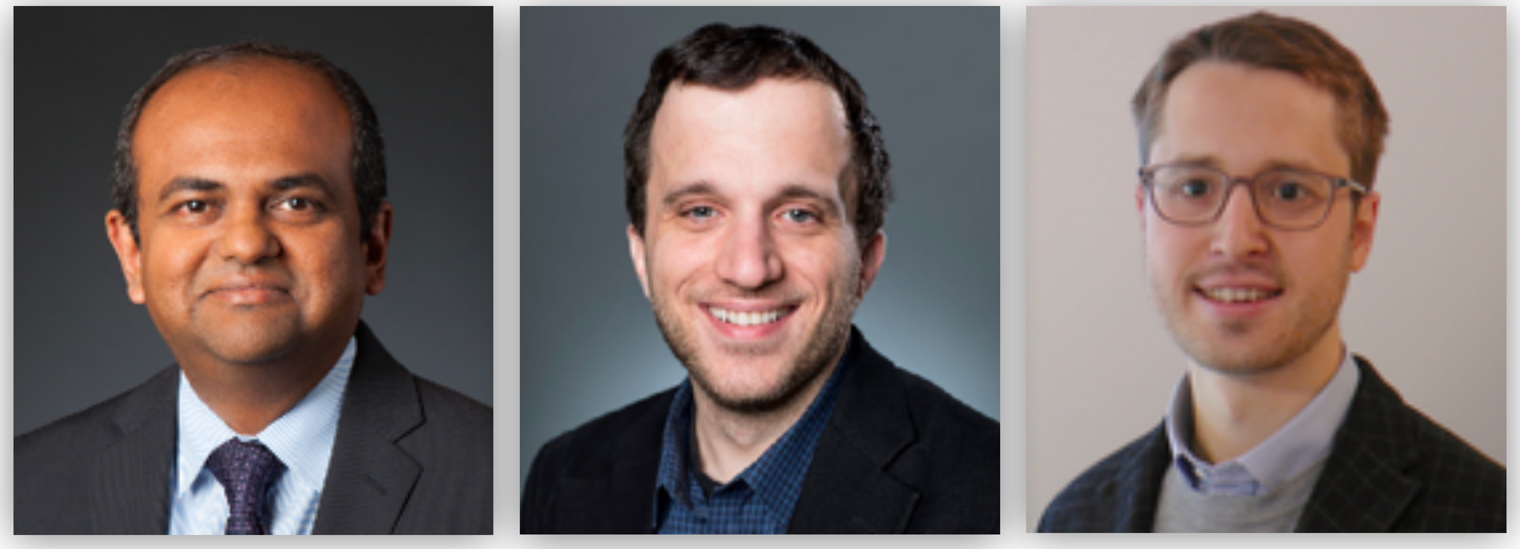
Joseph Lim | Shao-Hua Sun | Karl Pertsch | Ayush Jain | Jesse Zhang | Grace Zhang



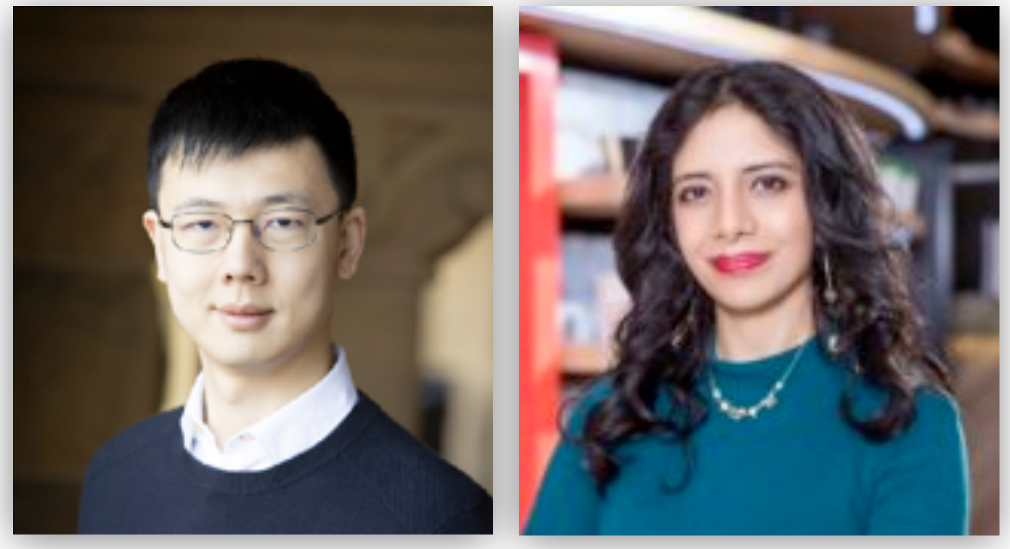
Edward Hu | Andrew Szot | Jingyun Yang | Sri Somasundaram | Zhengyu Yang | Alex Yin | I-Chun Liu

Jun Yamada | Han Zhong | Shagun Uppal | Lucy Shi | Shivin Dass | Minho Heo | Doohyun Lee

External collaborators



Gaurav Sukhatme | Stefanos Nikolaidis | Peter Englert



Yuke Zhu | Anima Anandkumar

USC

NVIDIA

SK telecom

NAVER AI LAB

Scaling Robot Learning with Skills

Towards Furniture Assembly and Beyond

Youngwoon Lee / youngwoon.github.io

USC