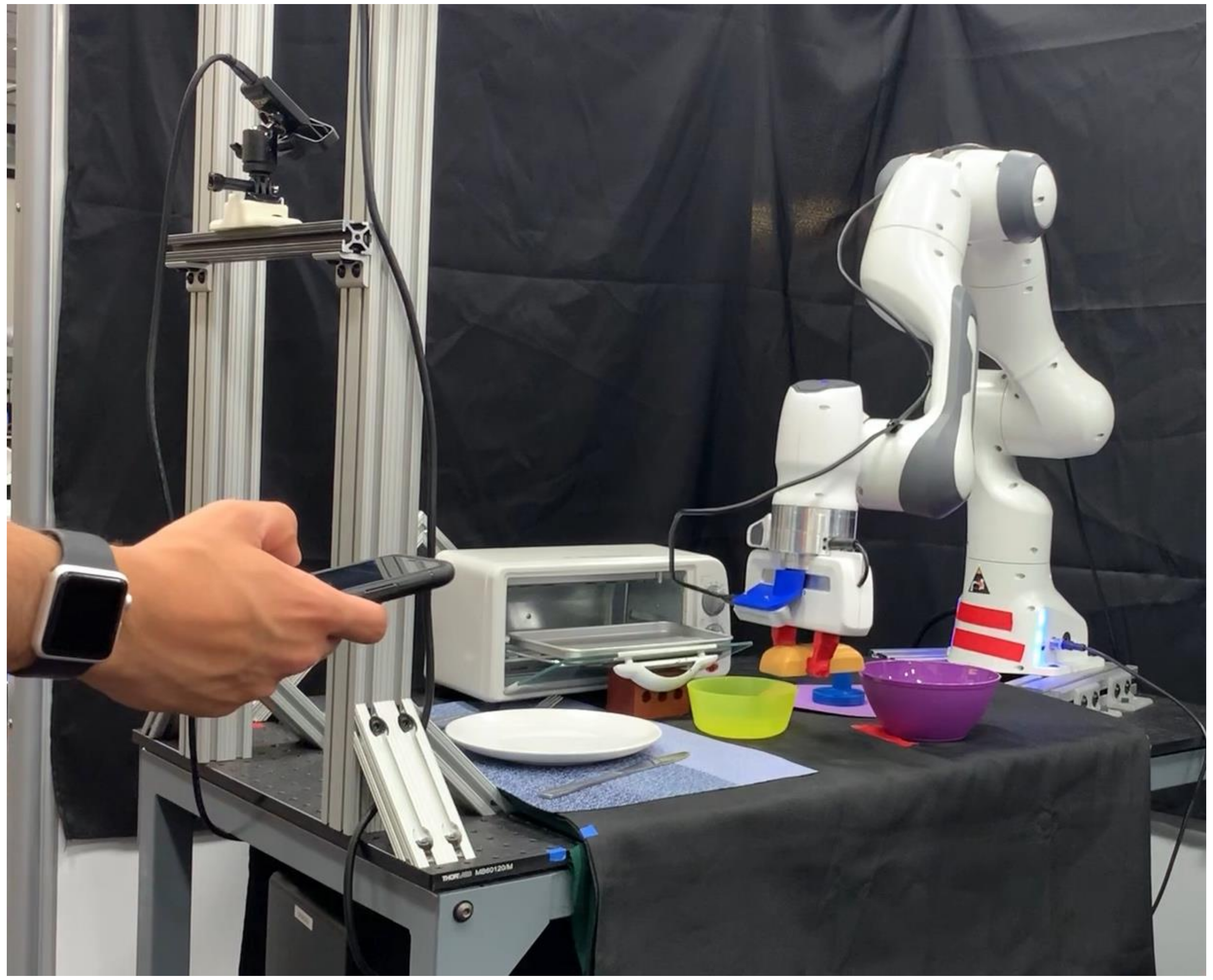
Moving from Data Collection to Data Generation: Addressing the Need for Data in Robotics

Ajay Mandlekar

NVIDIA AI Research Scientist

Imitation Learning from Human Demonstrations is a Promising Paradigm

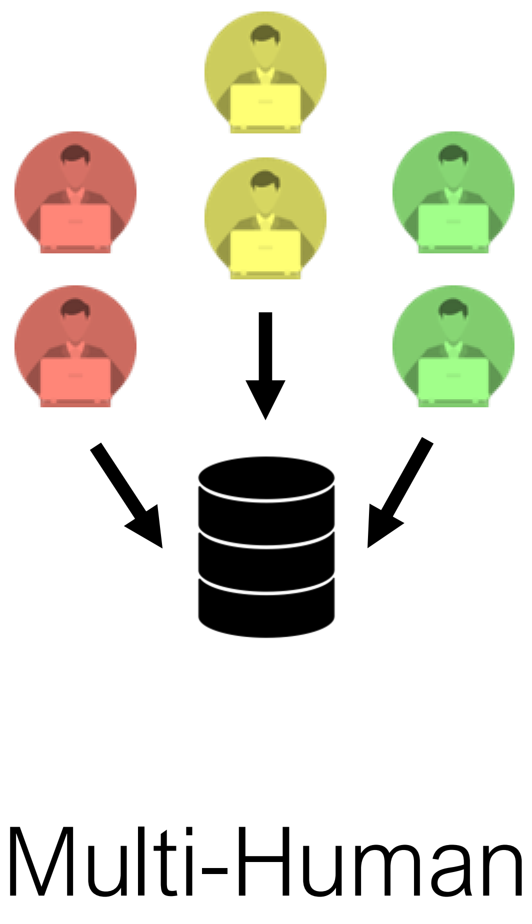
Human teleoperates robot arm to collect data for tasks of interest

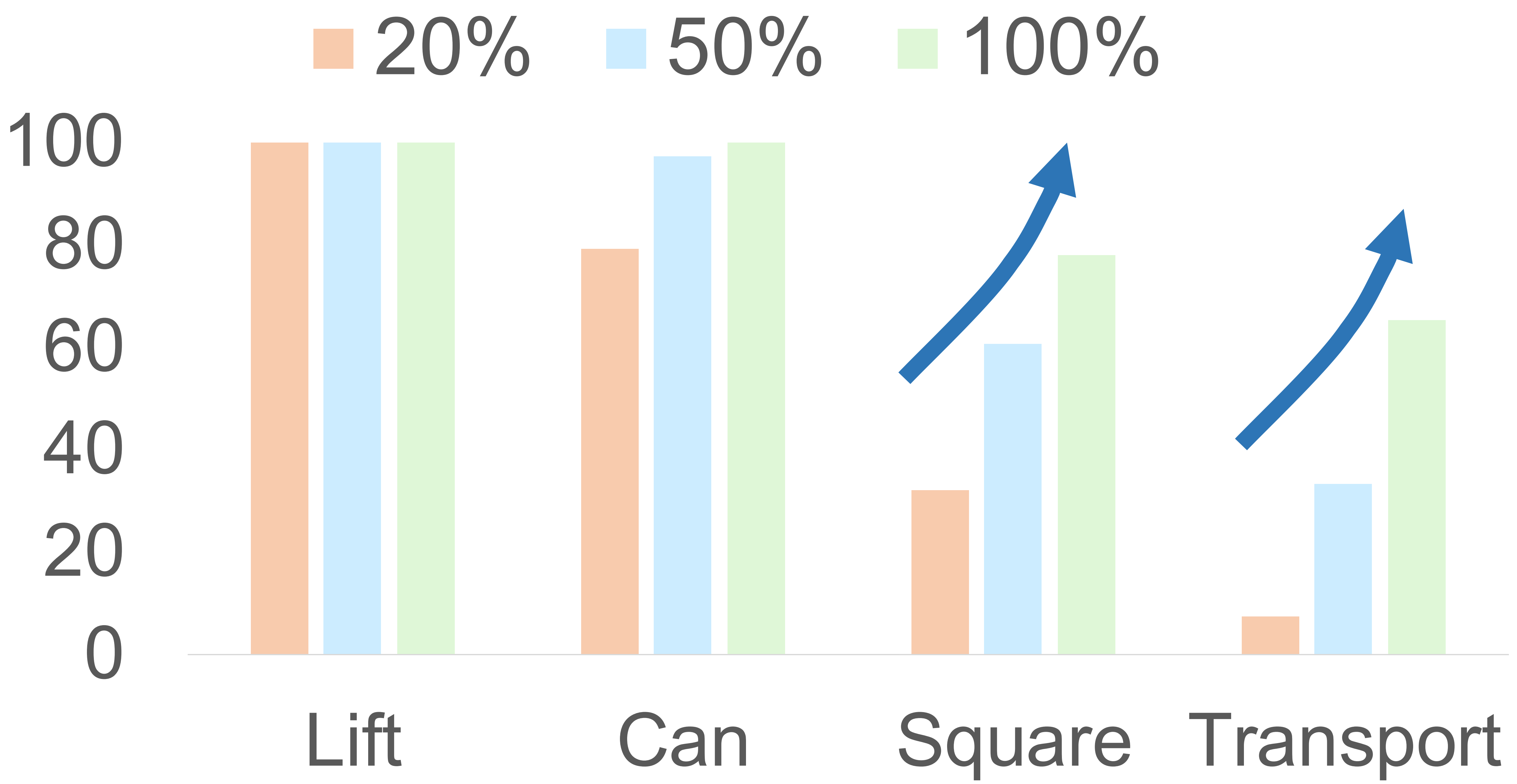


Robot learns from data to perform tasks autonomously

Mandlekar and Xu et al. “Learning to Generalize Across Long-Horizon Tasks from Human Demonstrations”, RSS 2020

Simple Recipe for Skill Learning: Scale Data to Scale Performance





Mandlekar et al. “What Matters in Learning from Offline Human Demonstrations for Robot Manipulation”, CoRL 2021

The robomimic study showed that robot performance scales with larger datasets

Simple Recipe for Skill Learning: Scale Data to Scale Performance

Brohan et al. “RT-1: Robotics Transformer for Real-World Control at Scale”, 2022

18 months of data with a large team of human contractors and robots to achieve 97% success on rearrangement tasks

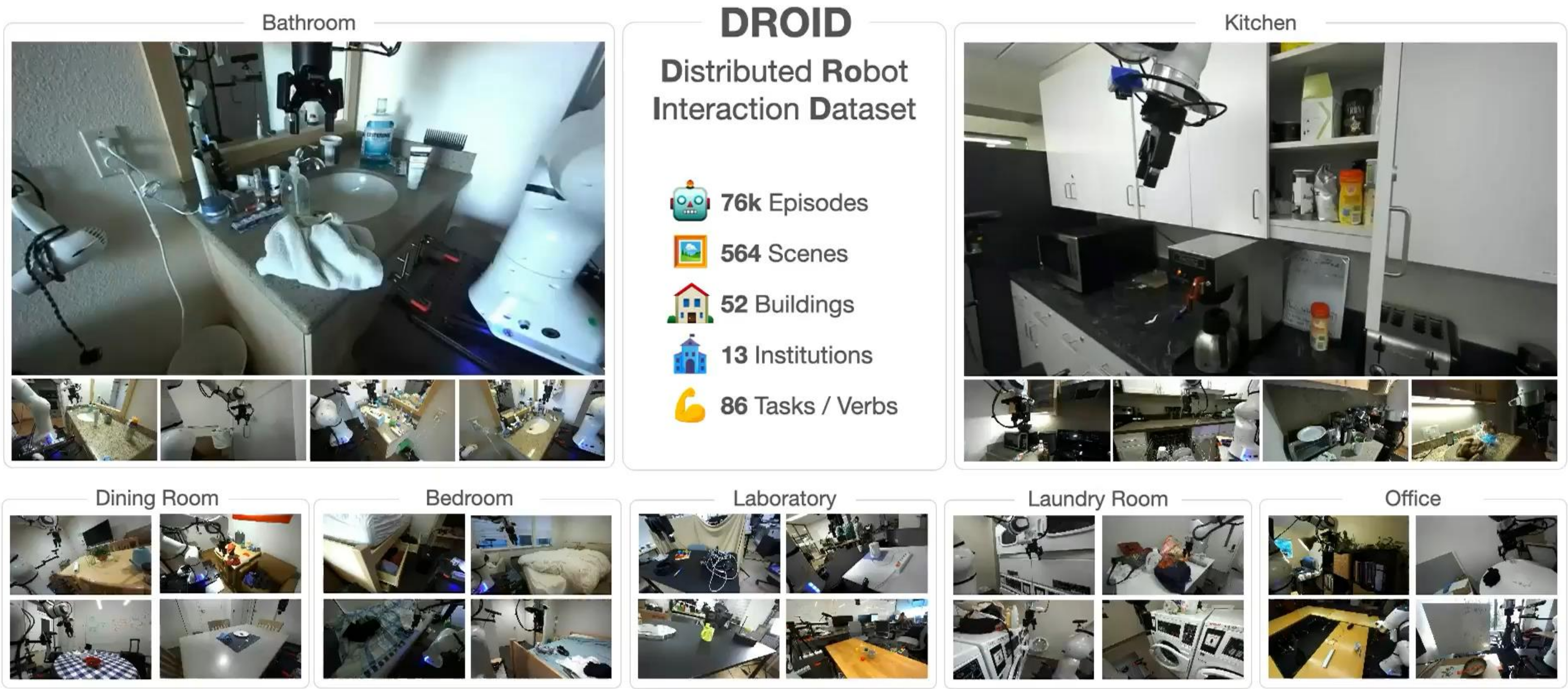
Robot learning is moving towards even larger data regimes

Open-X

Open X-Embodiment Collaboration “Open X-Embodiment: Robotic Learning Datasets and RT-X Models”, 2023

20+ academic institutions, 22 robot embodiments, 500 skills, 150,000 tasks

DROID



Khazatsky et al. “DROID: A Large-Scale In-the-Wild Robot Manipulation Dataset”, 2024

76,000 episodes, 564 scenes, 52 buildings, 13 institutions, 86 tasks

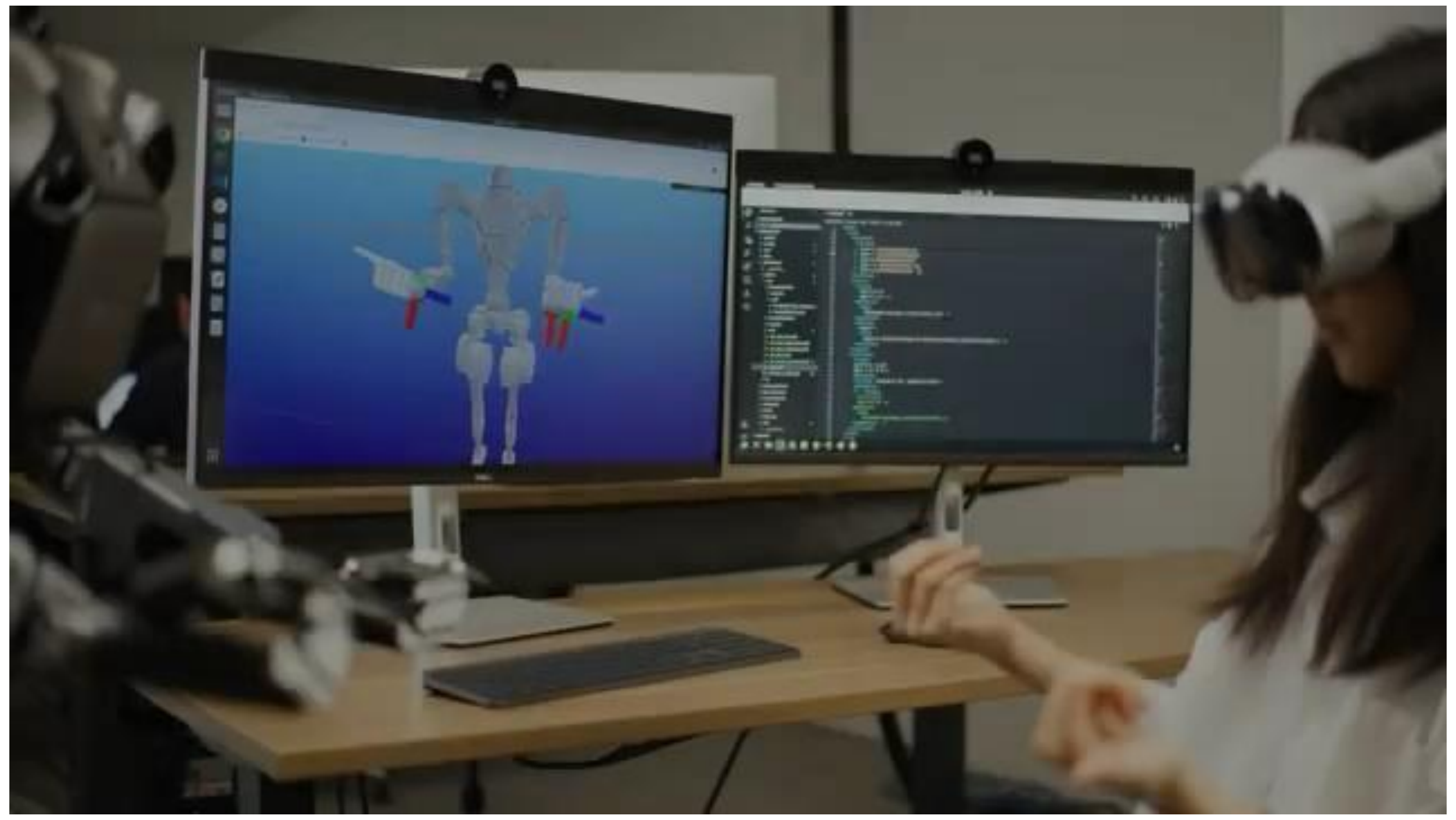
Robot learning is moving towards even larger data regimes

Tesla

1X

Physical Intelligence

NVIDIA Project GR00T



Scaling data collection requires extreme amounts of human effort

Tesla Optimus Robot Demo

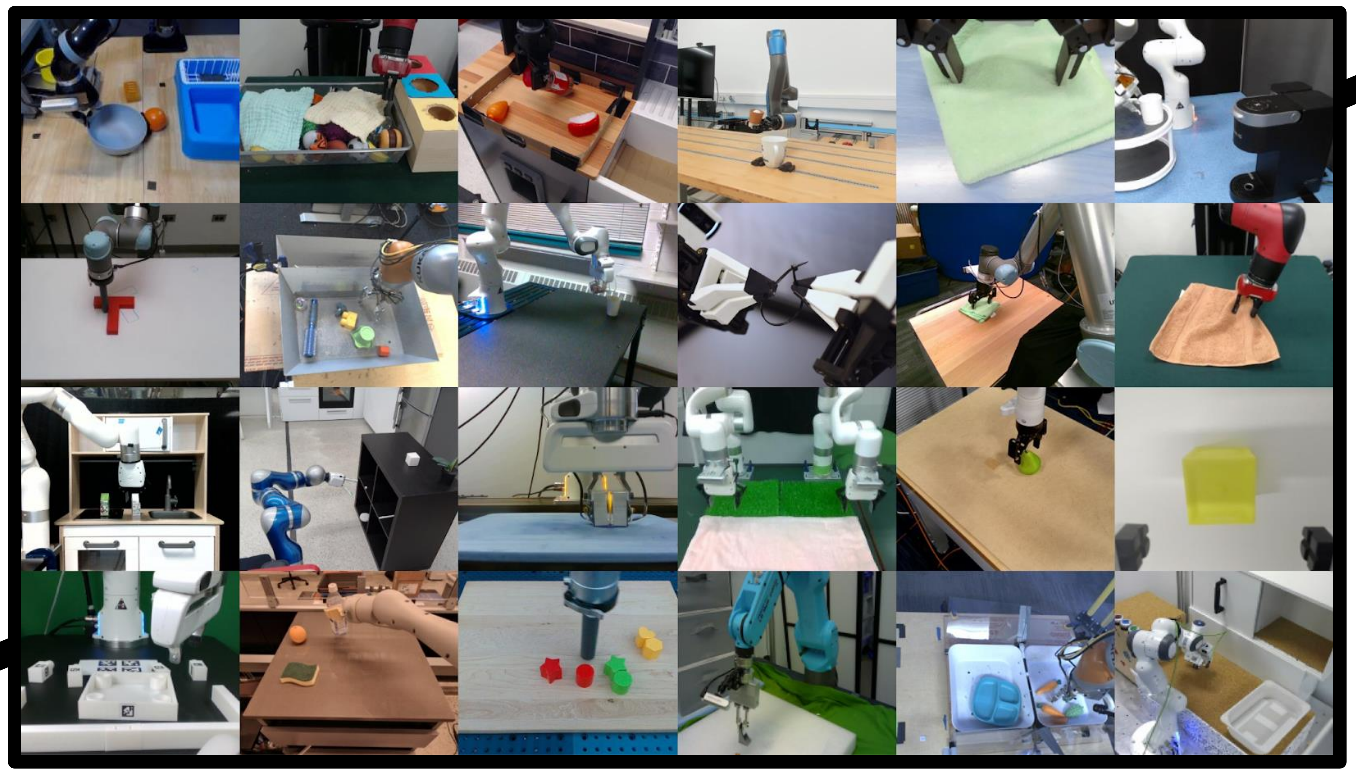
Requirements:

- Must be able to walk 7+ hours a day while carrying up to 30 lbs.
- Ability to wear and operate a motion capture suit and VR headset for extended periods of time.
- Continuous hand/eye coordination and fine manipulation, body coordination, and kinesthetic awareness and ability to walk up/downstairs.
- Must have the ability to stand, sit, walk, stoop, bend, reach, crouch, and twist throughout the day.
- Ability to work a flexible schedule: day/night shift and 1 weekend day + overtime when needed.

Source: Tesla Data Collection Operator job posting

Can we scale data collection without scaling human effort?

?



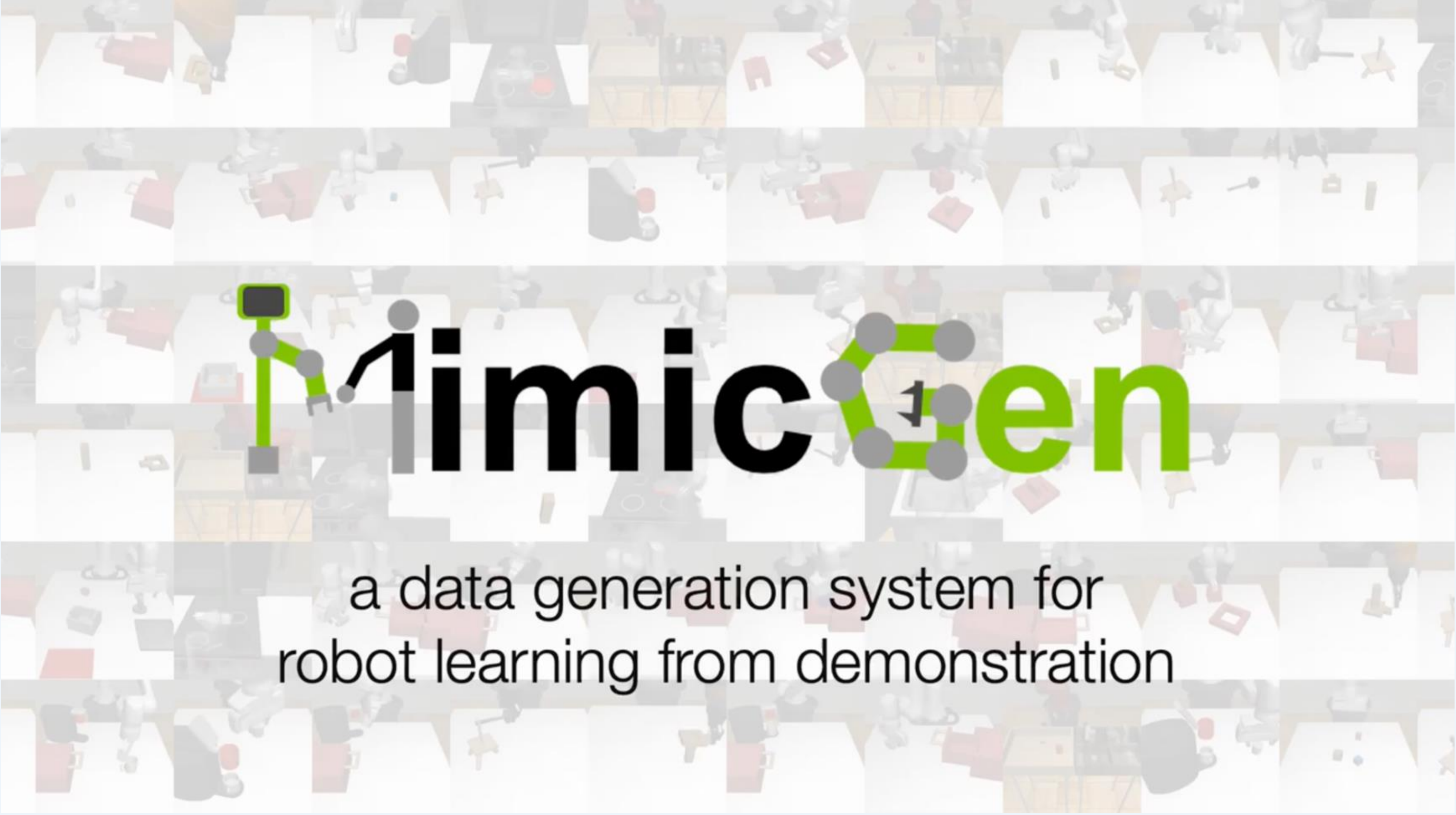
Years with multiple
academic institutions

18 months with several
human operators

Simulation is a compelling alternative to real-world data collection

Scalable data generation

Synthesize diverse, high-quality robot demonstrations autonomously



[Mandlekar et al. 2023]

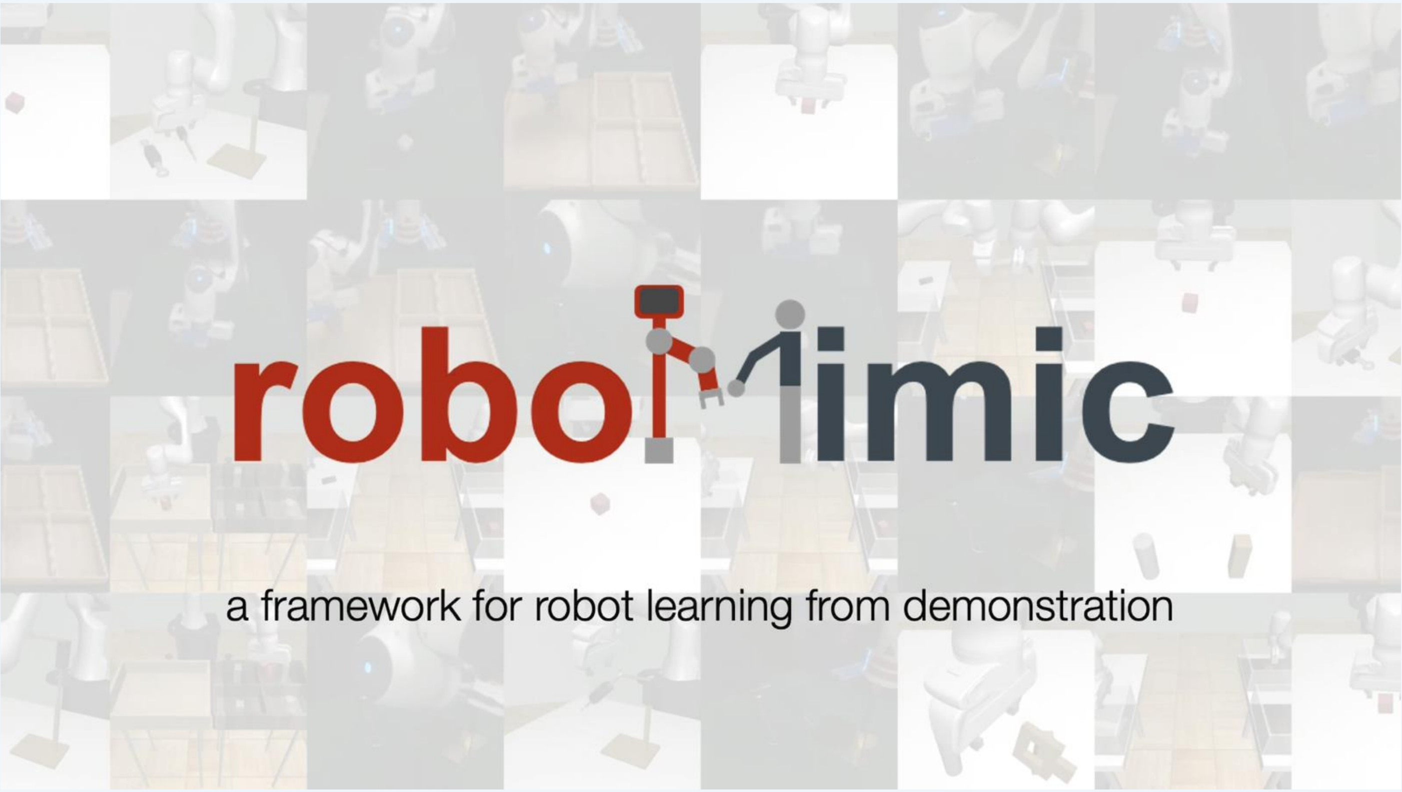
Infinite procedural generation

Procedurally generate scenes, objects, tasks with the aid of generative AI tools

[Nasiriany et al. 2024]

Ease-of-use and reproducibility

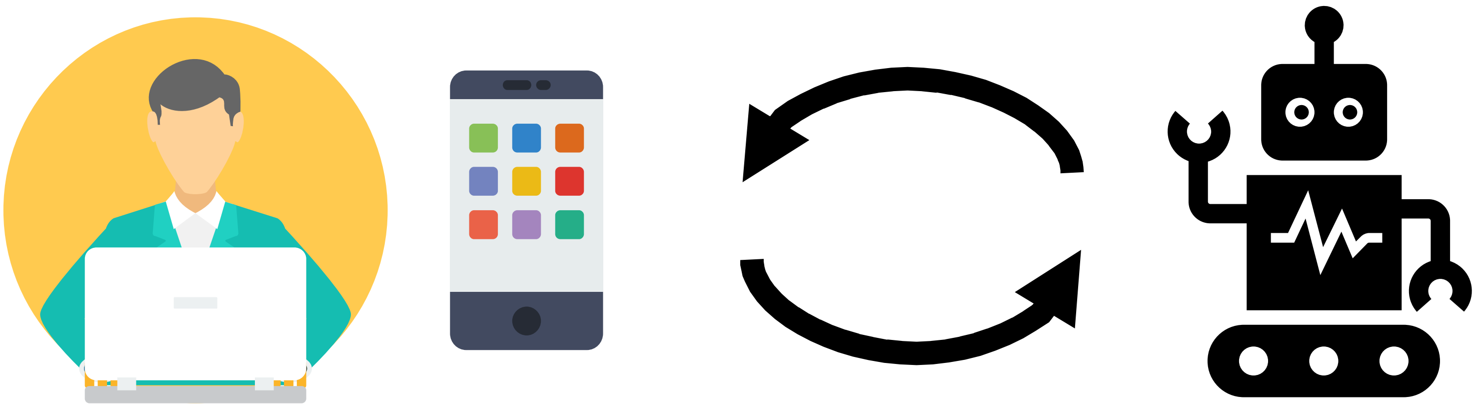
Reliably prototype algorithms, create benchmarks, and reproduce results



[Mandlekar et al. 2021]

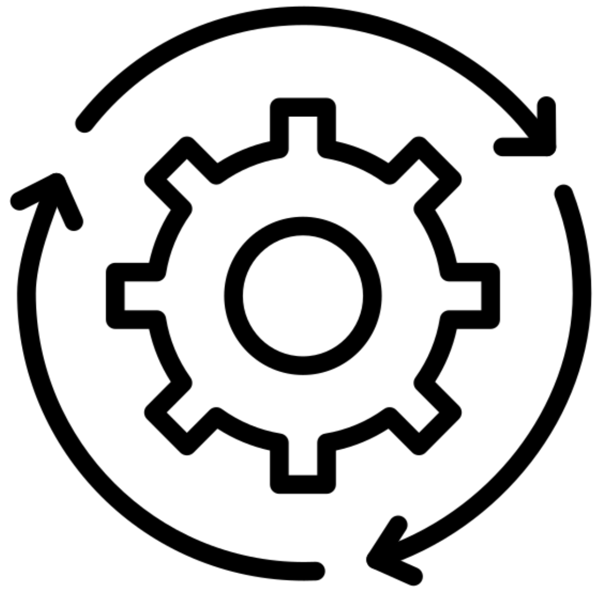
Moving from data collection to data generation

Data Collection



Months of persistent and costly human effort

Data Generation

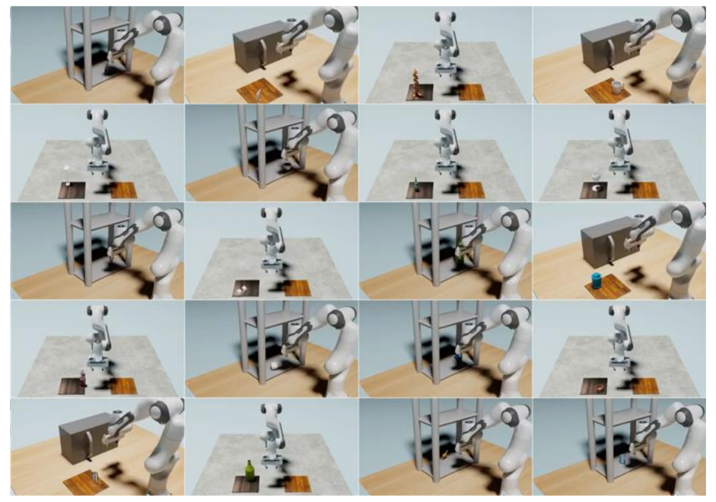


Little to no human effort

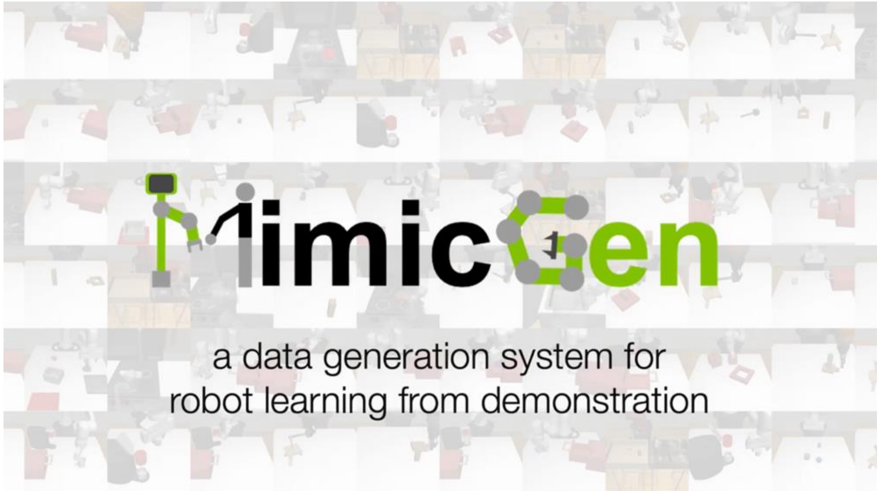
Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations



OPTIMUS (CoRL 2023)

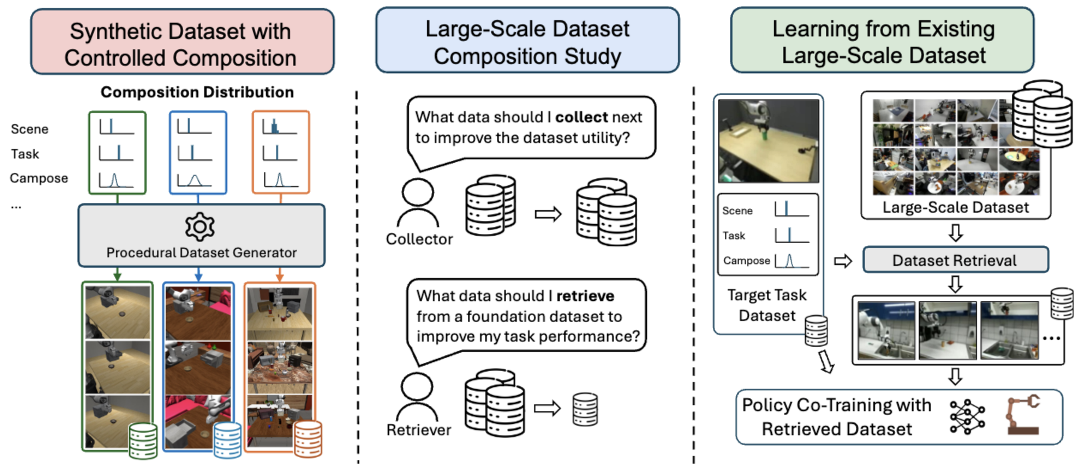


MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)



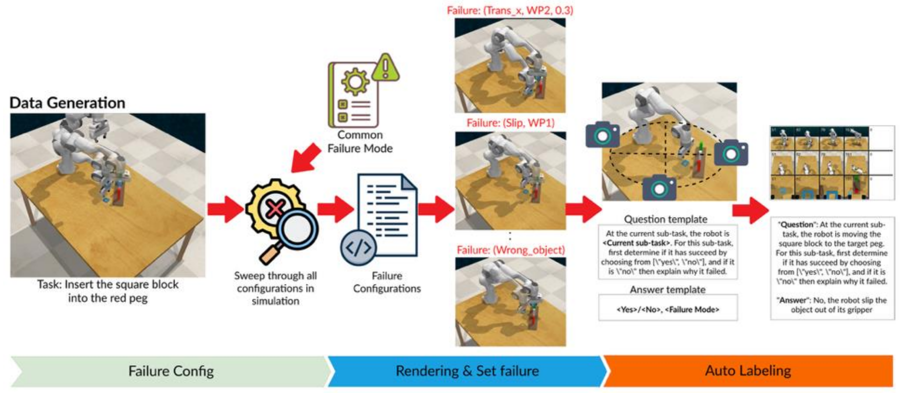
MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen (ICRA 2025)

SkillMimicGen (CoRL 2024)

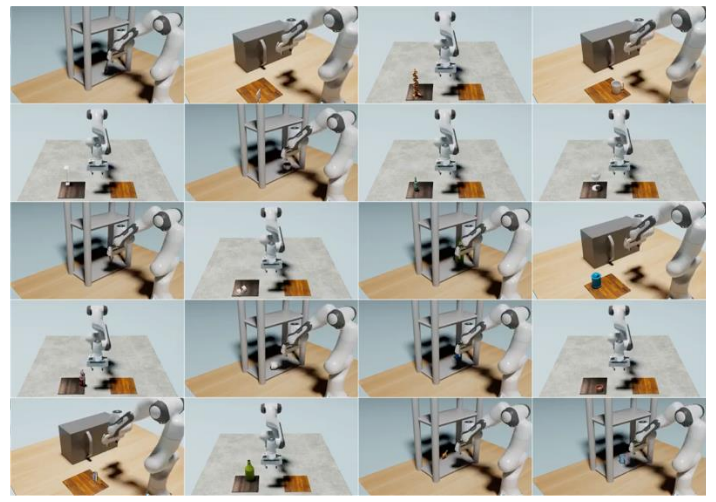


AHA (ICLR 2025)

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations



OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)

MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen (ICRA 2025)

SkillMimicGen (CoRL 2024)

AHA (ICLR 2025)

Task and Motion Planning (TAMP)

Solving Robot Manipulation with Optimization

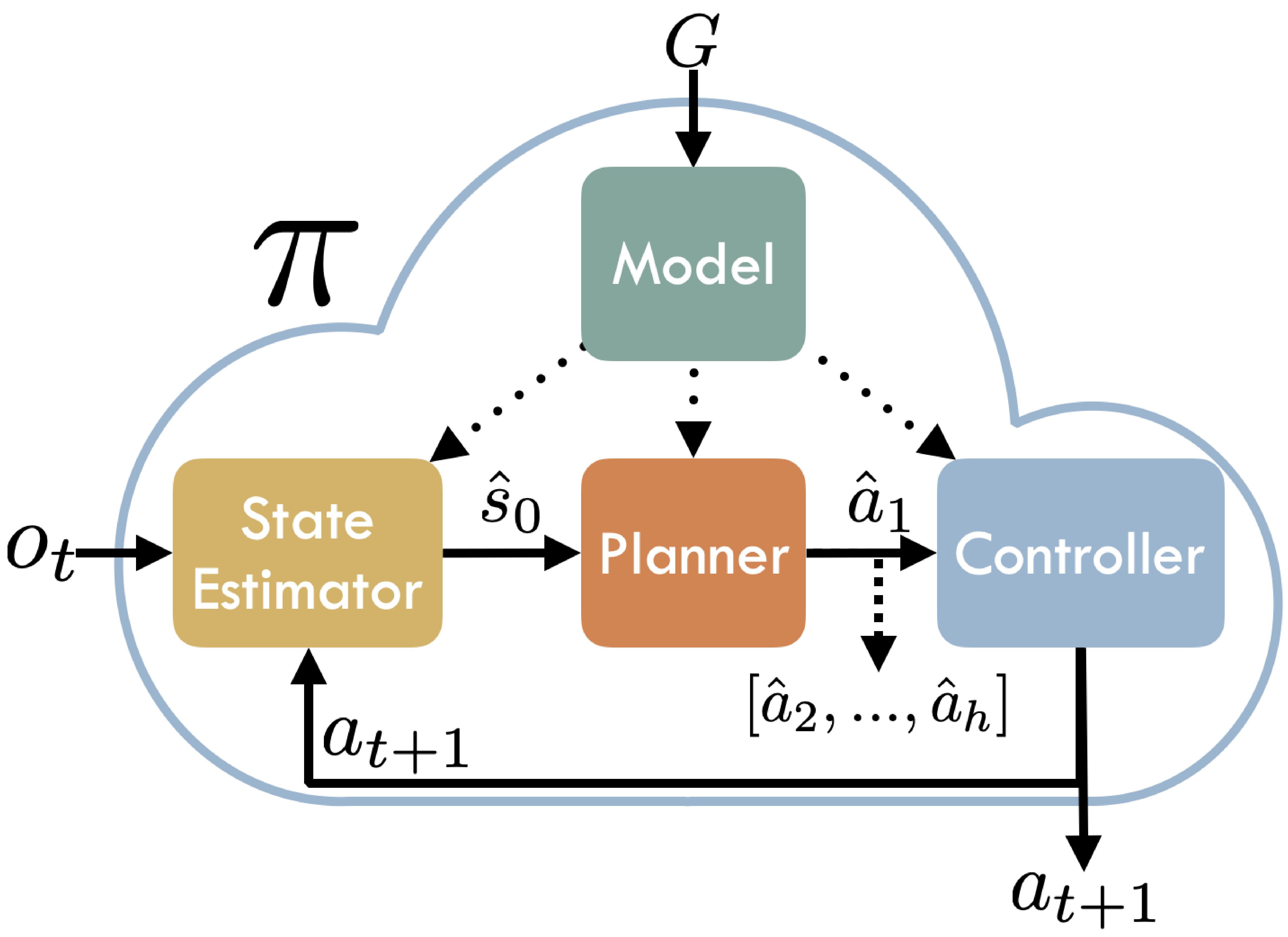
Use world models and optimization to solve long-term objectives

- Goal: move the can of spam from the drawer to the cabinet
- TAMP plans a sequence of robot motions to achieve this goal: open the cabinet, grasp the can, place can on table, re-grasp the can, place can in cabinet
- TAMP understands that different grasps are required at different parts of the task, so the can is placed on the table temporarily

Garrett et al. “Integrated Task and Motion Planning”, 2021

Task and Motion Planning (TAMP)

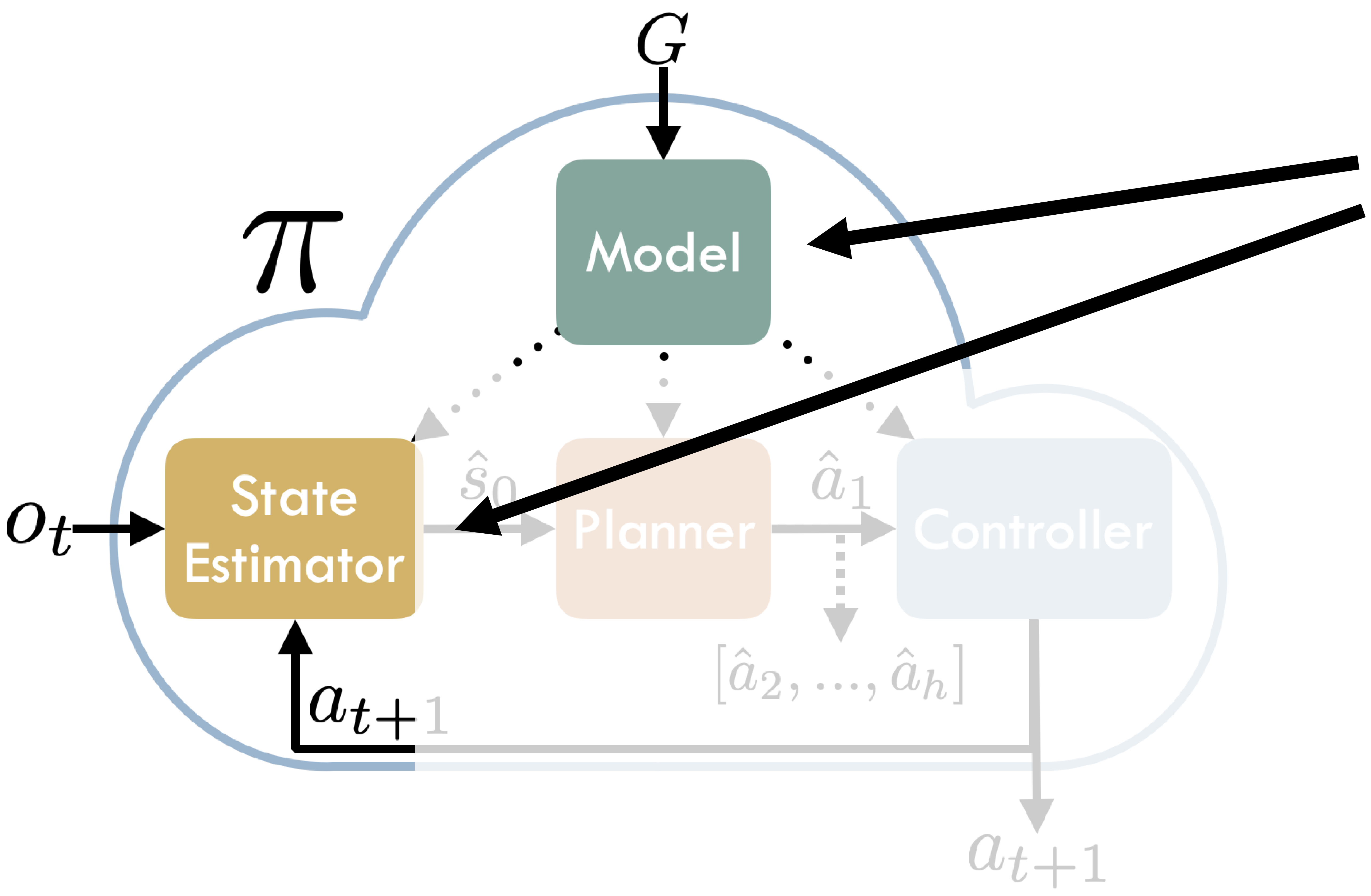
Solving Robot Manipulation with Optimization



- Positives: easily applicable to a wide range of initial states, goal conditions, long-horizon tasks, and can handle substantial geometry variation.
- Requirements: a set of parametrized skills and knowledge of their effects (e.g. move, grasp, place, open, close), and some form of state estimation as well (meshes, poses, etc)

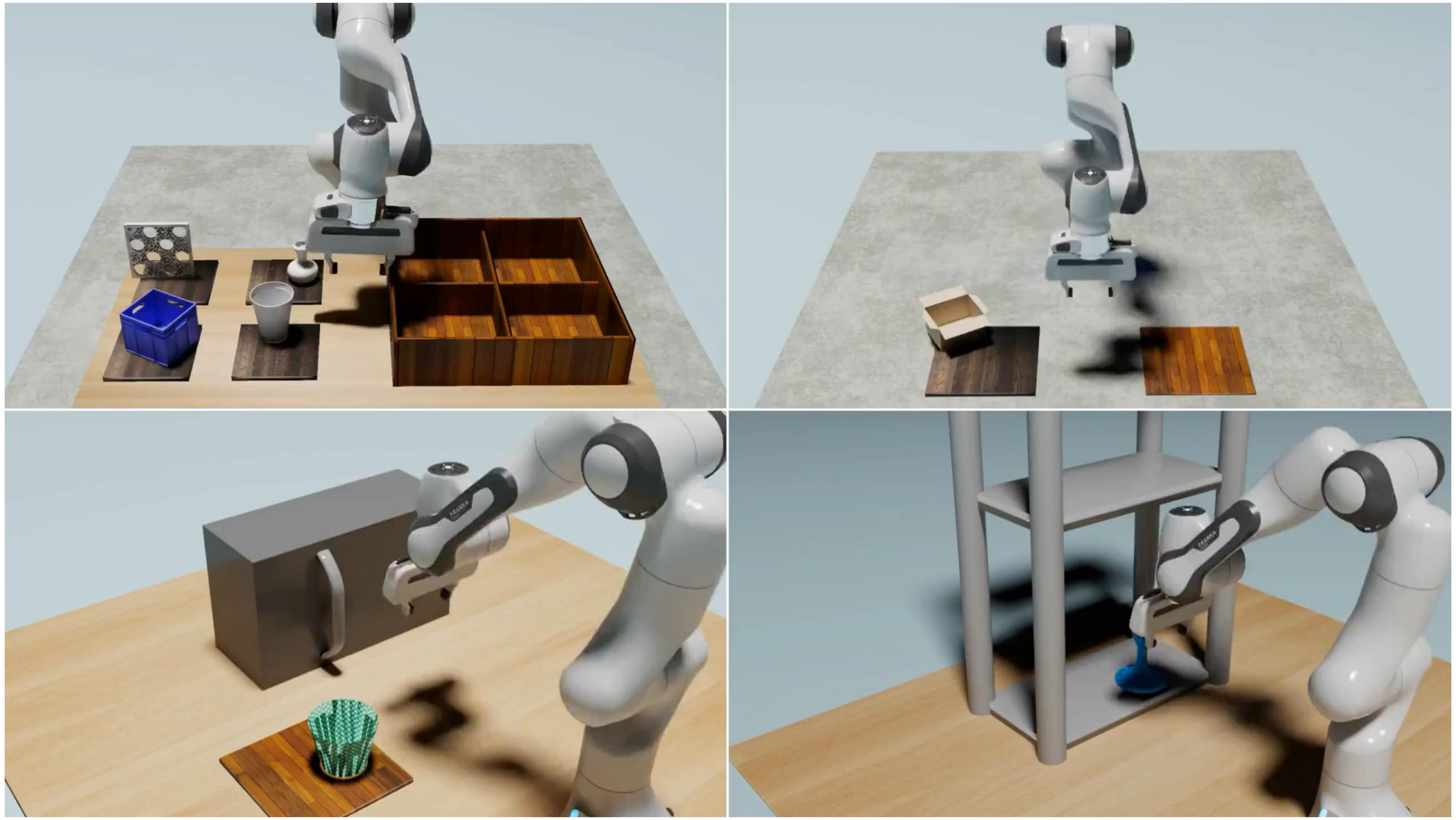
Task and Motion Planning (TAMP)

Solving Robot Manipulation with Optimization



Readily available in simulation!

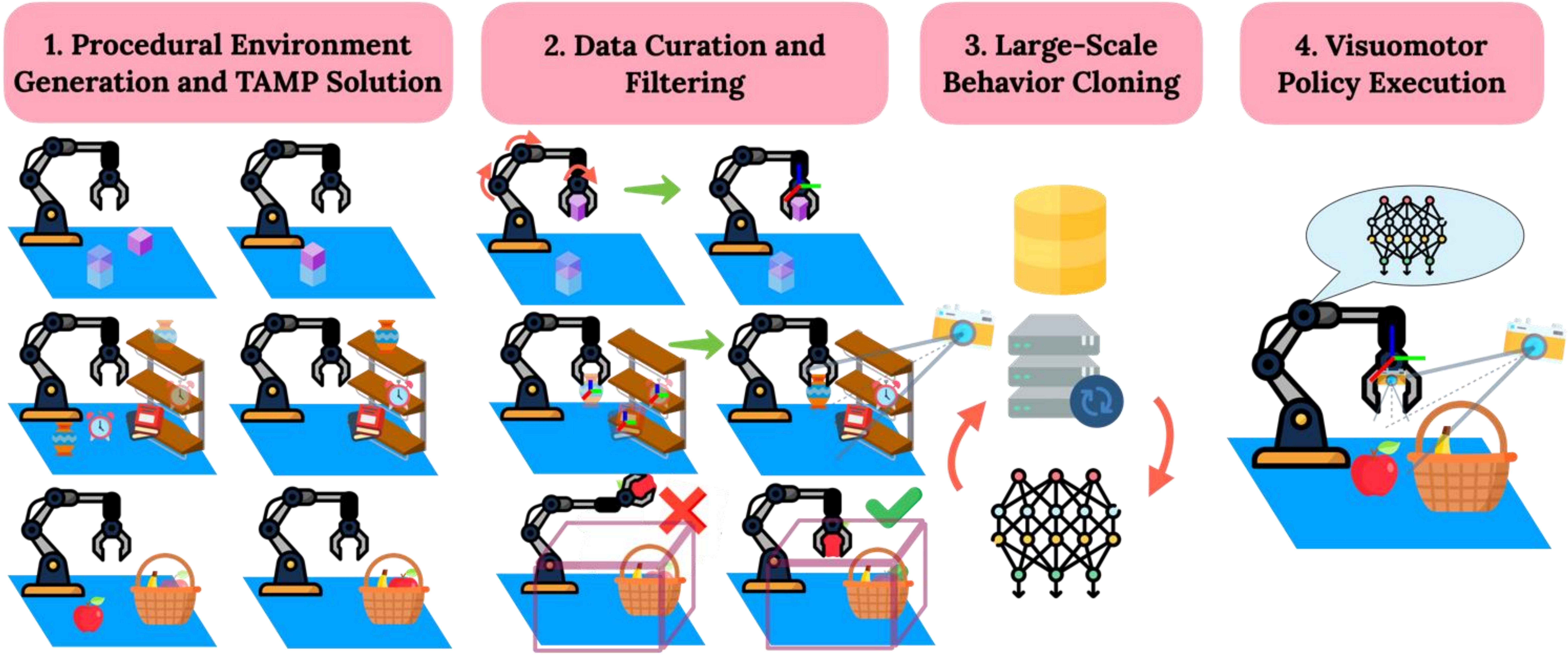
Idea: Use TAMP as a data generator in simulation



Apply TAMP to several procedurally generated simulation environments for a rich source of demonstration data!

Dalal et al. “Imitating Task and Motion Planning with Visuomotor Transformers”, CoRL 2023

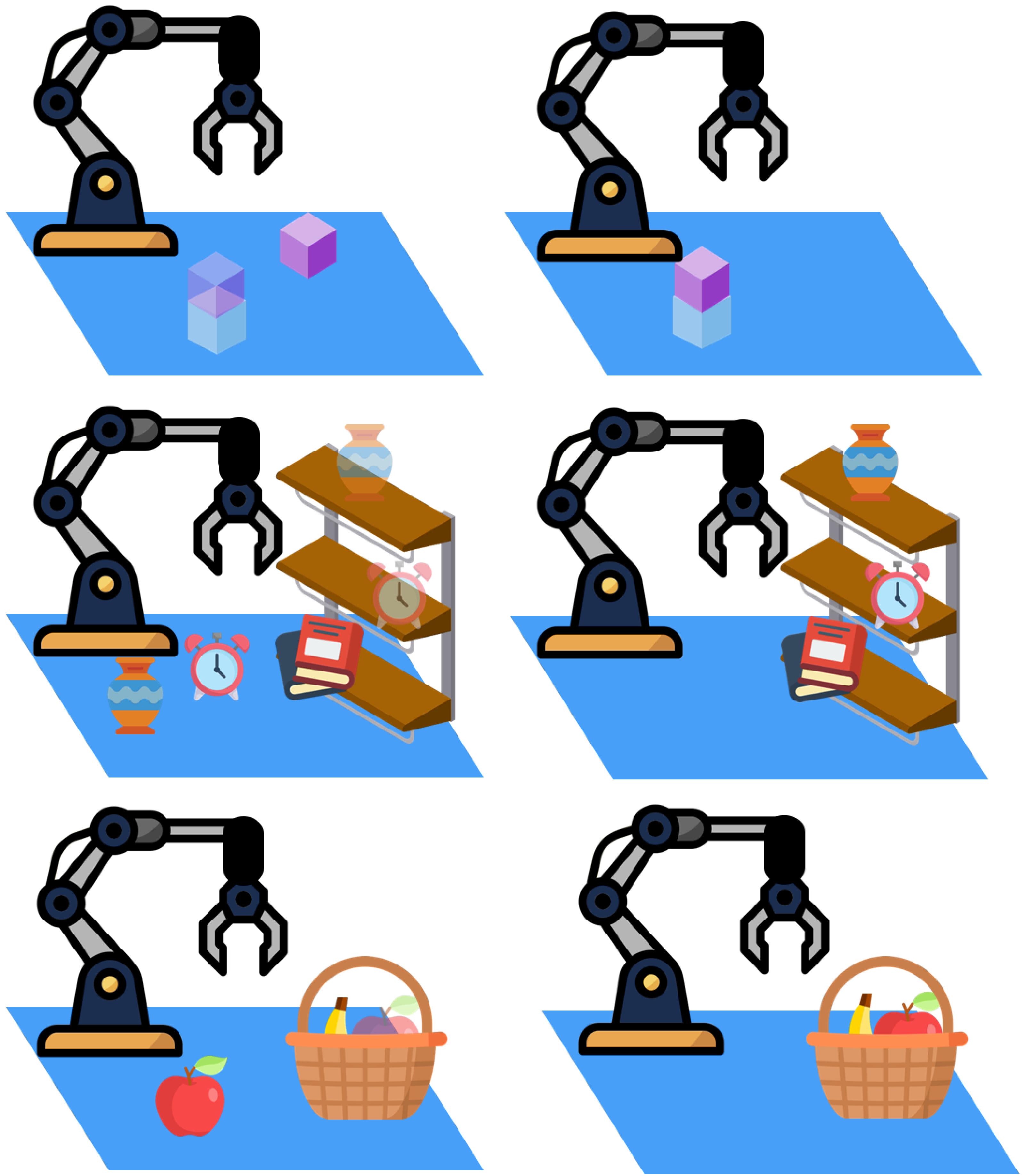
OPTIMUS: Offline Pretrained TAMP Imitation System

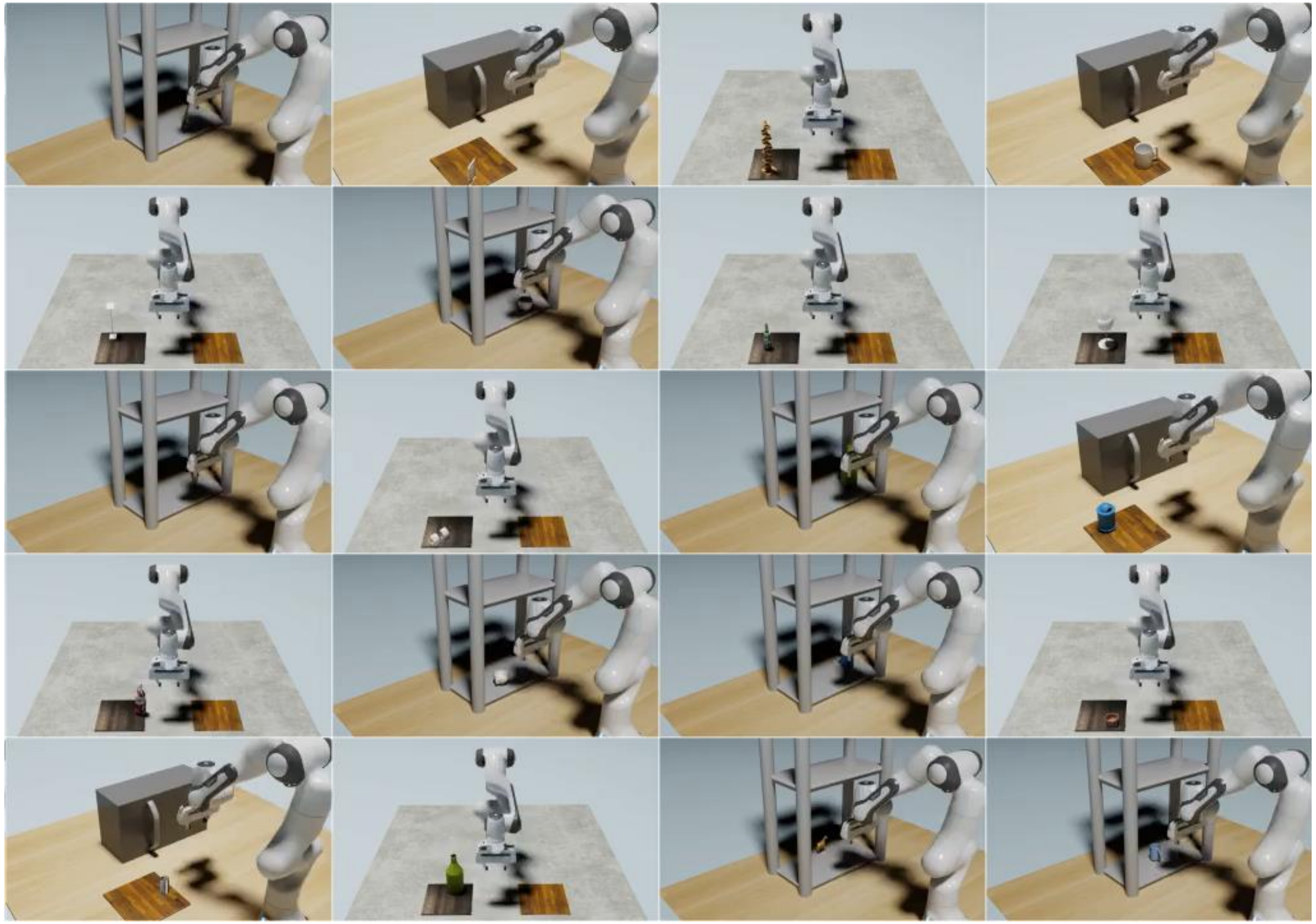


Dalal et al. “Imitating Task and Motion Planning with Visuomotor Transformers”, CoRL 2023

Procedural Environment and TAMP Solution Generation

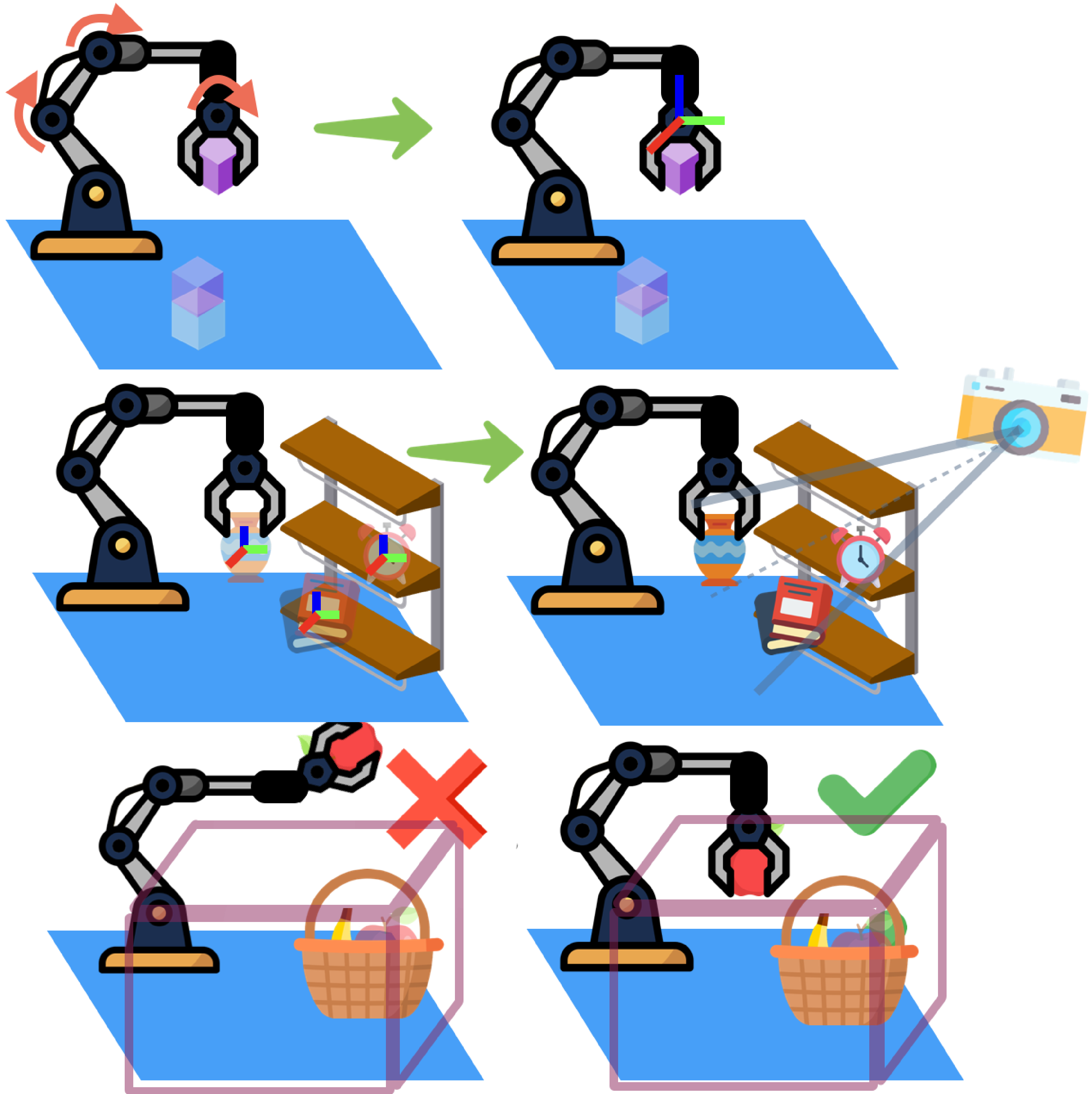
1. Procedural Environment Generation and TAMP Solution





Curating TAMP demonstrations to enable Imitation Learning

2. Data Curation and Filtering



1. Transform TAMP demonstrations into task space

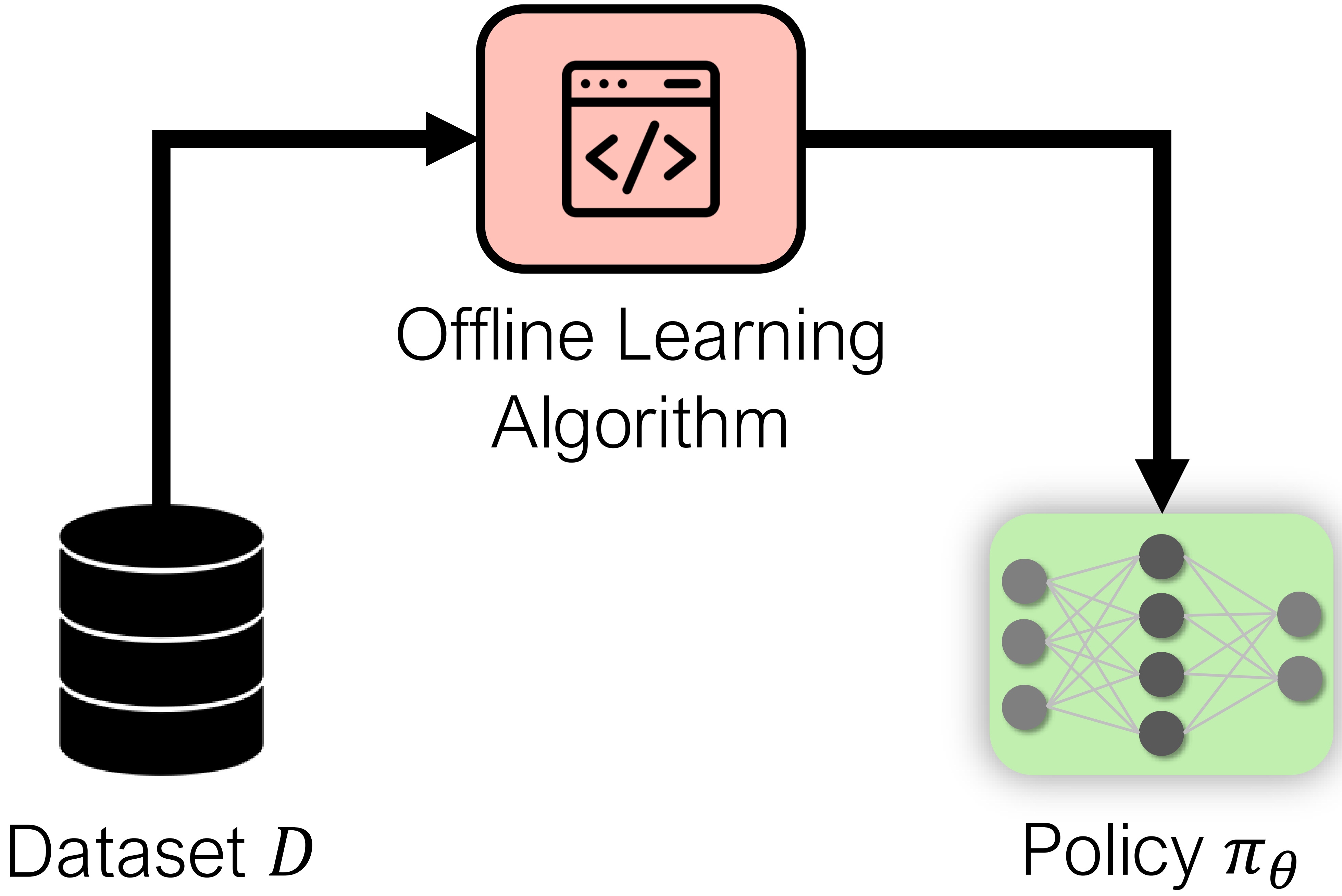
2. Use well-tuned camera views + wrist camera

3. Filter out out-of-distribution TAMP trajectories

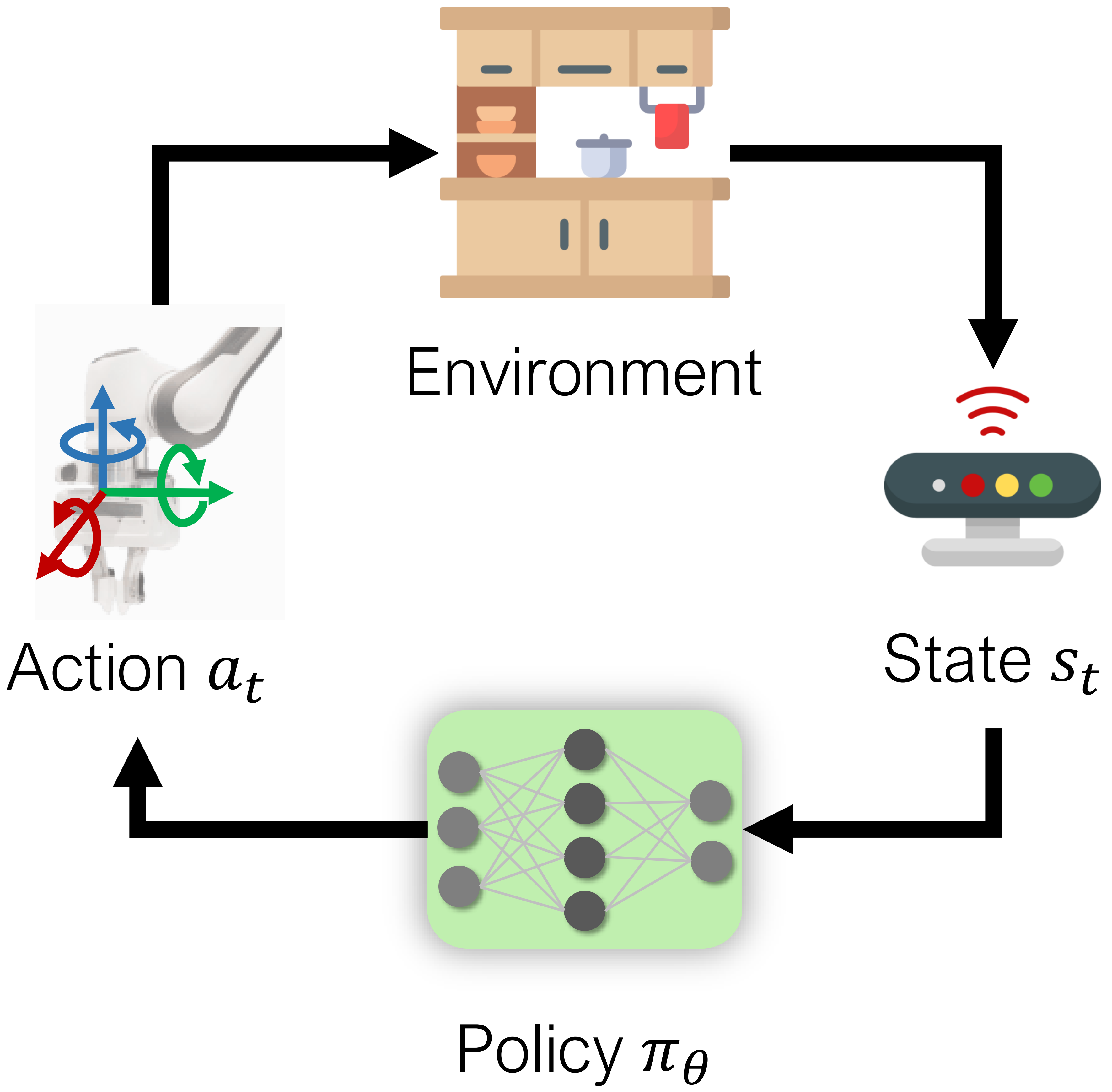
Offline Policy Learning

Goal: train closed-loop policy that performs well on task

Offline Training

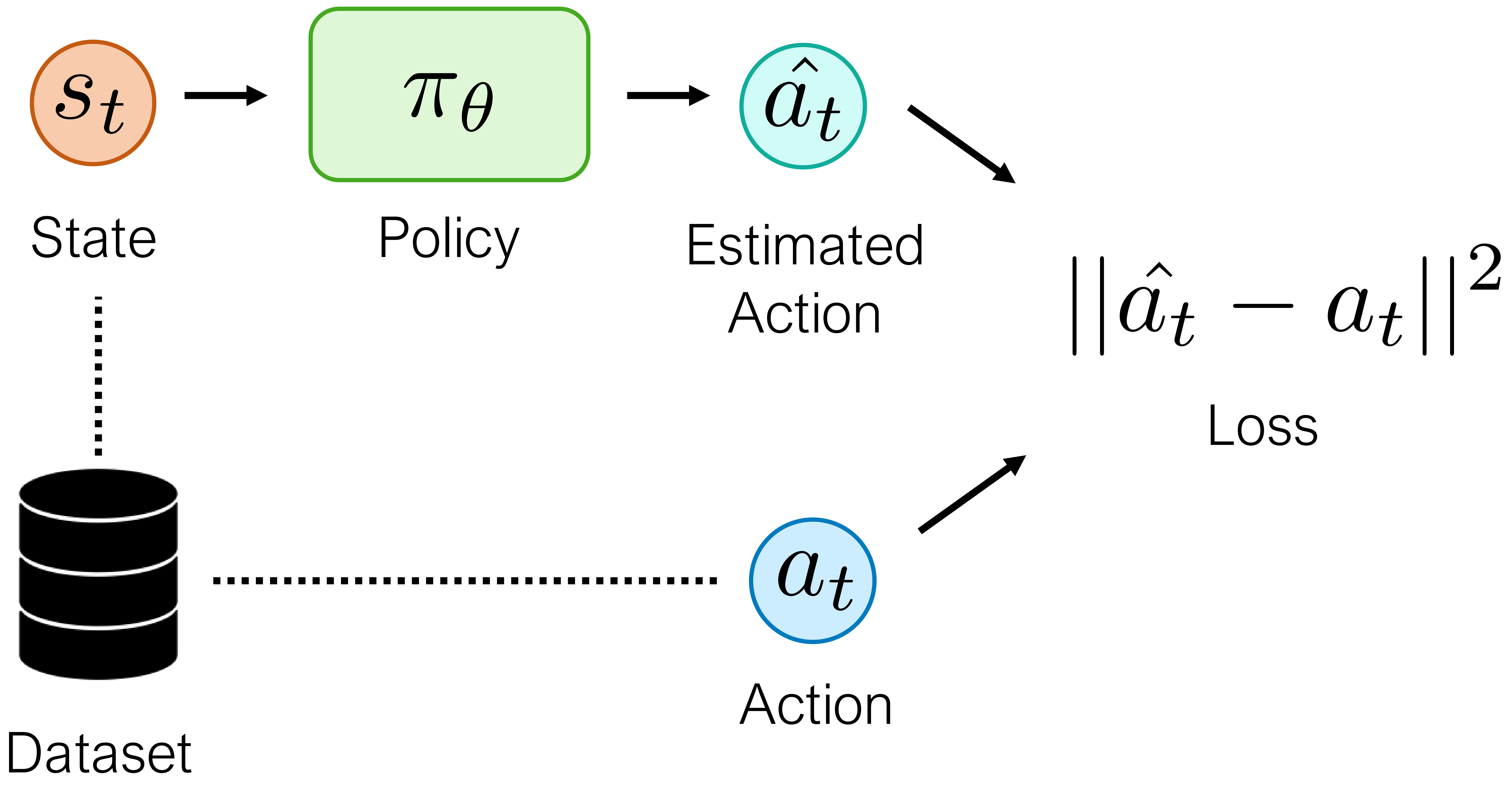


Online Evaluation



Imitation Learning: Behavioral Cloning

Key Idea: Copy all actions

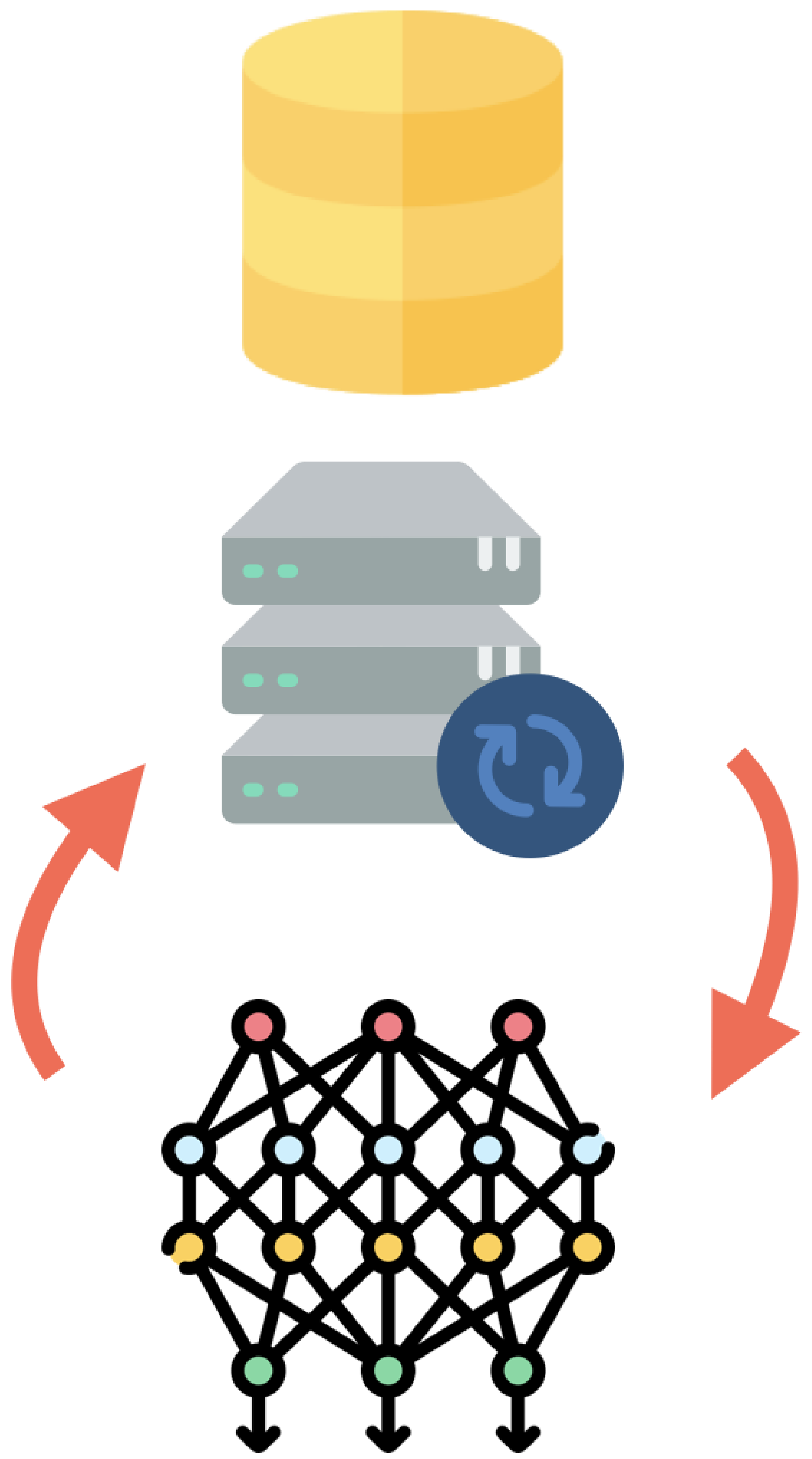


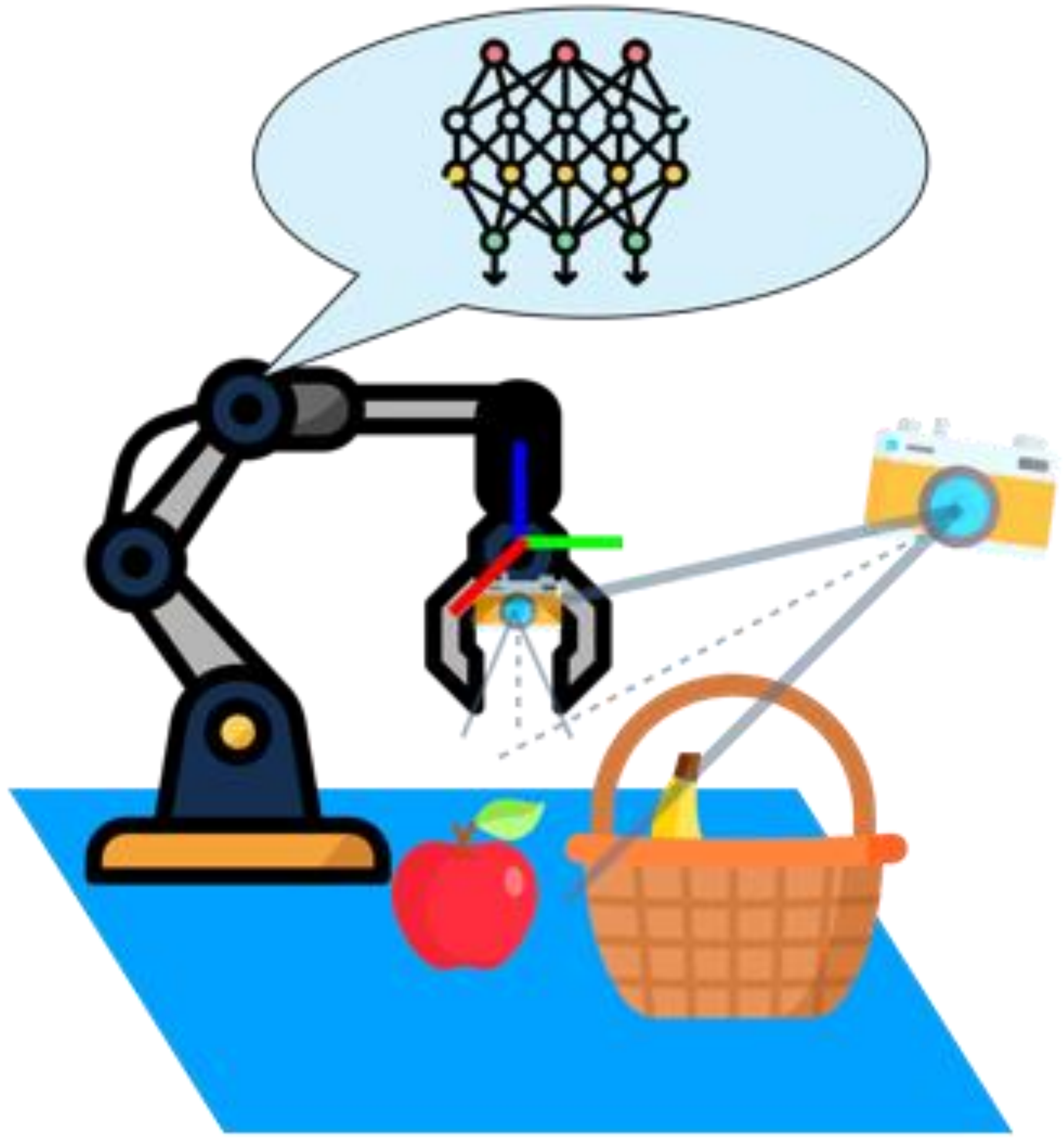
[Pomerleau et al. (1989)]

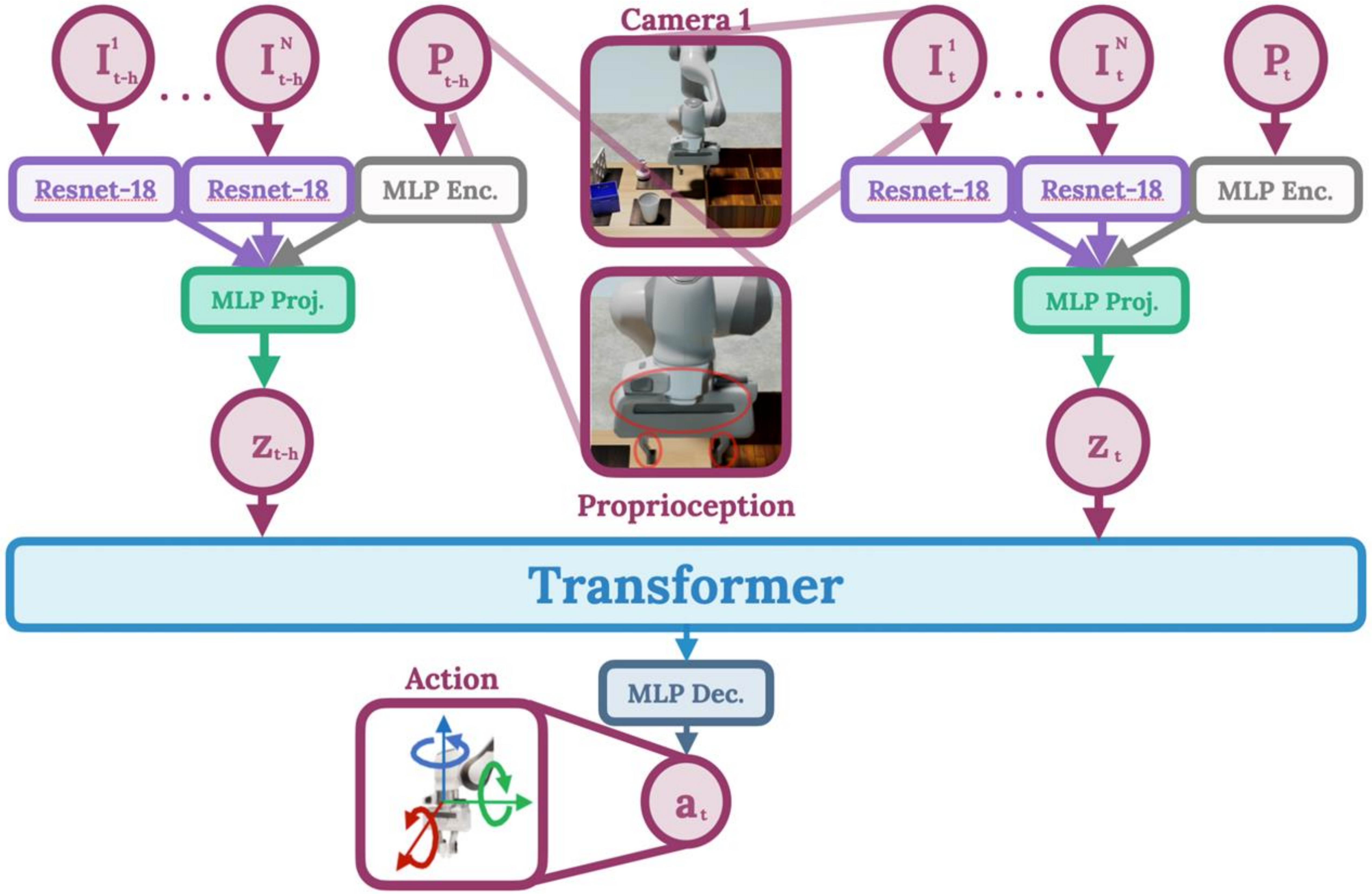
OPTIMUS Policy Architecture

3. Large-Scale Behavior Cloning

4. Visuomotor Policy Execution







Dalal et al. “Imitating Task and Motion Planning with Visuomotor Transformers”, CoRL 2023

OPTIMUS produces policies that can solve over 300 manipulation tasks with up to 8 stages and 72 different objects

OPTIMUS can solve manipulation tasks with up to 8 stages

PickPlaceFour: 60% Success Rate

OPTIMUS can distill TAMP's task planning capabilities

MicrowaveAdapt: 75%

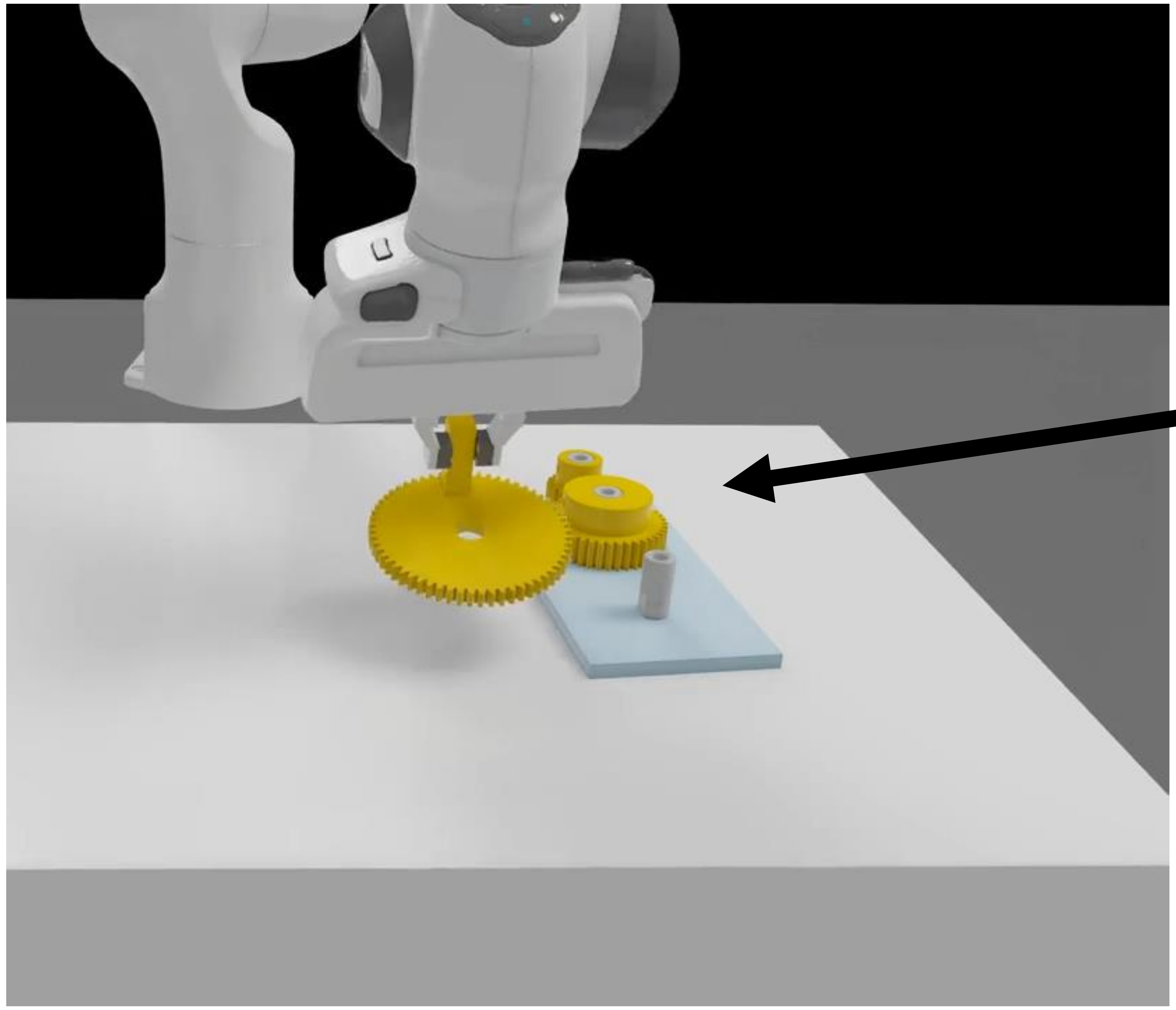
OPTIMUS Scales to Many Different Objects and Tasks

Dataset	BC-MLP	BC-RNN	BeT	OPTIMUS
PickPlace-1	94	97	85	**100**
PickPlace-19	61	58	50	**85**
PickPlace-72	50	49	41	**75**
Shelf-1	**91**	88	70	**91**
Shelf-19	48	31	26	**66**
Shelf-72	30	36	13	**48**
Microwave-1	73	77	51	**86**
Microwave-19	24	41	31	**61**
Microwave-72	23	29	16	**47**

Dalal et al. “Imitating Task and Motion Planning with Visuomotor Transformers”, CoRL 2023

Limitation: TAMP can struggle with contact-rich tasks

Recall: TAMP needs accurate world models and set of skills

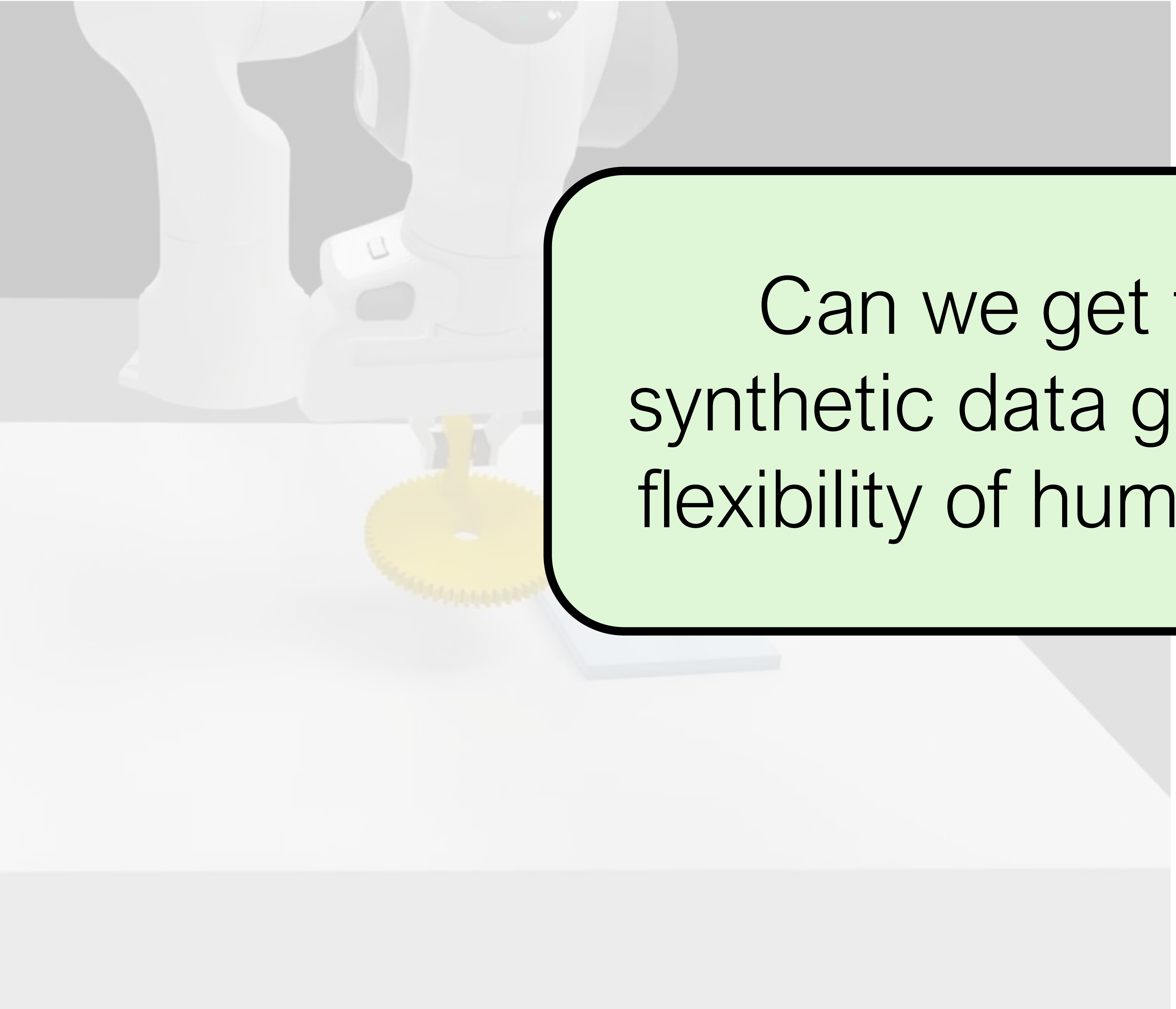


Requires designing skills capable of millimeter-level precision!

Human teleoperation is more flexible: just demonstrate the skill!

Limitation: TAMP can struggle with contact-rich tasks

Recall: TAMP needs accurate world models and set of skills



Can we get the benefits of synthetic data generation with the flexibility of human teleoperation?

esigning skills capable of
eter-level precision!

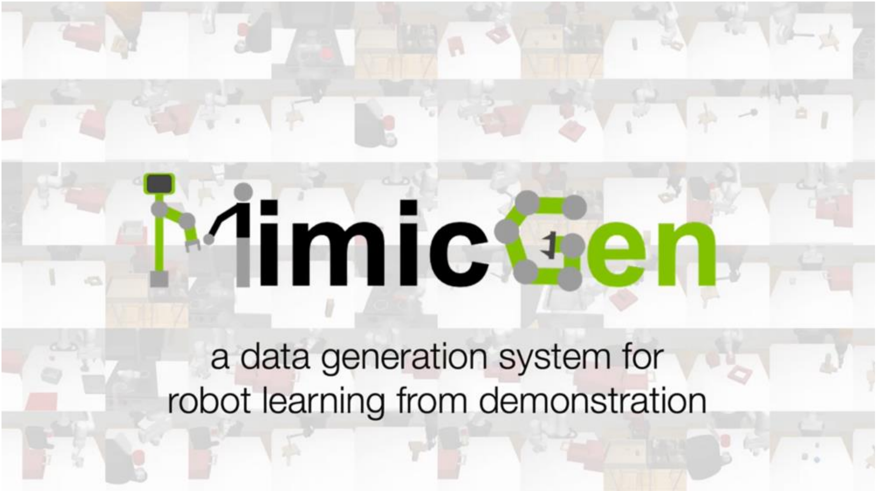
Human teleoperation is more flexible: just demonstrate the skill!

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations

OPTIMUS (CoRL 2023)



MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)

MimicLabs (ICLR 2025)

Building More Powerful Data Generators

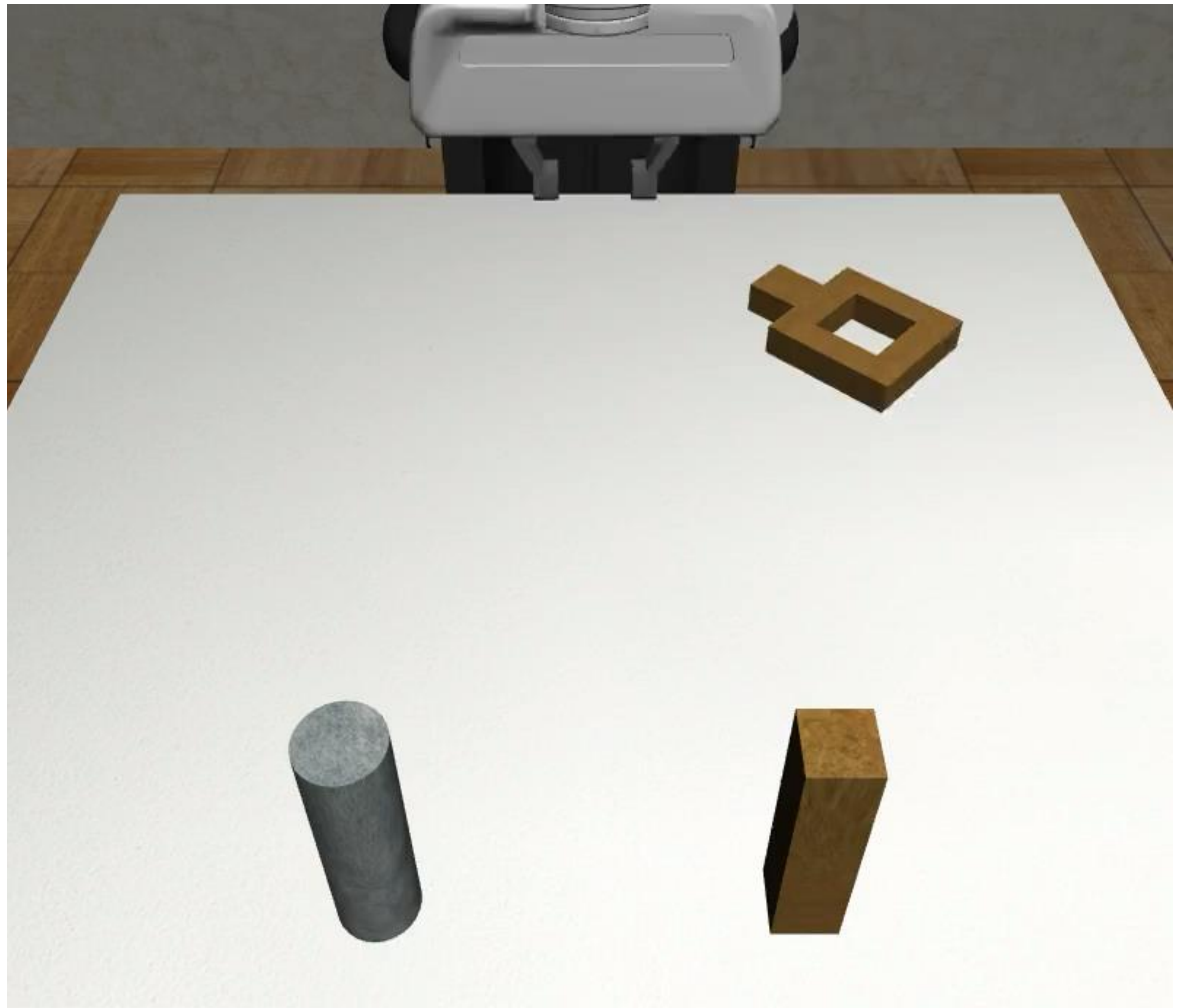
- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen (ICRA 2025)

SkillMimicGen (CoRL 2024)

AHA (ICLR 2025)

Single task learning requires too much human effort



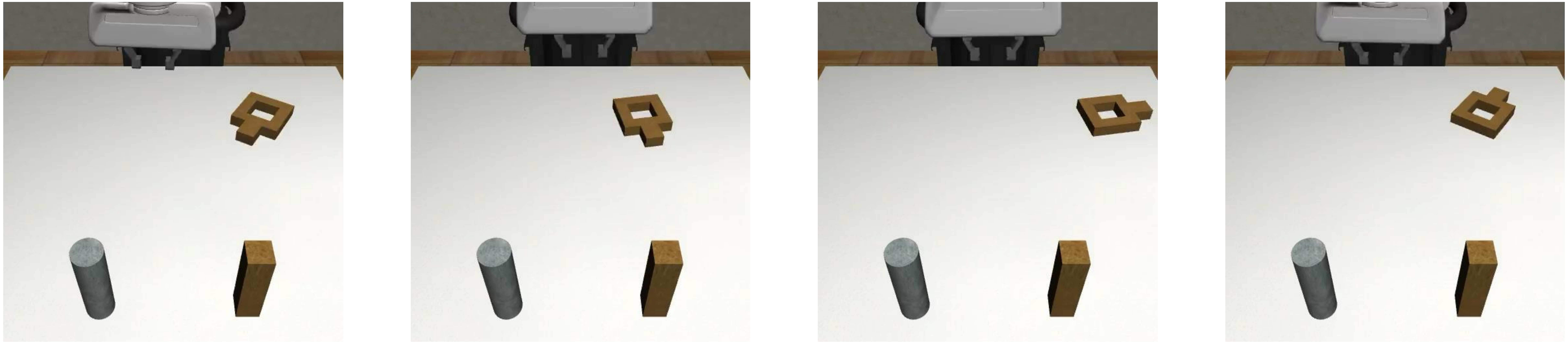
Mandlekar et al. “What Matters in Learning from Offline Human Demonstrations for Robot Manipulation”, CoRL 2021

200 human demonstrations

~1 hour of human operator time

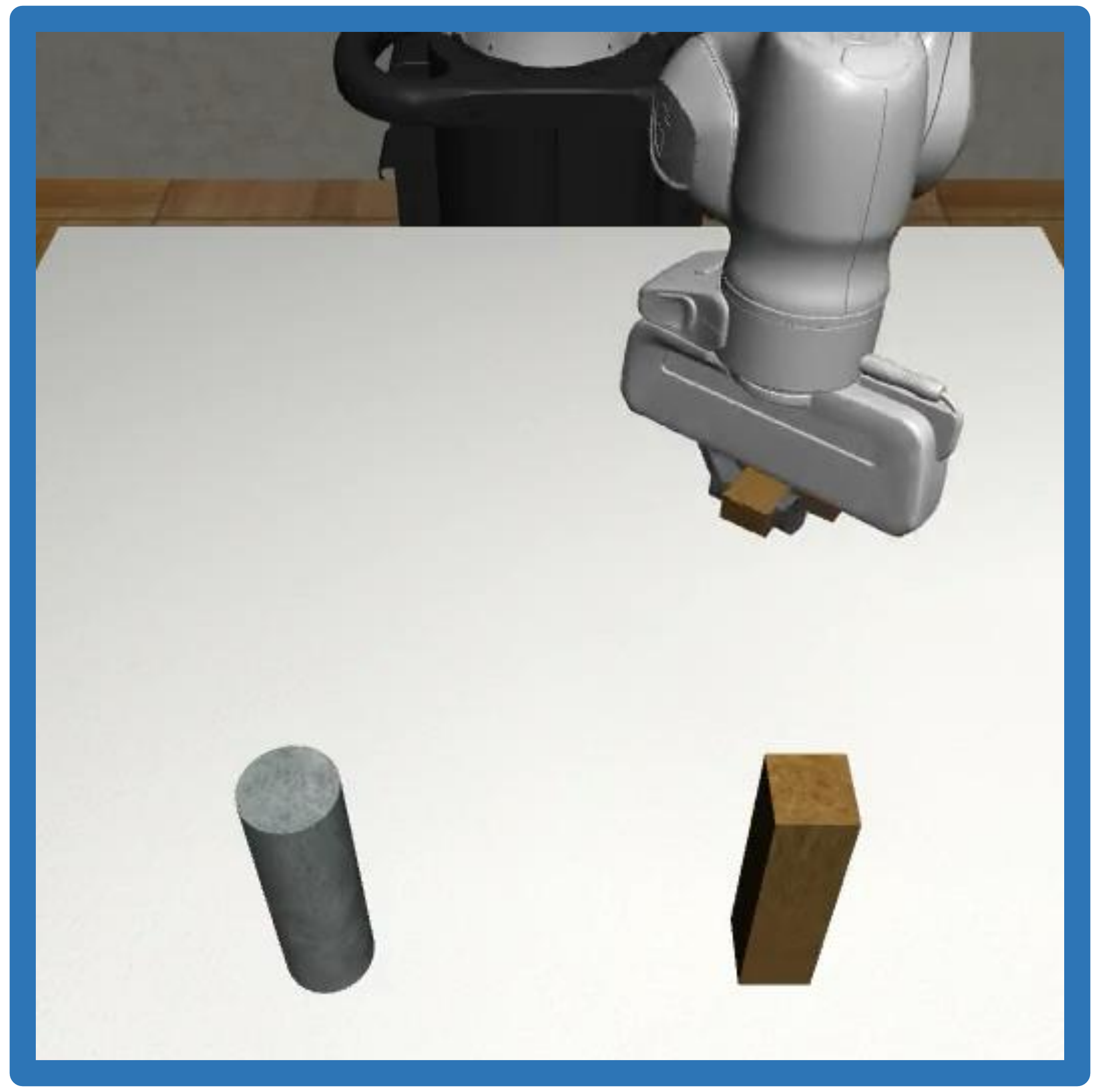
84% agent success rate

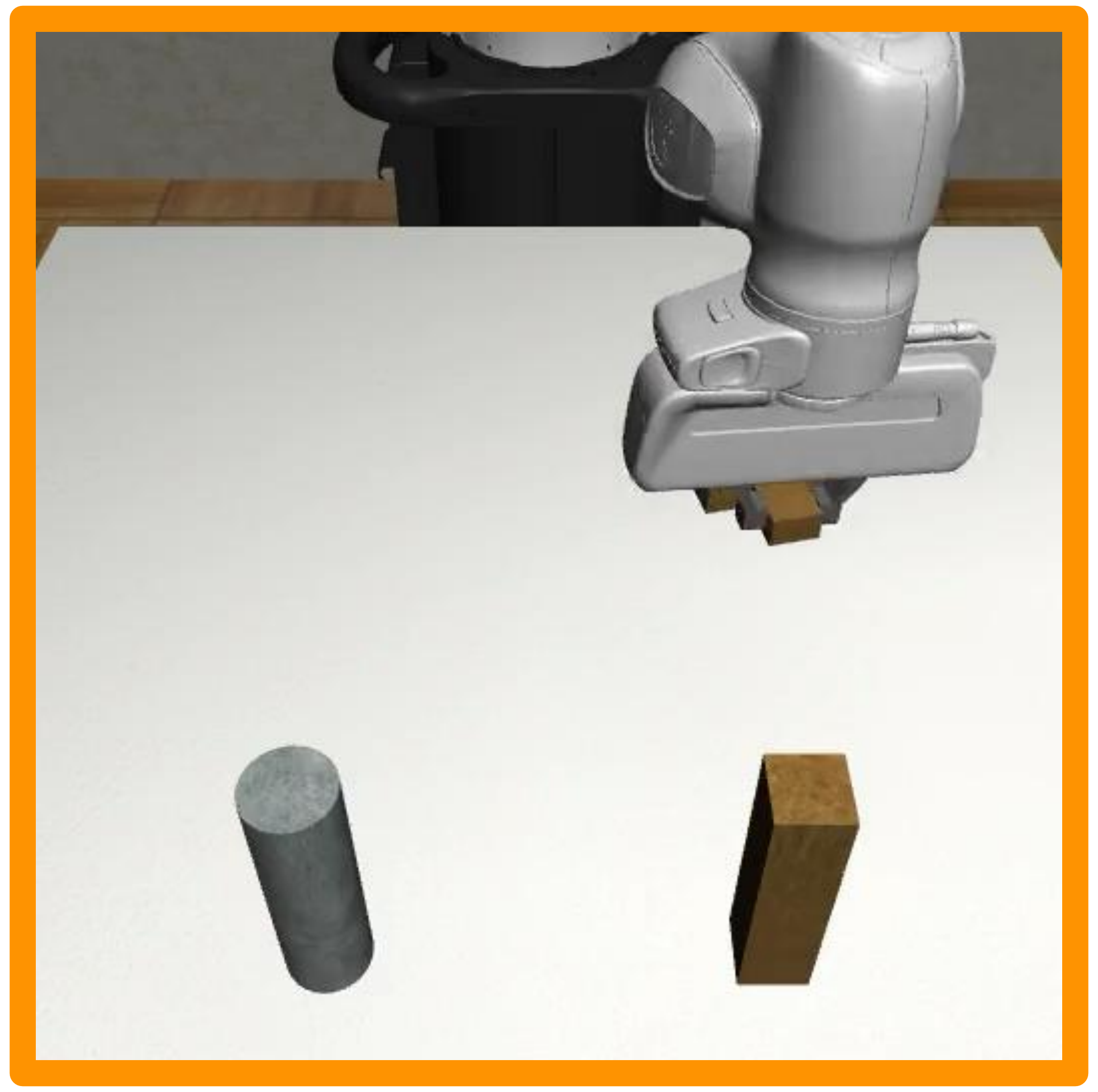
Key Insight: Human datasets can consist of similar, possibly redundant behaviors

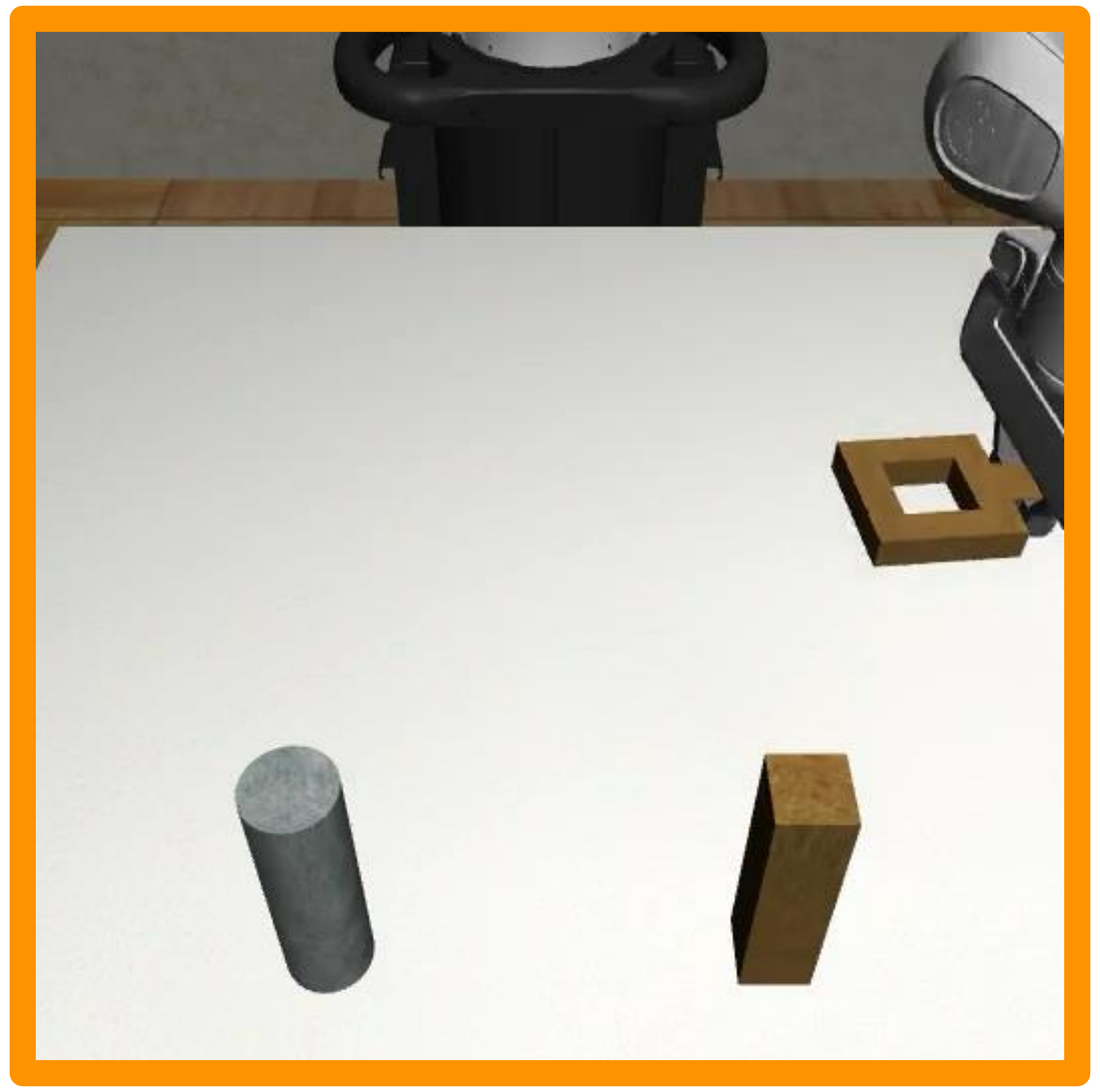


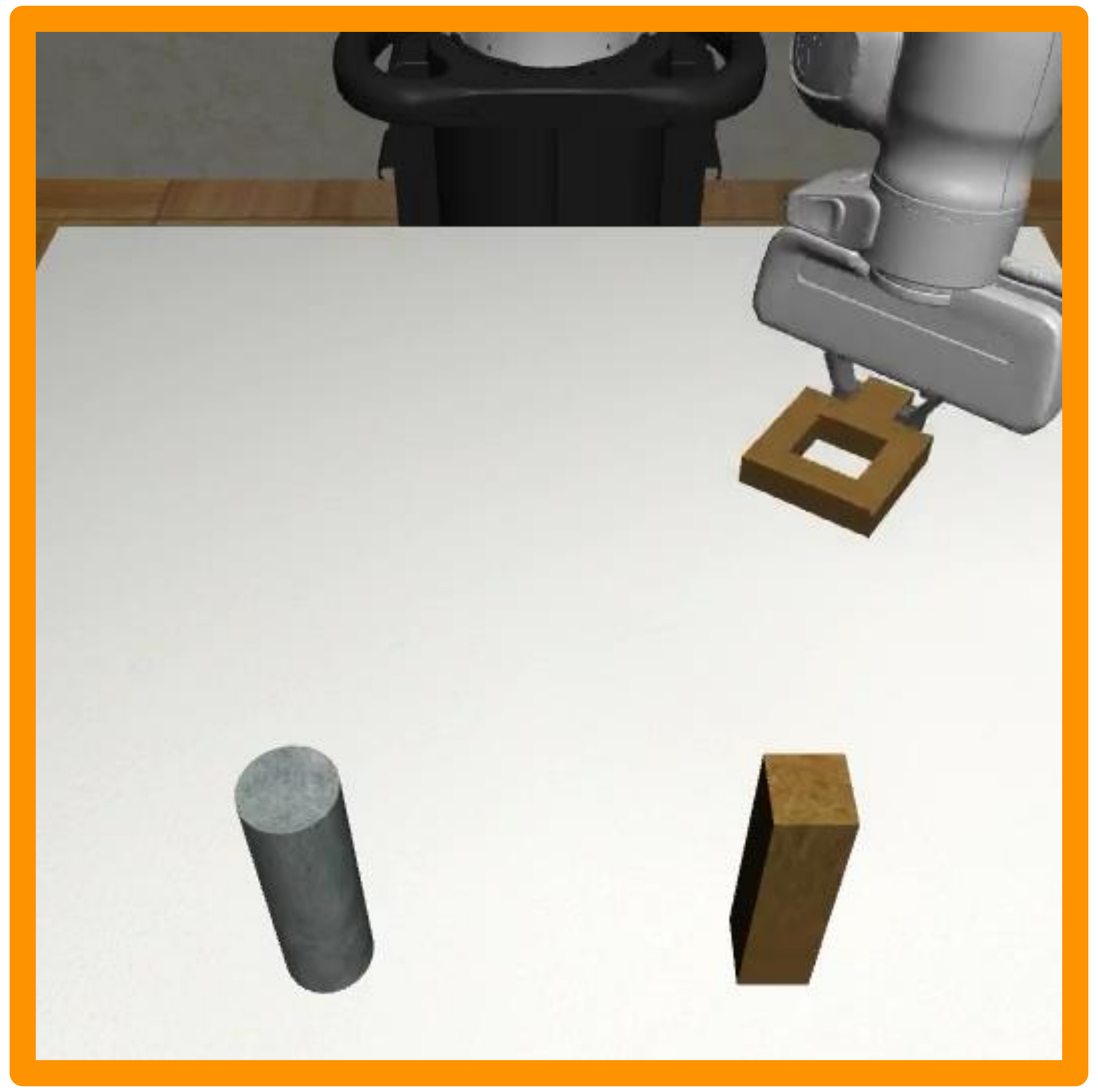
Mandlekar et al. “MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations”, CoRL 2023

Key Idea: Transform existing human data to generate new data with no human effort









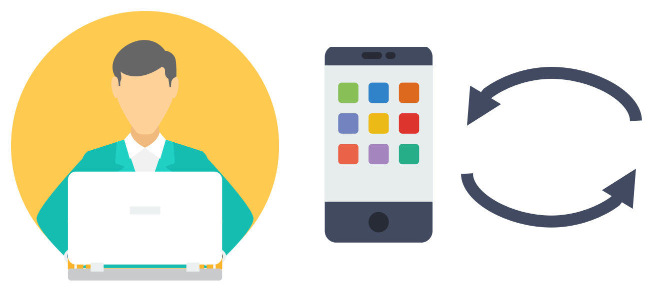
Mandlekar et al. “MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations”, CoRL 2023

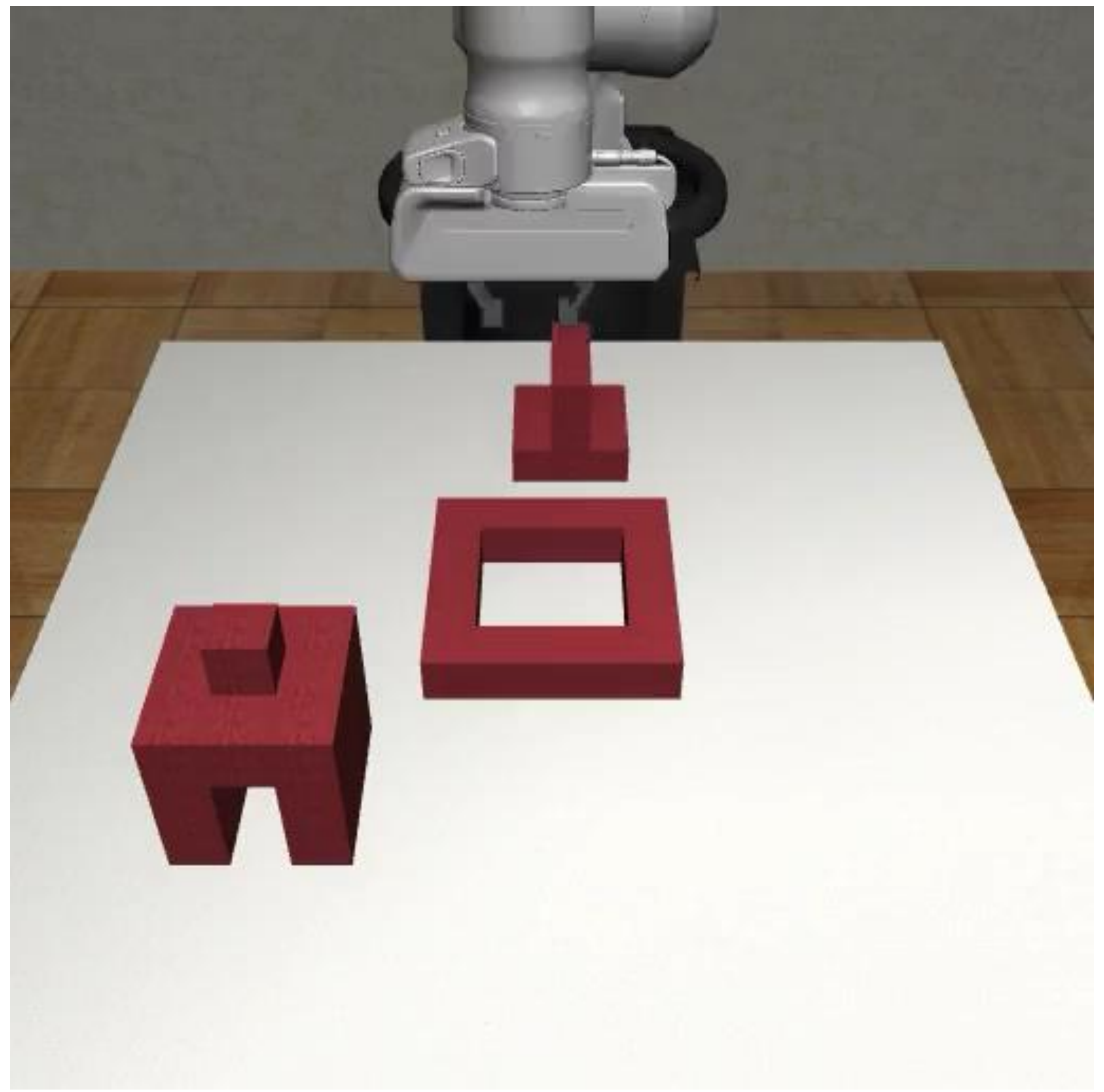
MimicGen

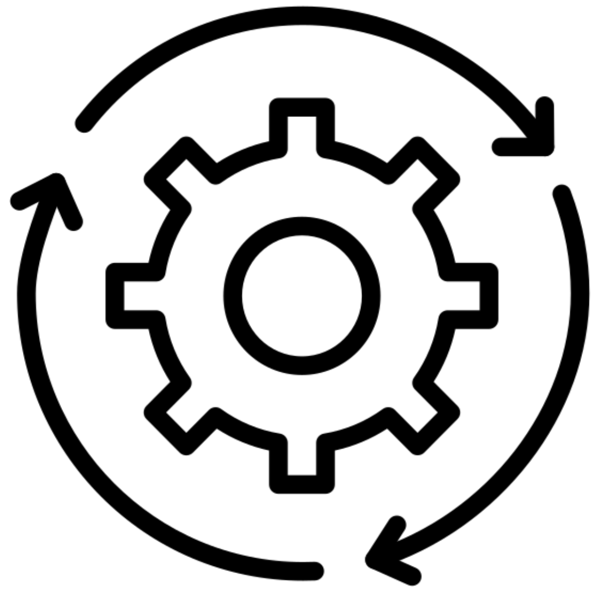
a data generation system for
robot learning from demonstration

MimicGen generates large datasets from few human demos

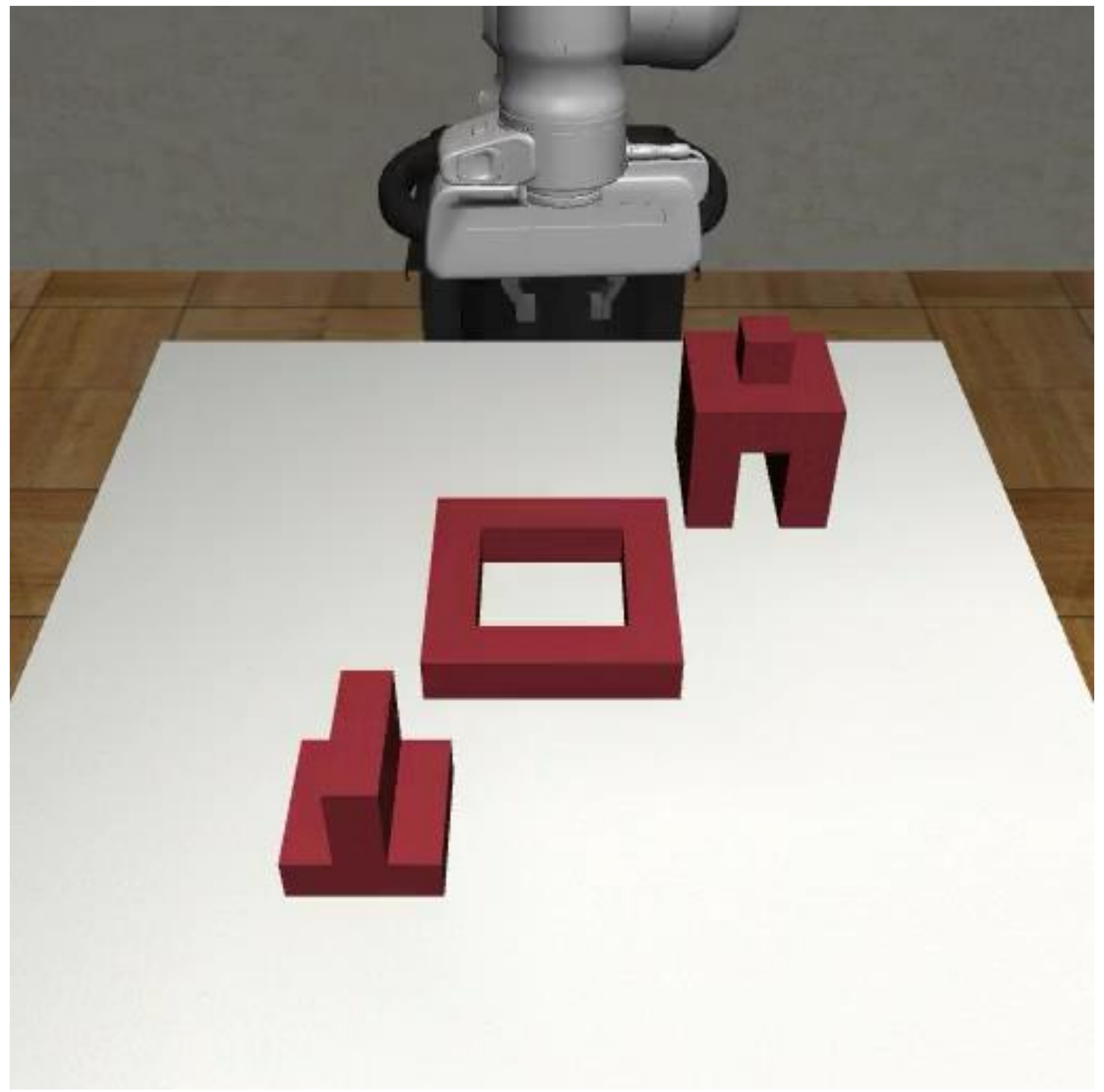
10 human demos







1000 generated demos



Human collects small number of teleoperated demonstrations

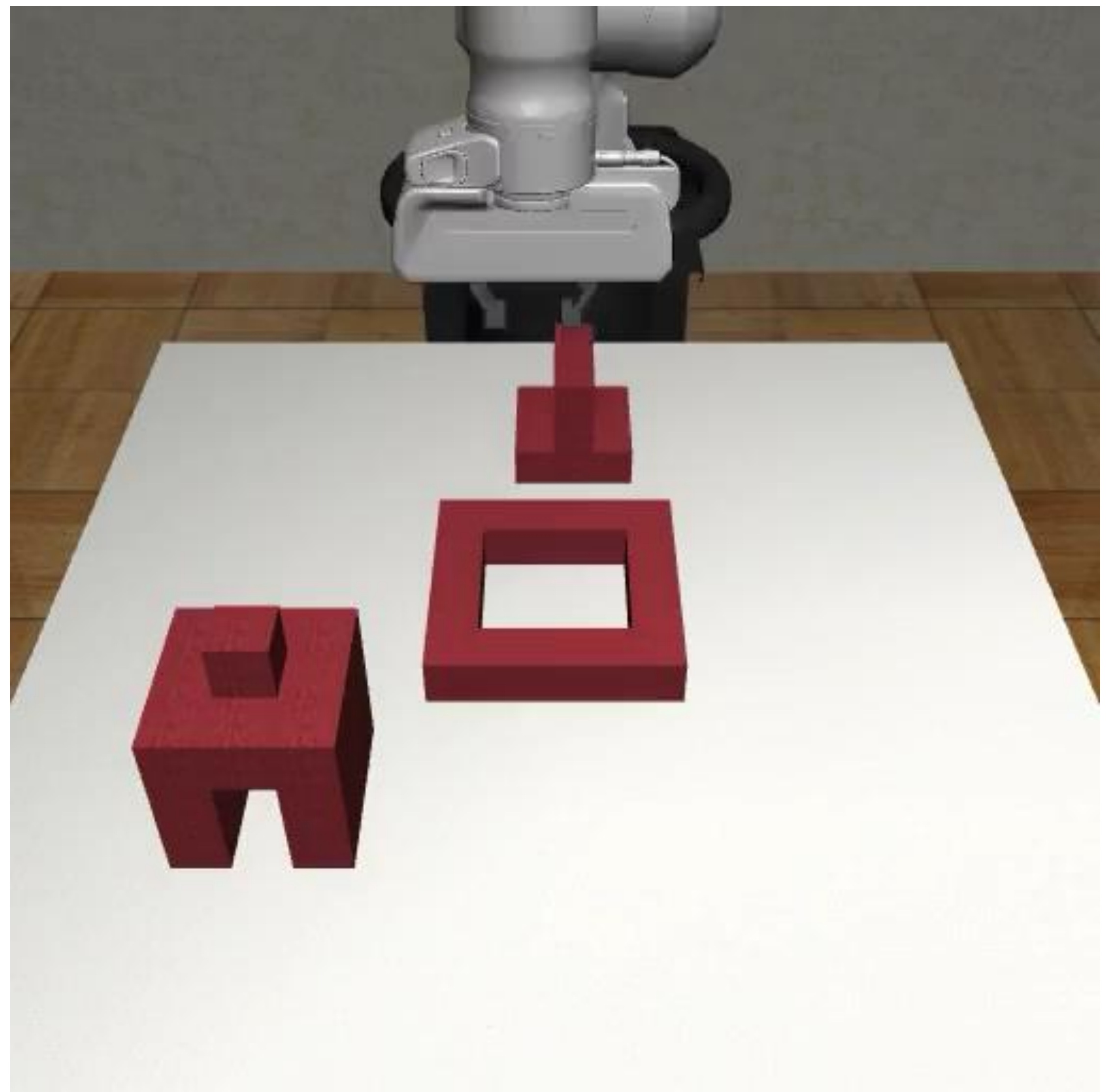
MimicGen generates lots more autonomously!

MimicGen: Large Datasets with Low Effort

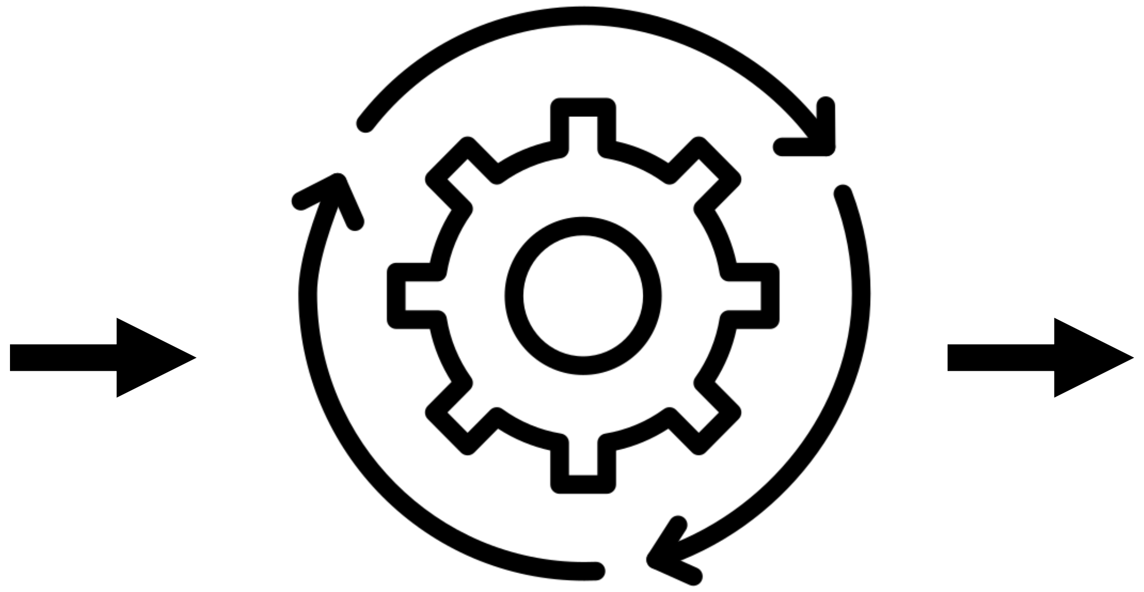
50K generated demos from 200 human demos across
18 tasks, 2 simulators, and real-world

MimicGen: Diverse Datasets from Handful of Human Demos

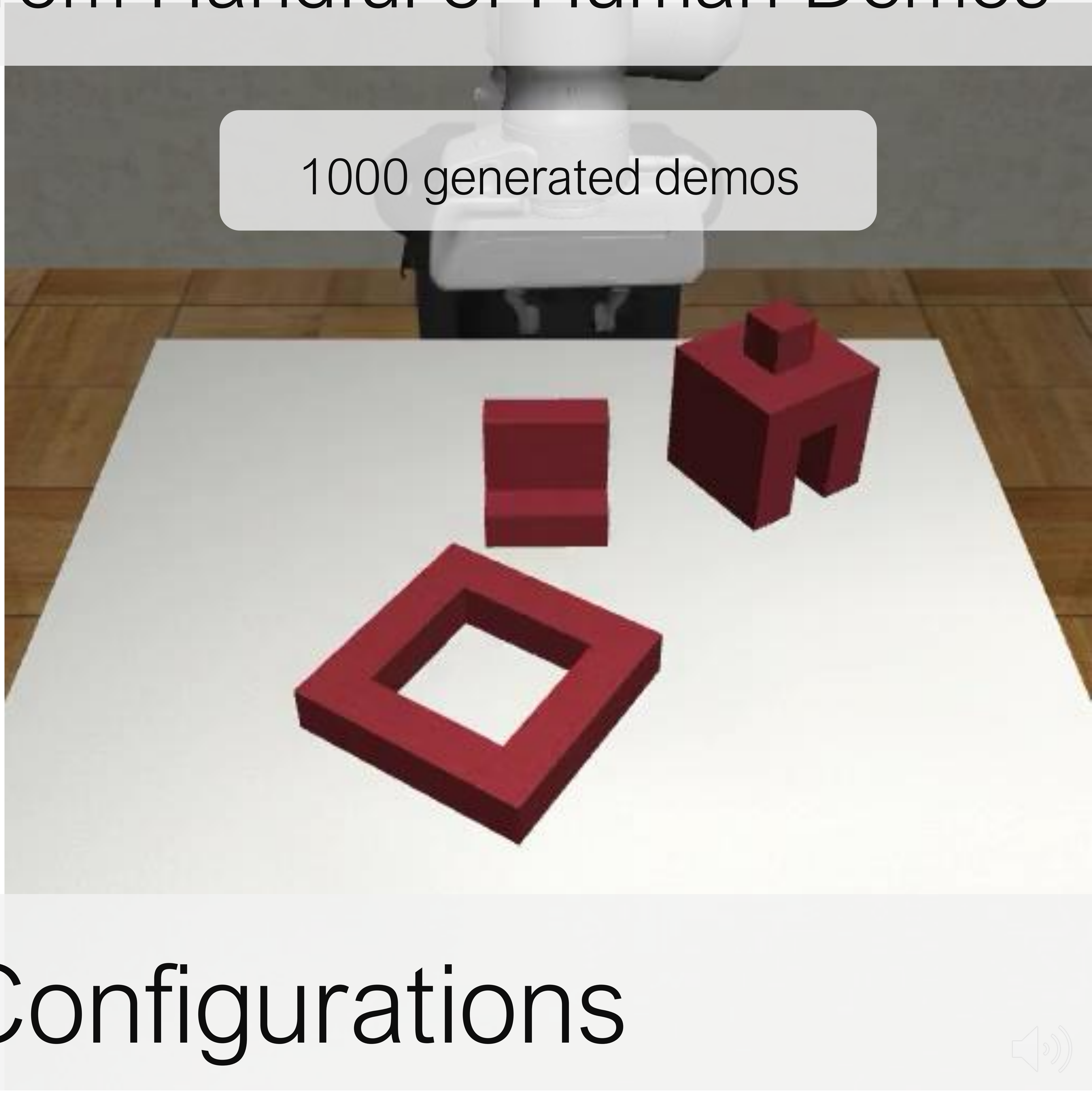
10 human demos



MimicGen



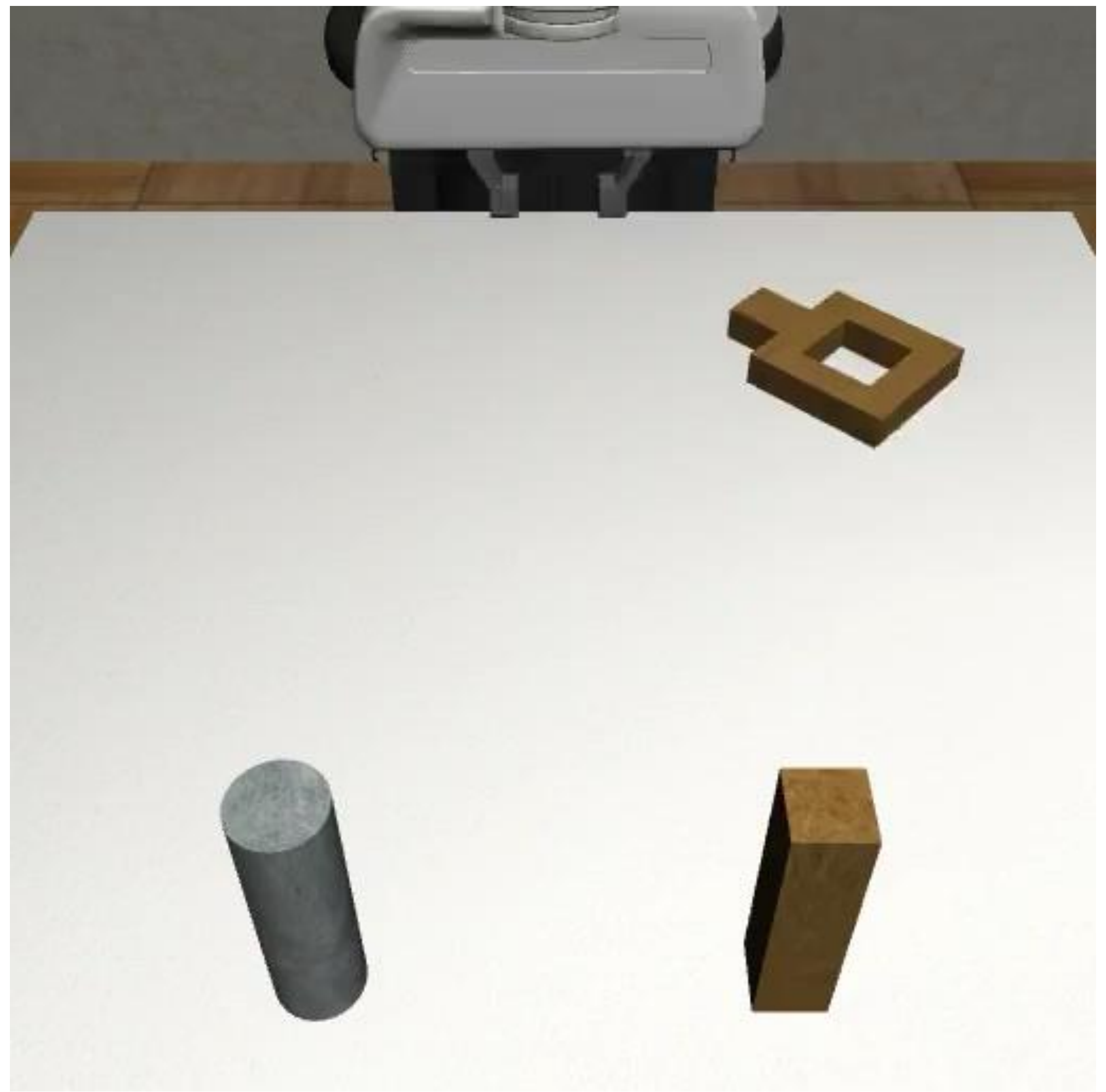
1000 generated demos



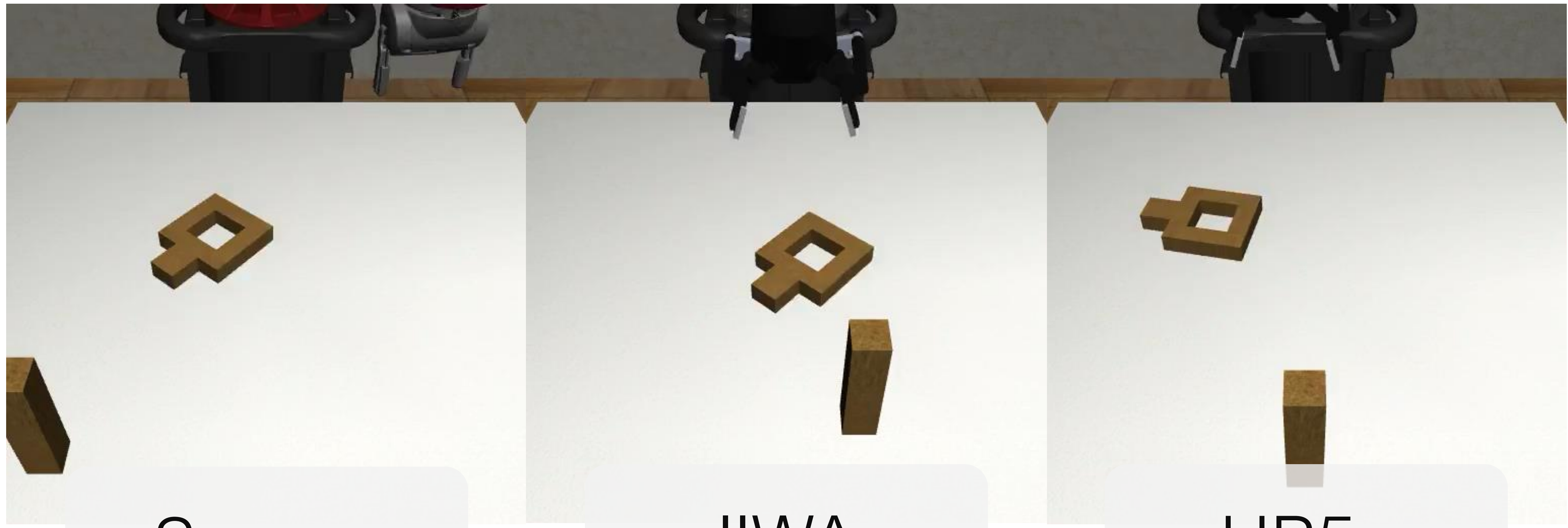
New Object Configurations

MimicGen: Diverse Datasets from Handful of Human Demos

10 human demos
on Panda



1000 demos on multiple robot arms



Sawyer

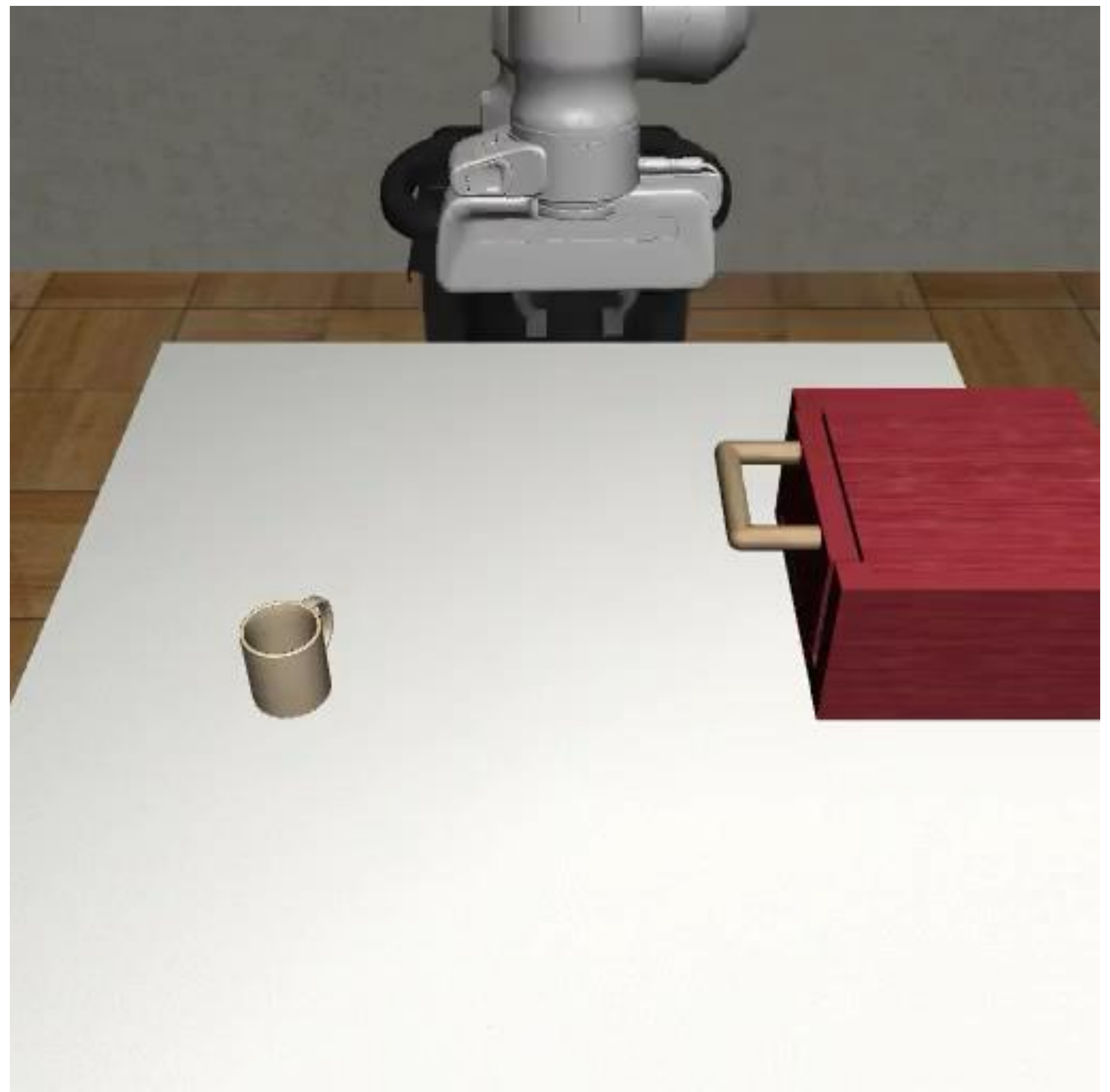
IIWA

UR5e

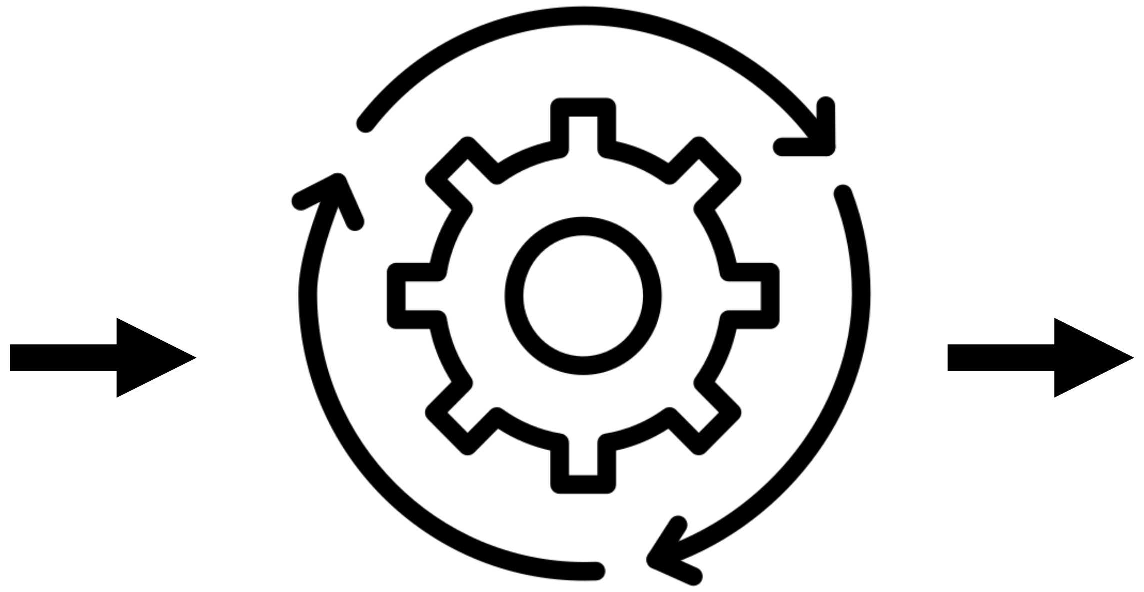
New Robot Arms

MimicGen: Diverse Datasets from Handful of Human Demos

10 human demos
(1 mug)



MimicGen



New Objects

MimicGen is Broadly Applicable to Different Tasks

Contact-Rich (mm-precision) tasks in Isaac Gym Factory

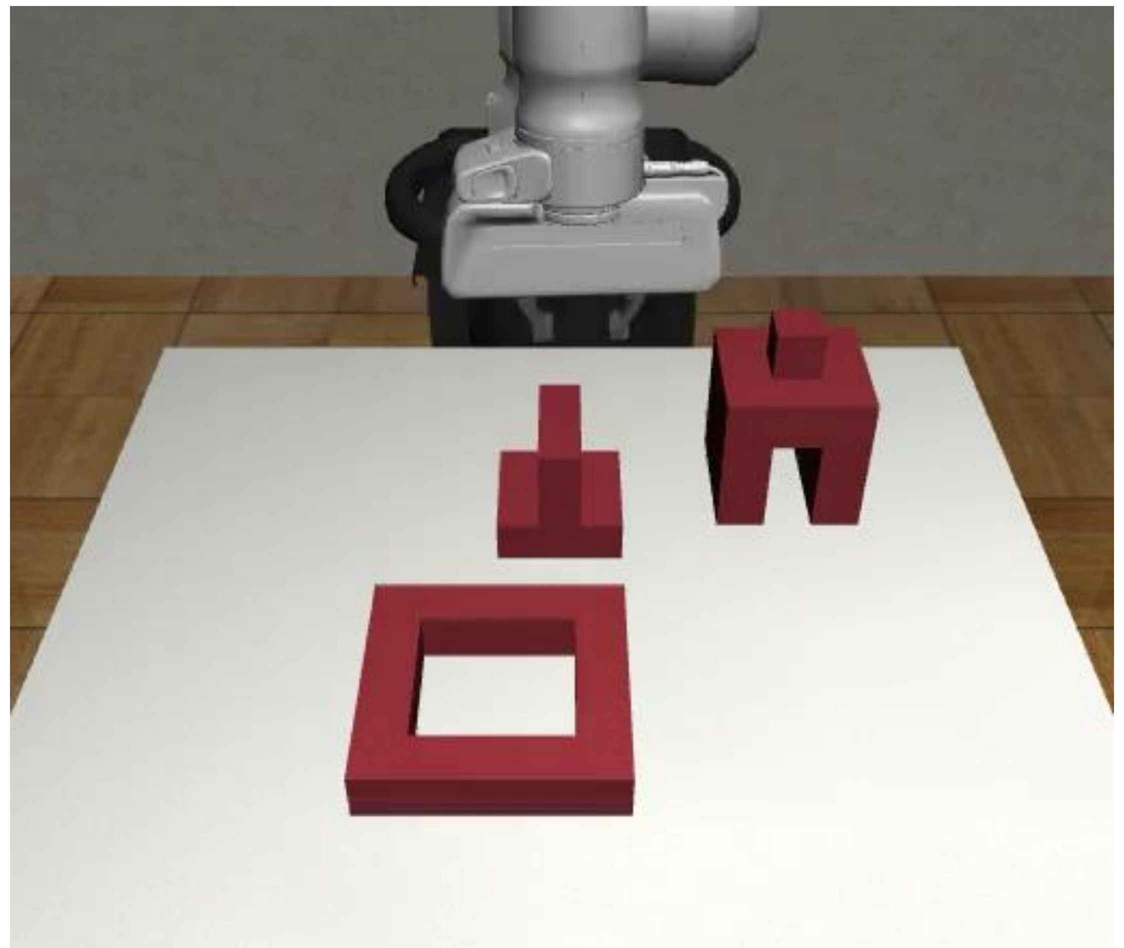
MimicGen is Broadly Applicable to Different Tasks

Contact-Rich (mm-precision) tasks in Isaac Gym Factory

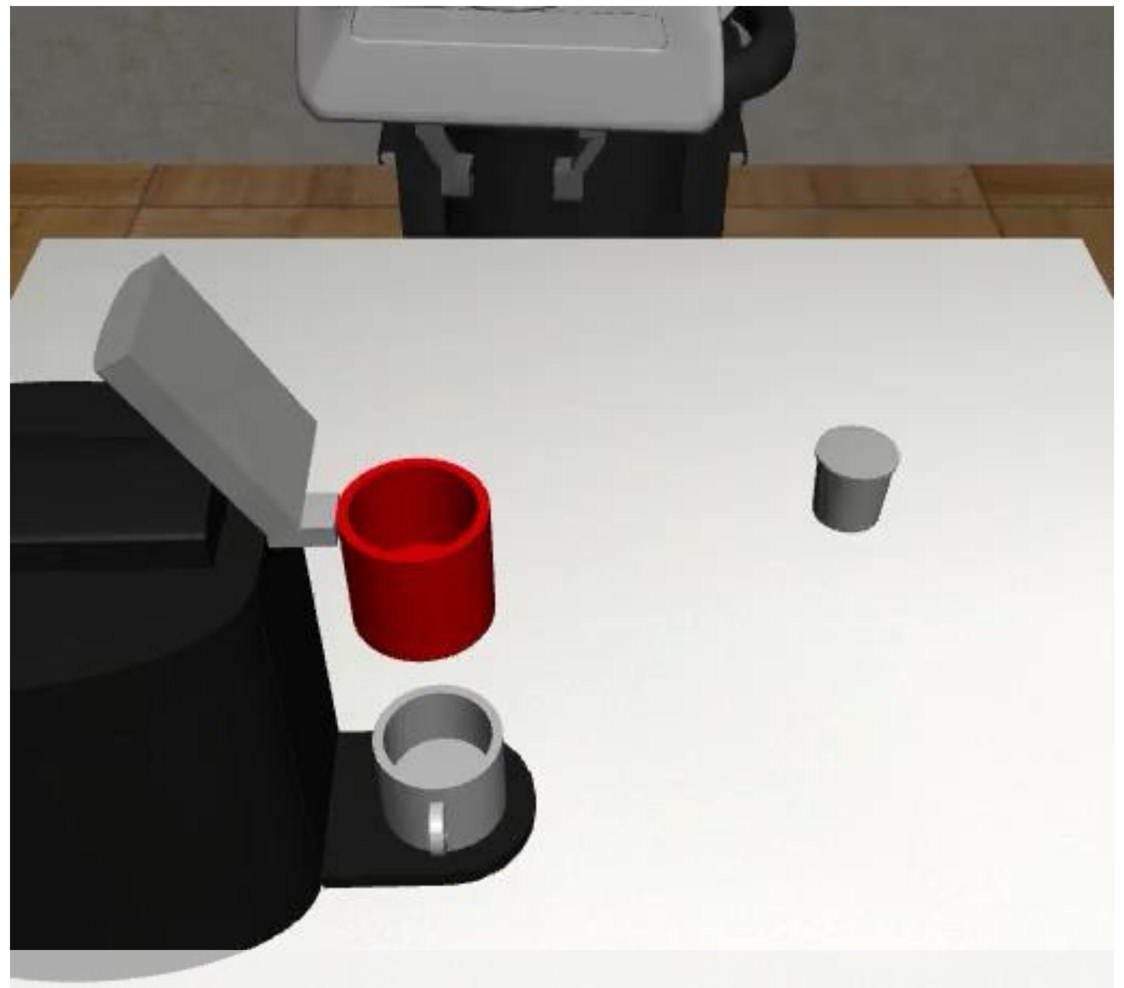
MimicGen is Broadly Applicable to Different Tasks

Long-Horizon tasks

Simple BC on MimicGen data trains performant policies

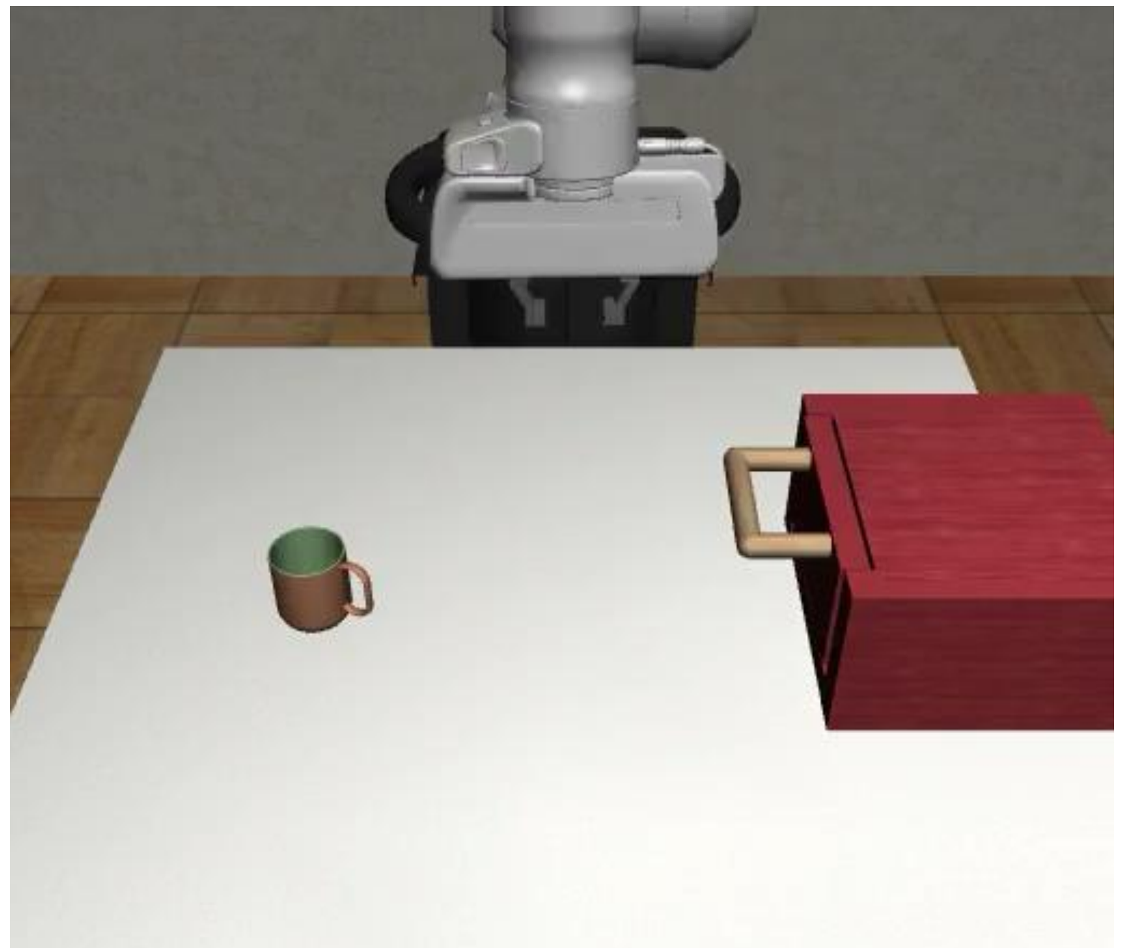


TPA: 61%

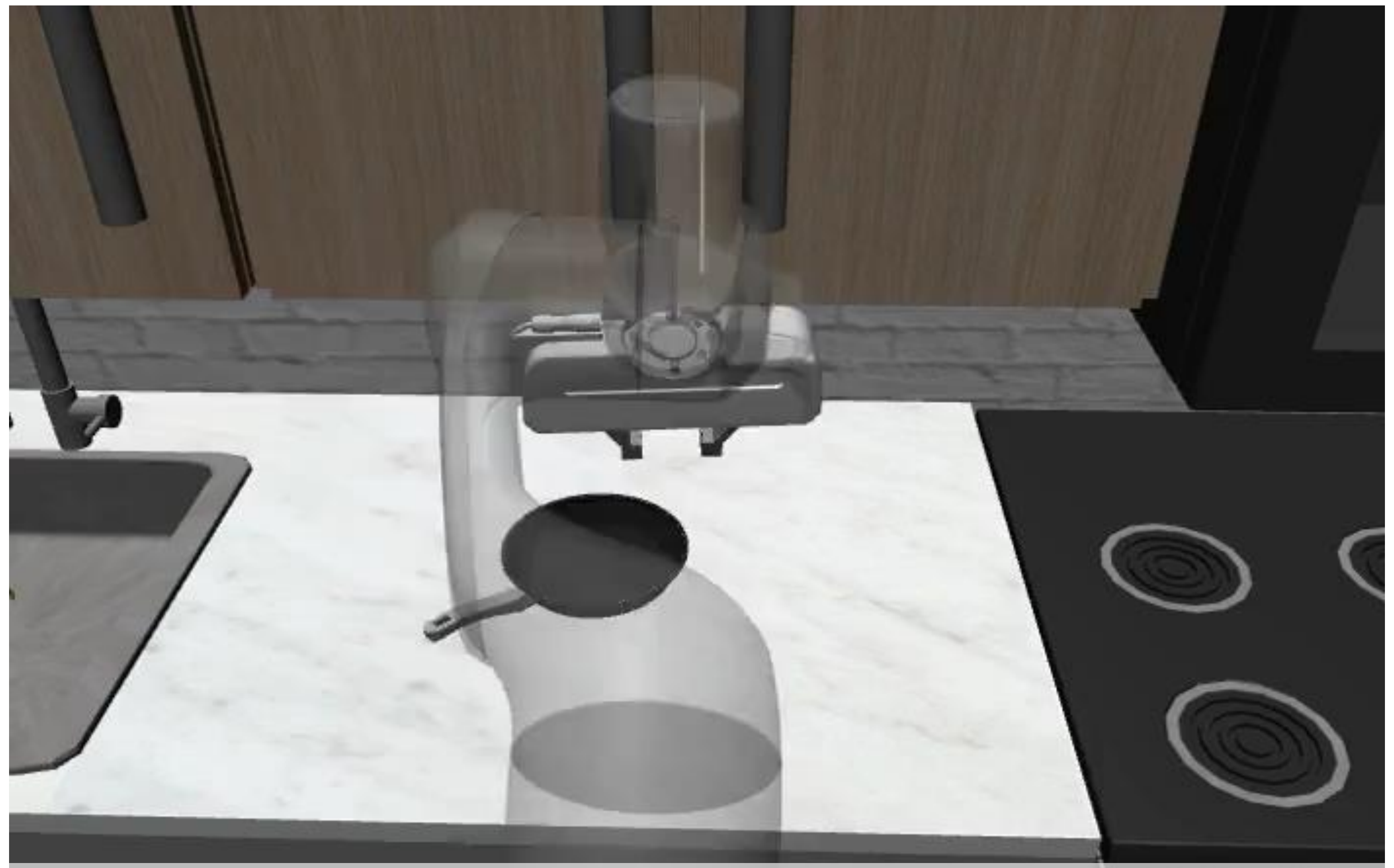


Coffee: 93%

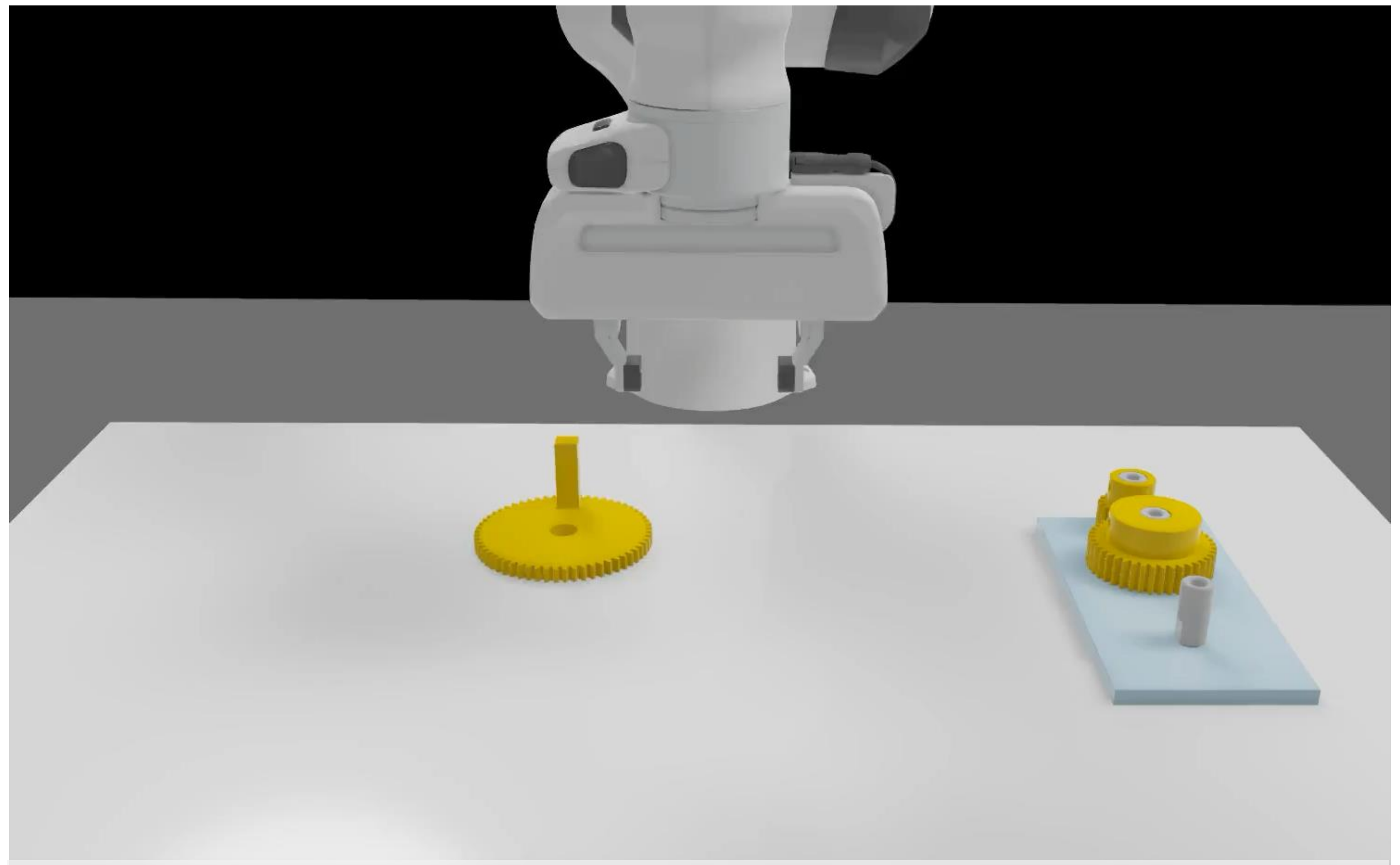
Threading: 72%



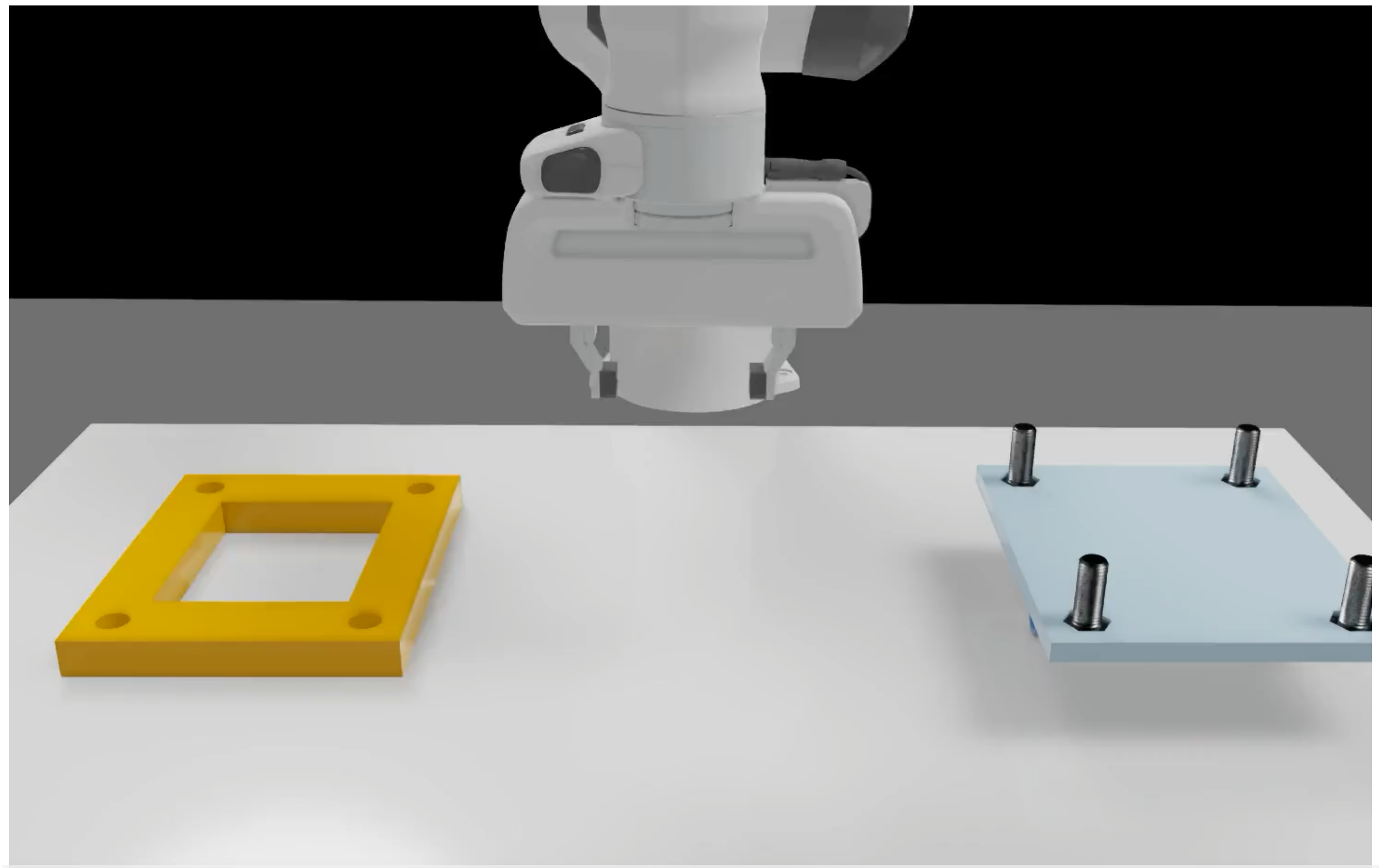
Mug Cleanup: 82%



Mobile Kitchen: 77%

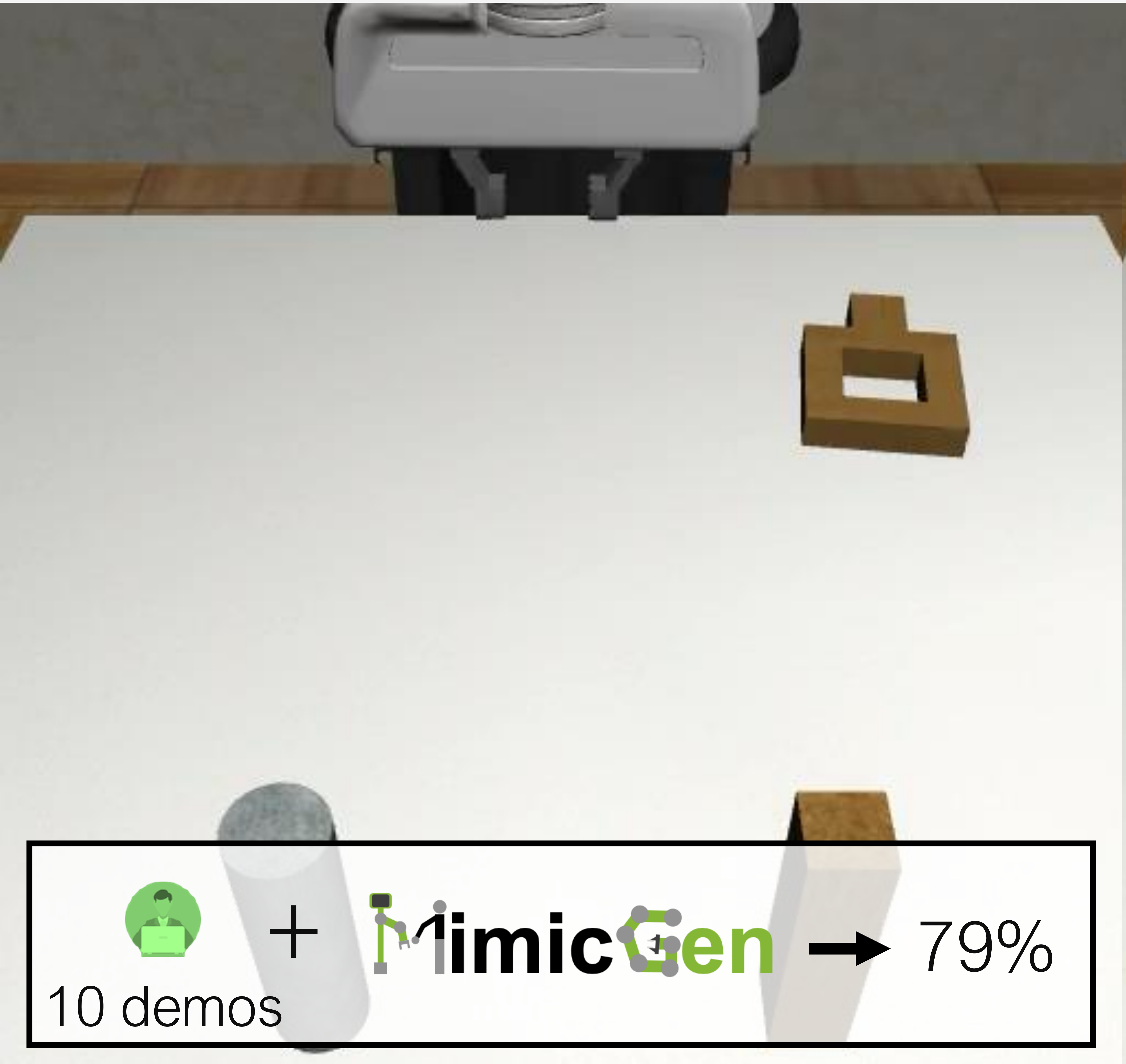


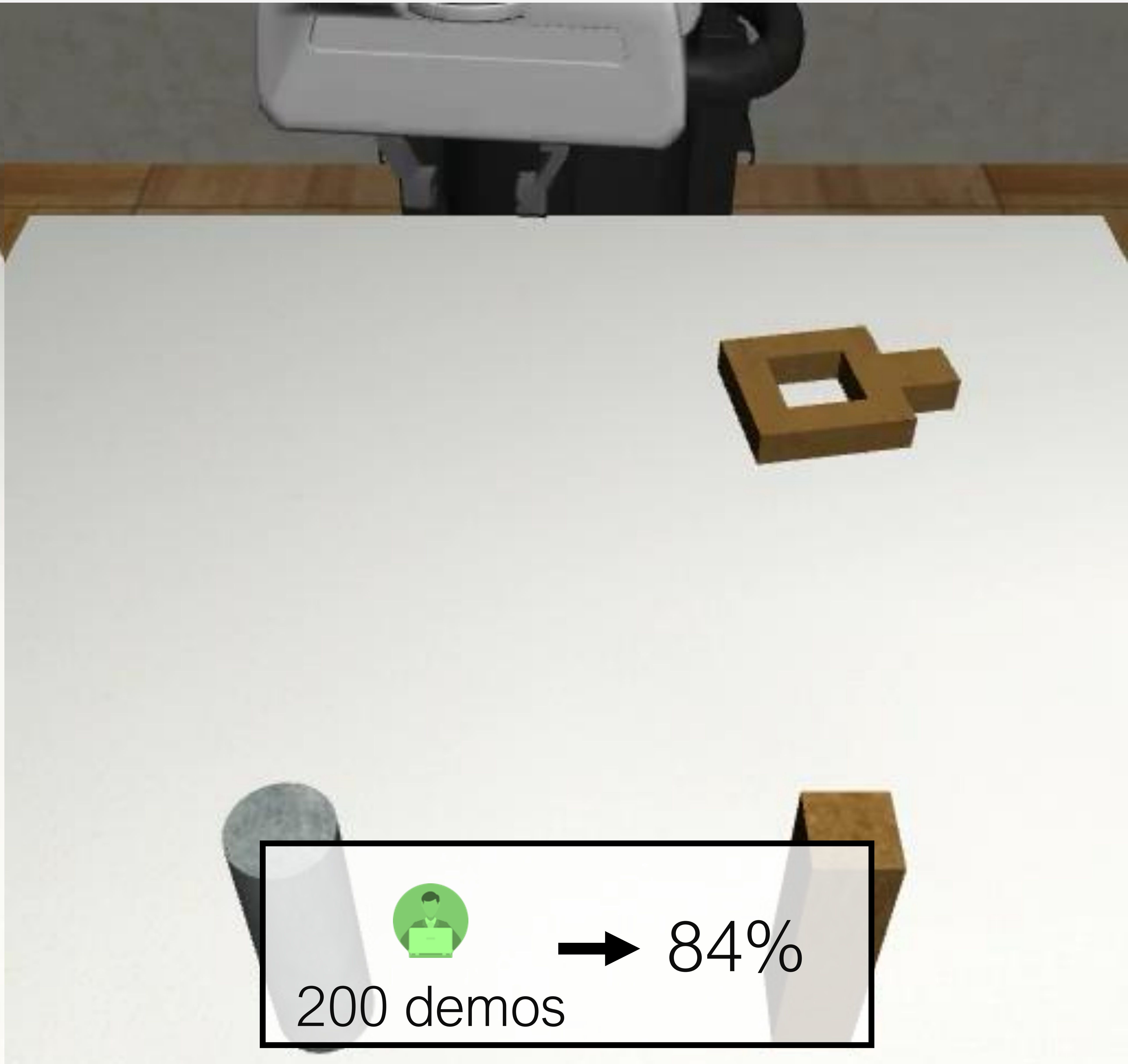
Gear Assembly: 76%



Frame Assembly: 71%

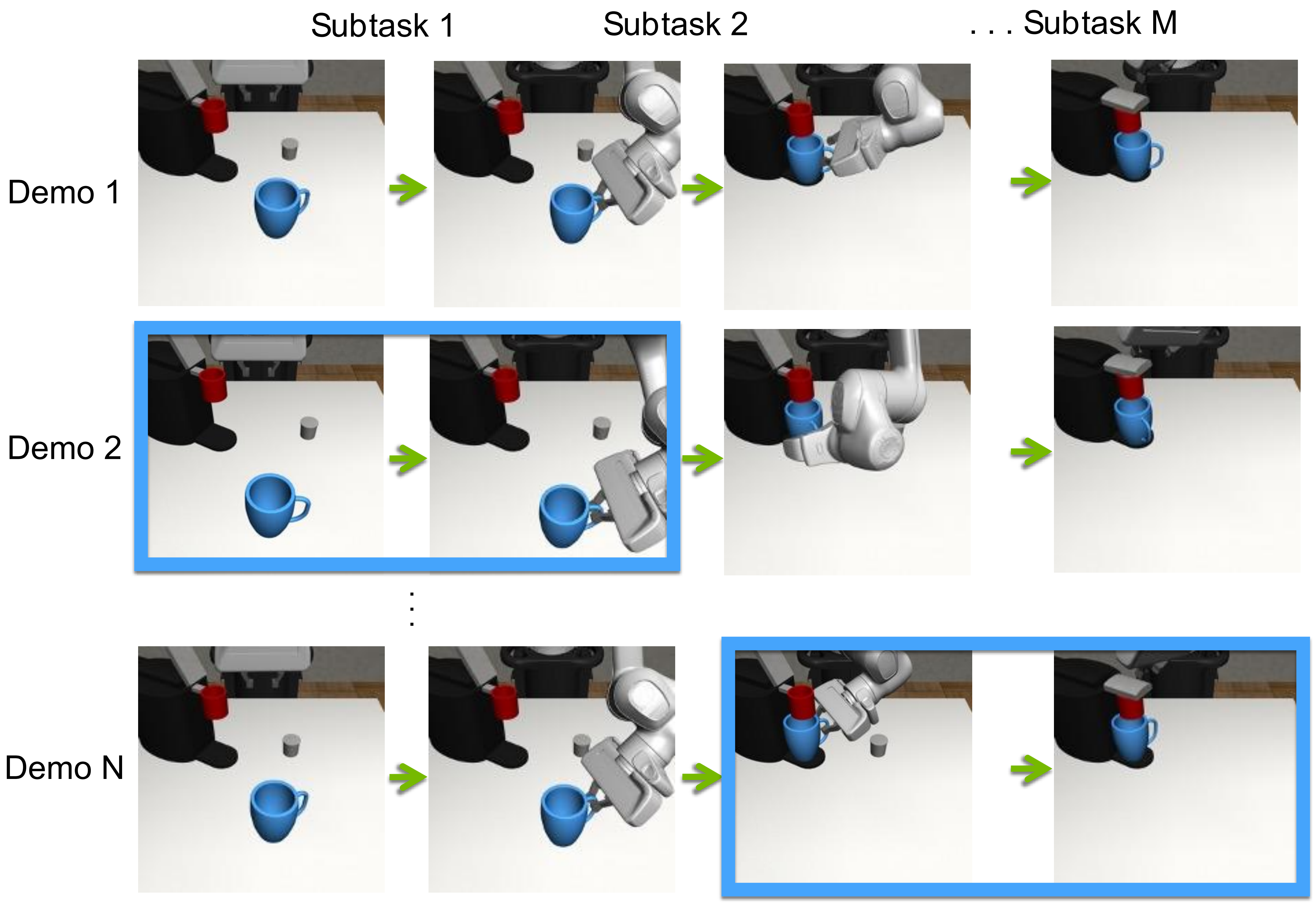
MimicGen produces comparable results to larger human datasets



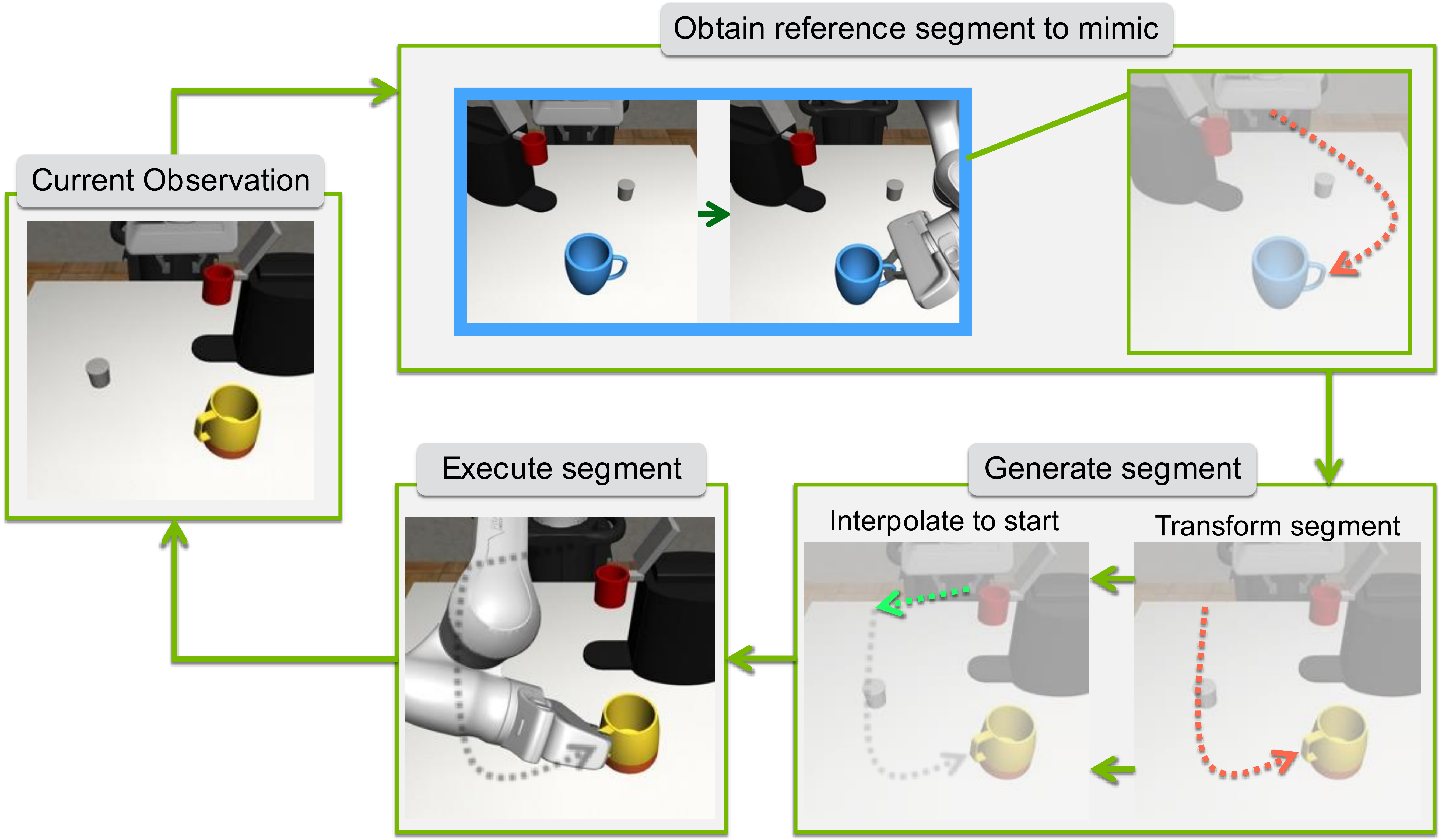


MimicGen Data Generation Overview

Parse source demonstrations into segments



Pipeline for generating new trajectories



Source demos are split into object-centric pieces

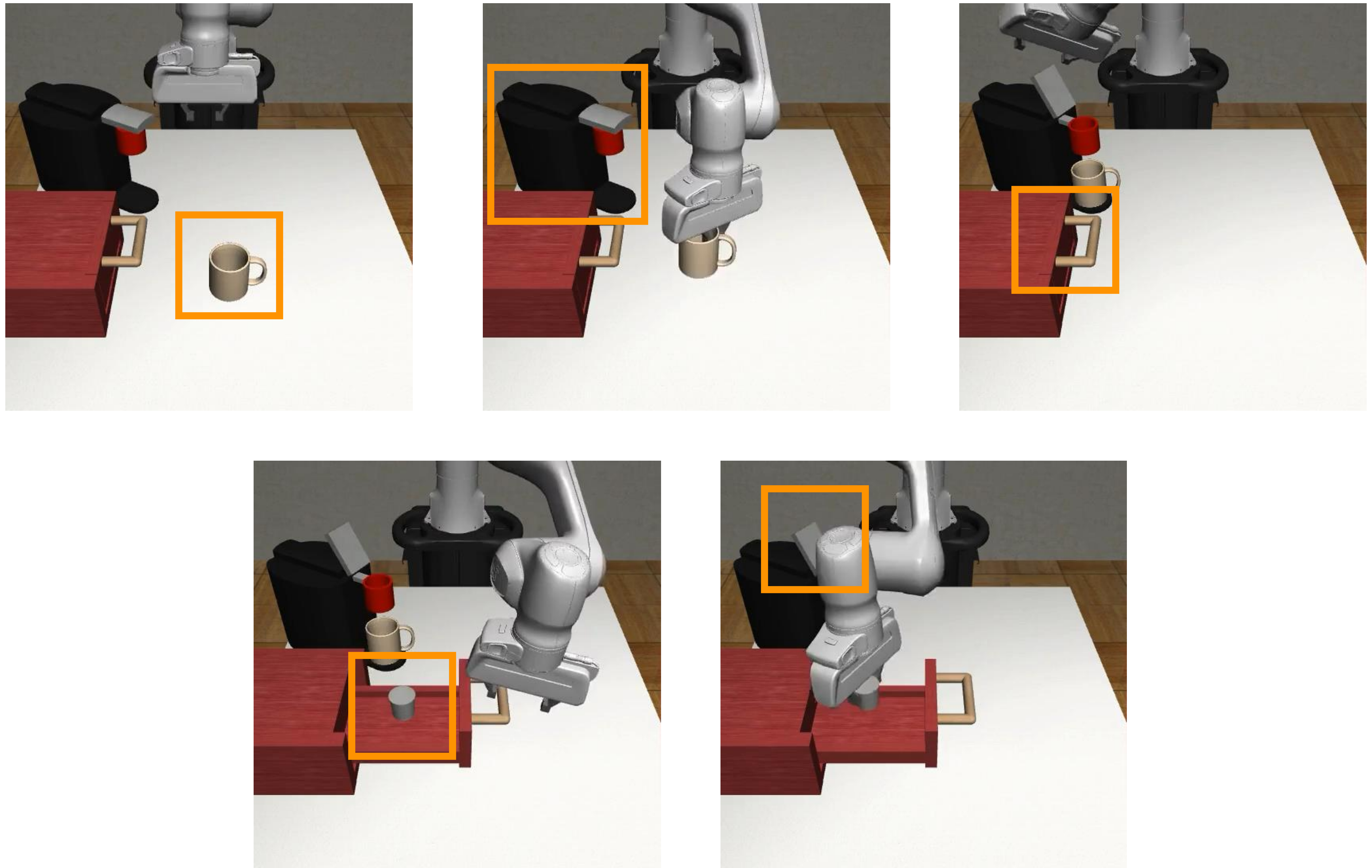
Source demo pieces are transformed and replayed in the new scene one by one

MimicGen: Data Generation Example

Object-Centric Subtask Segments

Source Dataset Trajectory

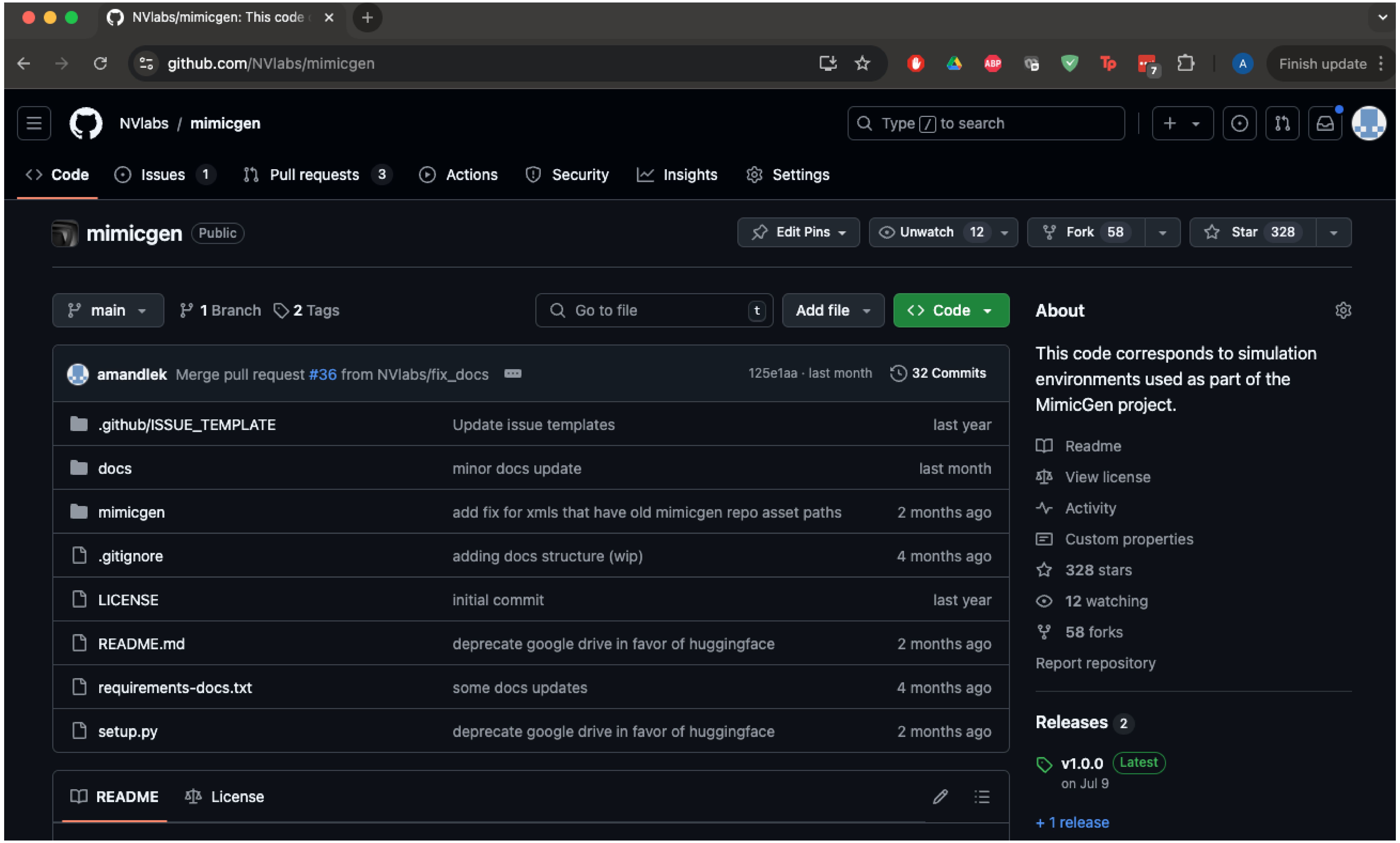
Split using subtask boundaries



Mandlekar et al. “MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations”, CoRL 2023

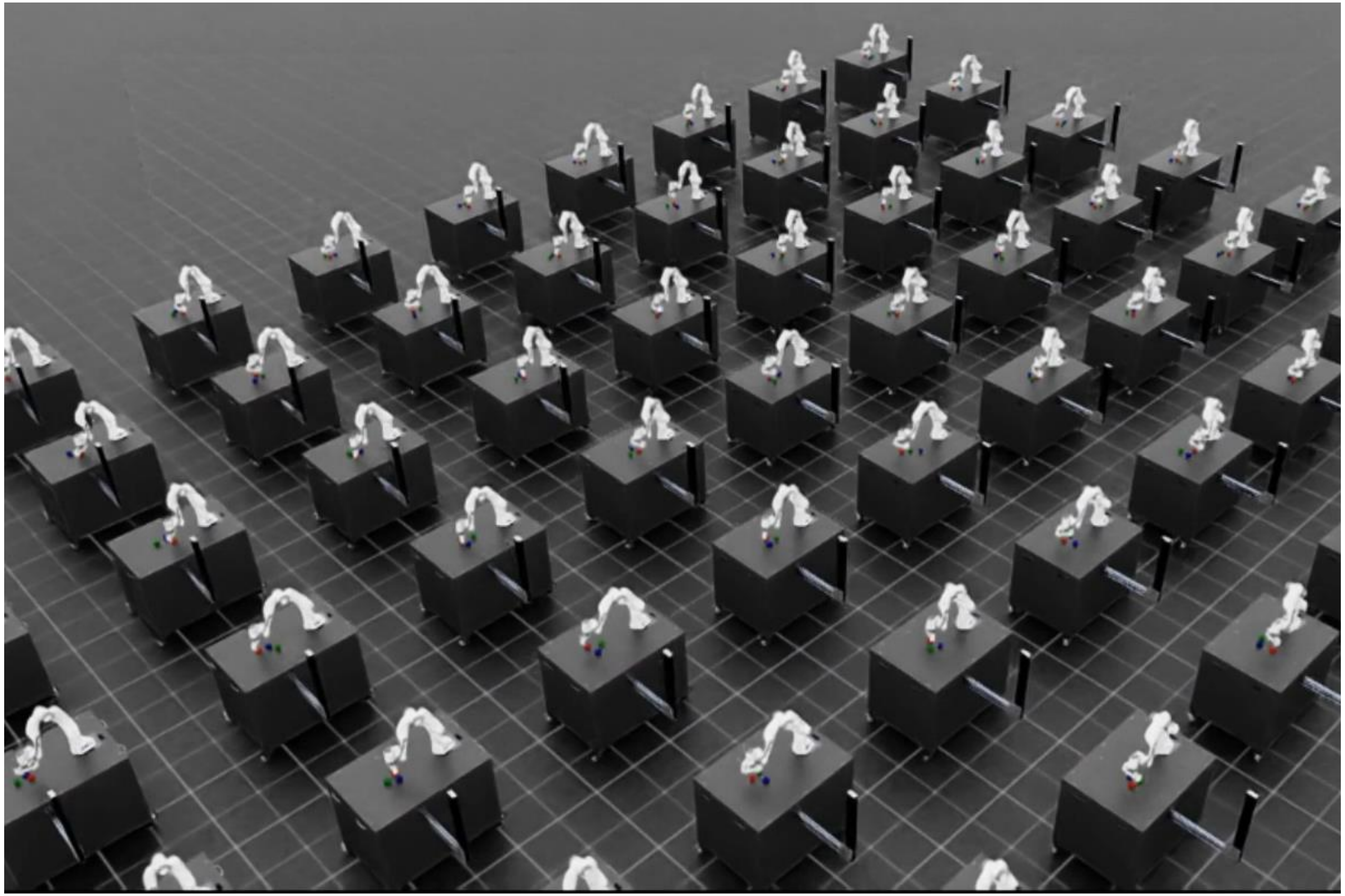
MimicGen: Data Generation Example

Try MimicGen out yourself!



Mandlekar et al. “MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations”, CoRL 2023

MimicGen is in NVIDIA Isaac Lab too!

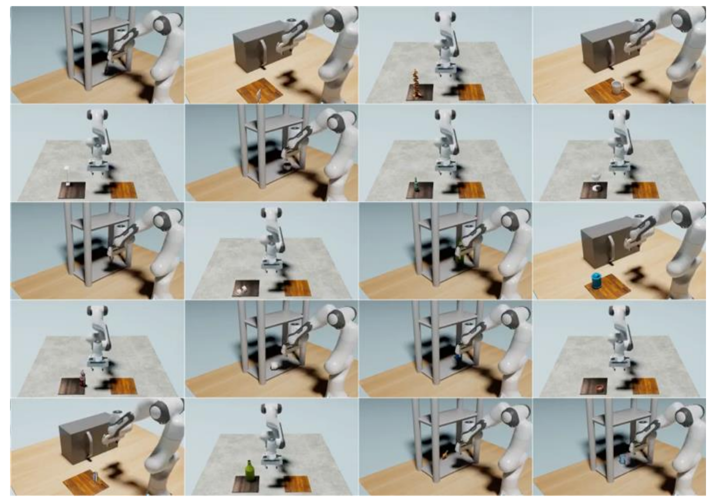


NVIDIA Isaac Lab v2.0

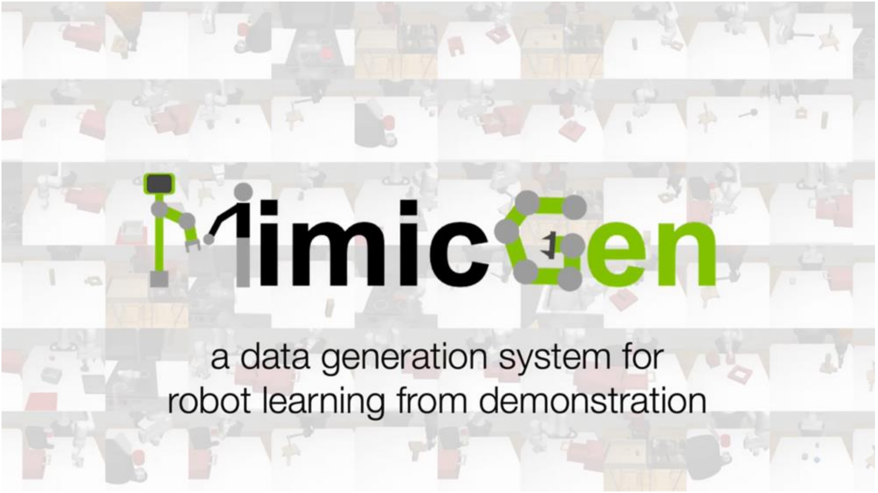
Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations



OPTIMUS (CoRL 2023)



MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)

MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen (ICRA 2025)

SkillMimicGen (CoRL 2024)

AHA (ICLR 2025)

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations

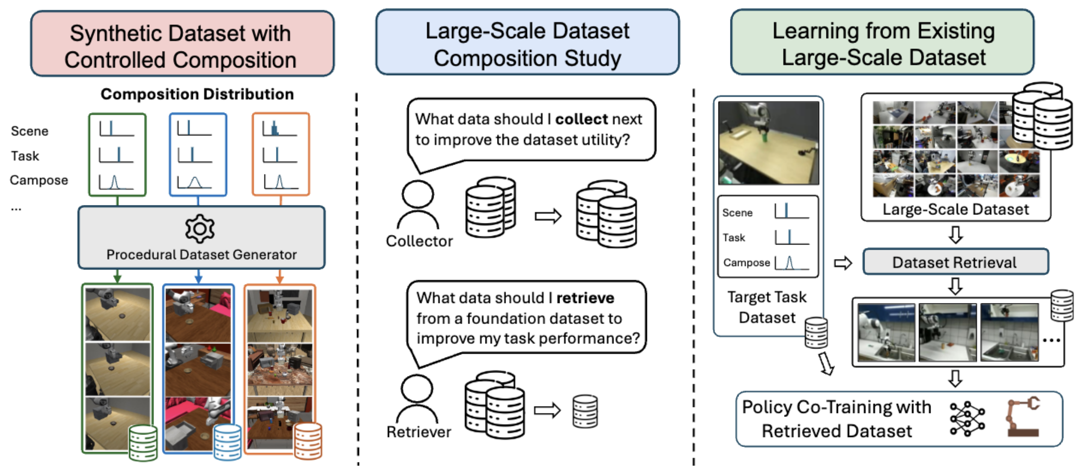
OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)



MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen
(ICRA 2025)

SkillMimicGen
(CoRL 2024)

AHA
(ICLR 2025)

Introducing RoboCasa

Interactable Furniture and Appliances

Augmenting scene diversity with text-to-image models

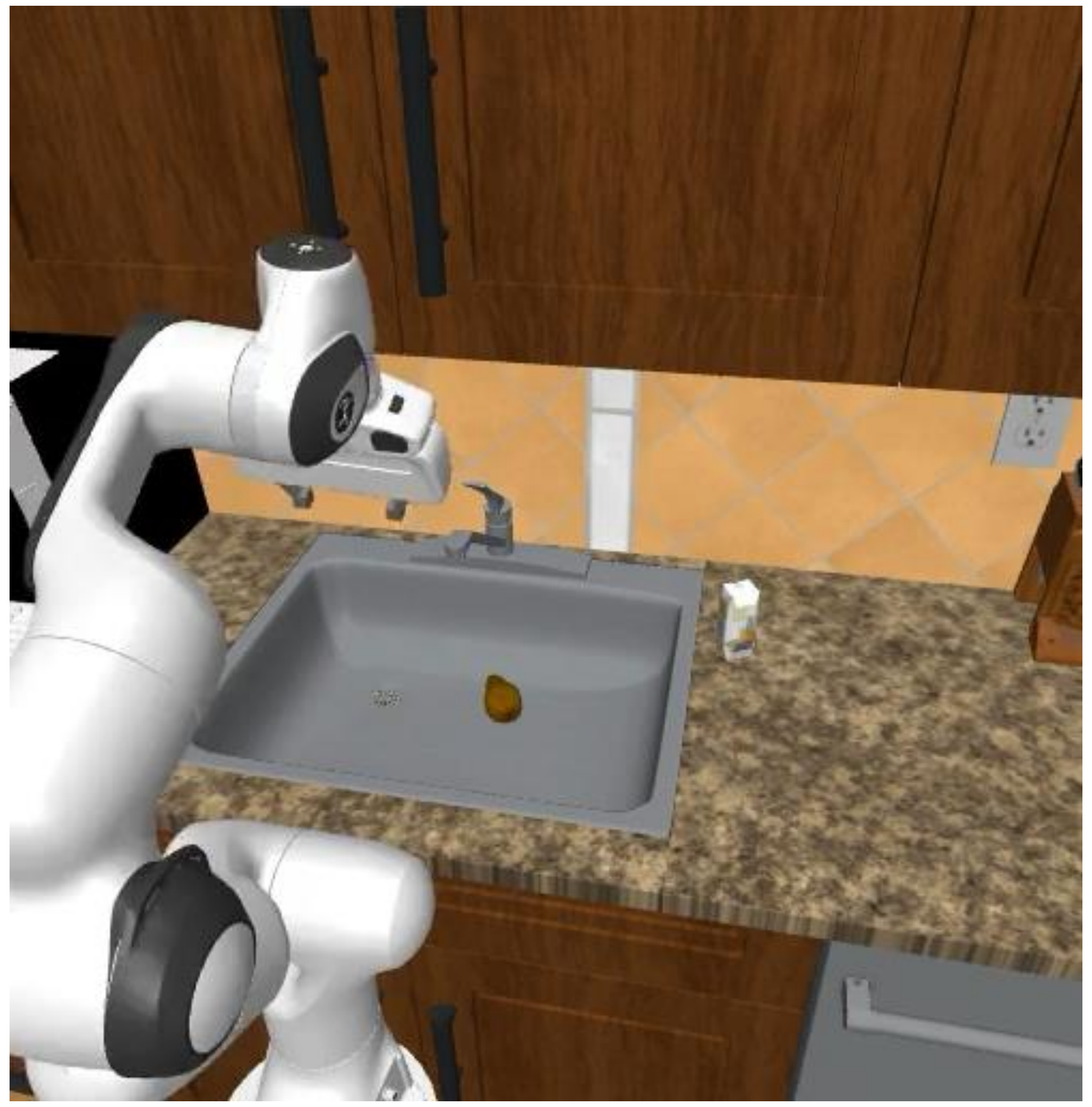
Creating diverse object assets with text-to-3D models

Combining diverse simulation with scalable synthetic data generation

50 human demos

MimicGen

3000 generated demos



Human collects small number of teleoperated demonstrations

MimicGen generates lots more autonomously!

Nasiriany et al. "RoboCasa: Large-Scale Simulation of Everyday Tasks for Generalist Robots", RSS 2024

Training foundational robot skills

Training foundational robot skills

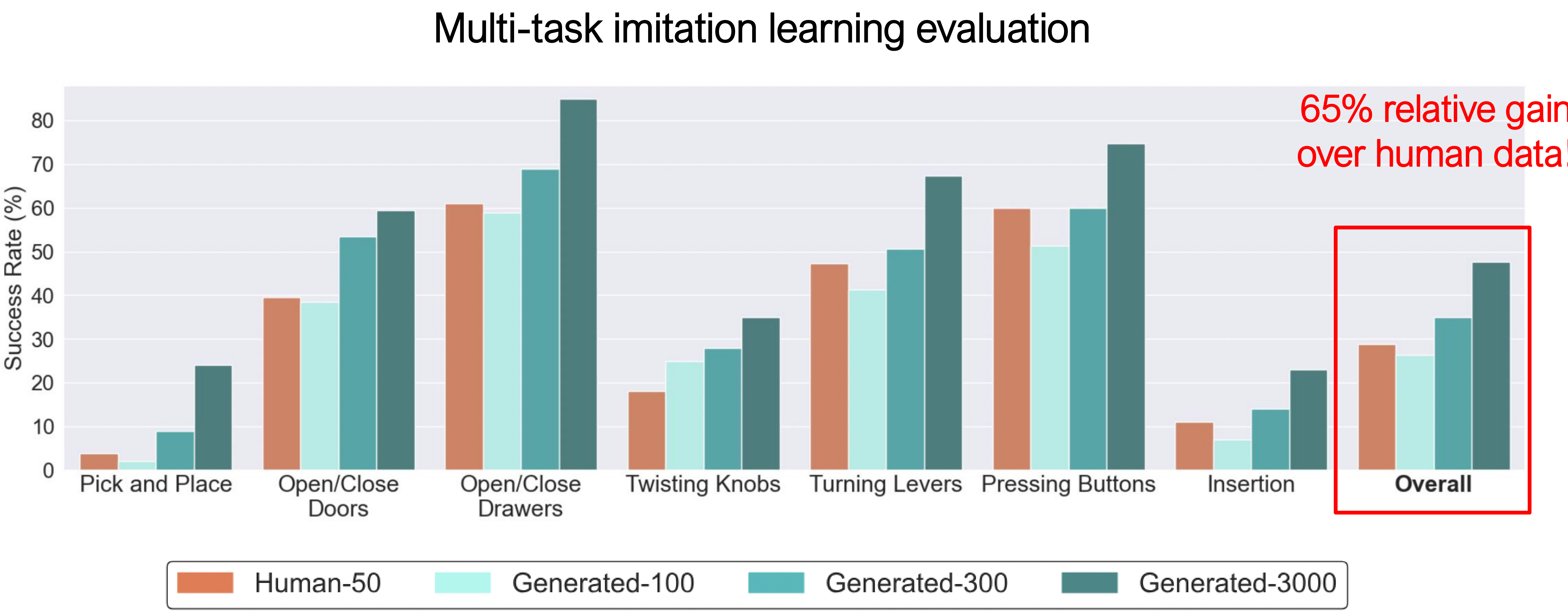
Opening and closing doors

Training foundational robot skills

Training foundational robot skills

Training foundational robot skills

Multi-Task Policy Performance Scales with Synthetic Data



Nasiriany et al. “RoboCasa: Large-Scale Simulation of Everyday Tasks for Generalist Robots”, RSS 2024

Synthetic MimicGen data can aid in transfer to real world tasks

Training on real-world tasks with 50 demonstrations

Real-world only training: 13.6%

Co-training with MimicGen data: 24.4%

Nasiriany et al. “RoboCasa: Large-Scale Simulation of Everyday Tasks for Generalist Robots”, RSS 2024

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations

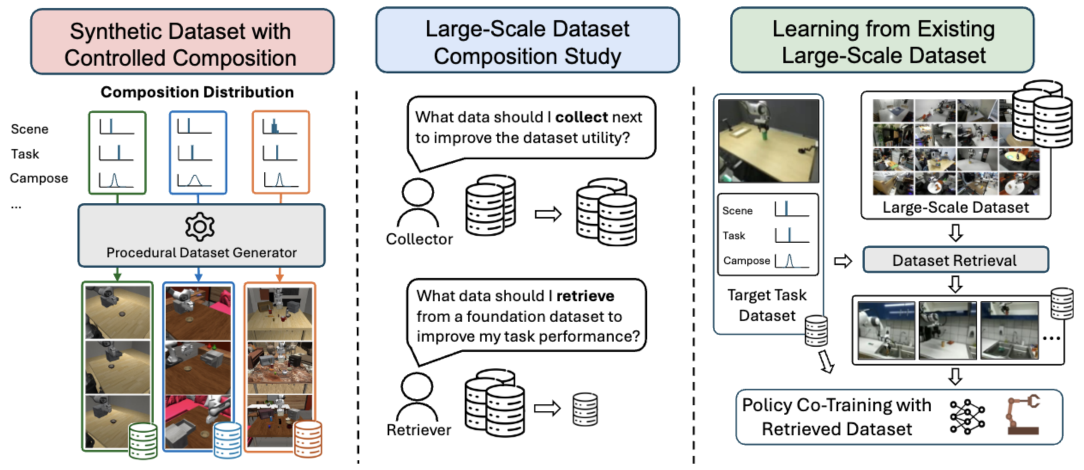
OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)



MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen (ICRA 2025)

SkillMimicGen (CoRL 2024)

AHA (ICLR 2025)

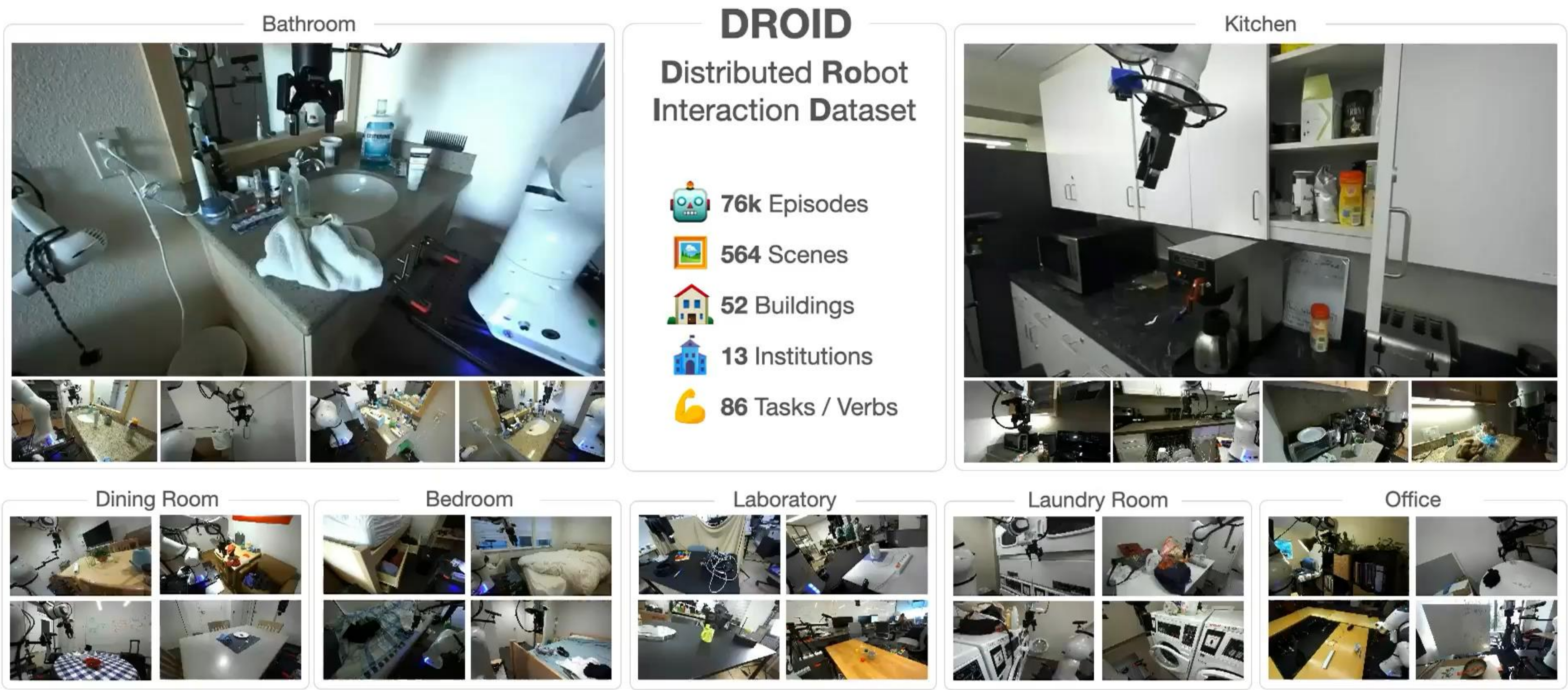
Imitation Learning from Large-Scale Multi-Task Datasets

Open-X

Open X-Embodiment Collaboration "Open X-Embodiment: Robotic Learning Datasets and RT-X Models", 2023

20+ academic institutions, 22 robot embodiments, 500 skills, 150,000 tasks

DROID



Khazatsky et al. "DROID: A Large-Scale In-the-Wild Robot Manipulation Dataset", 2024

76,000 episodes, 564 scenes, 52 buildings, 13 institutions, 86 tasks

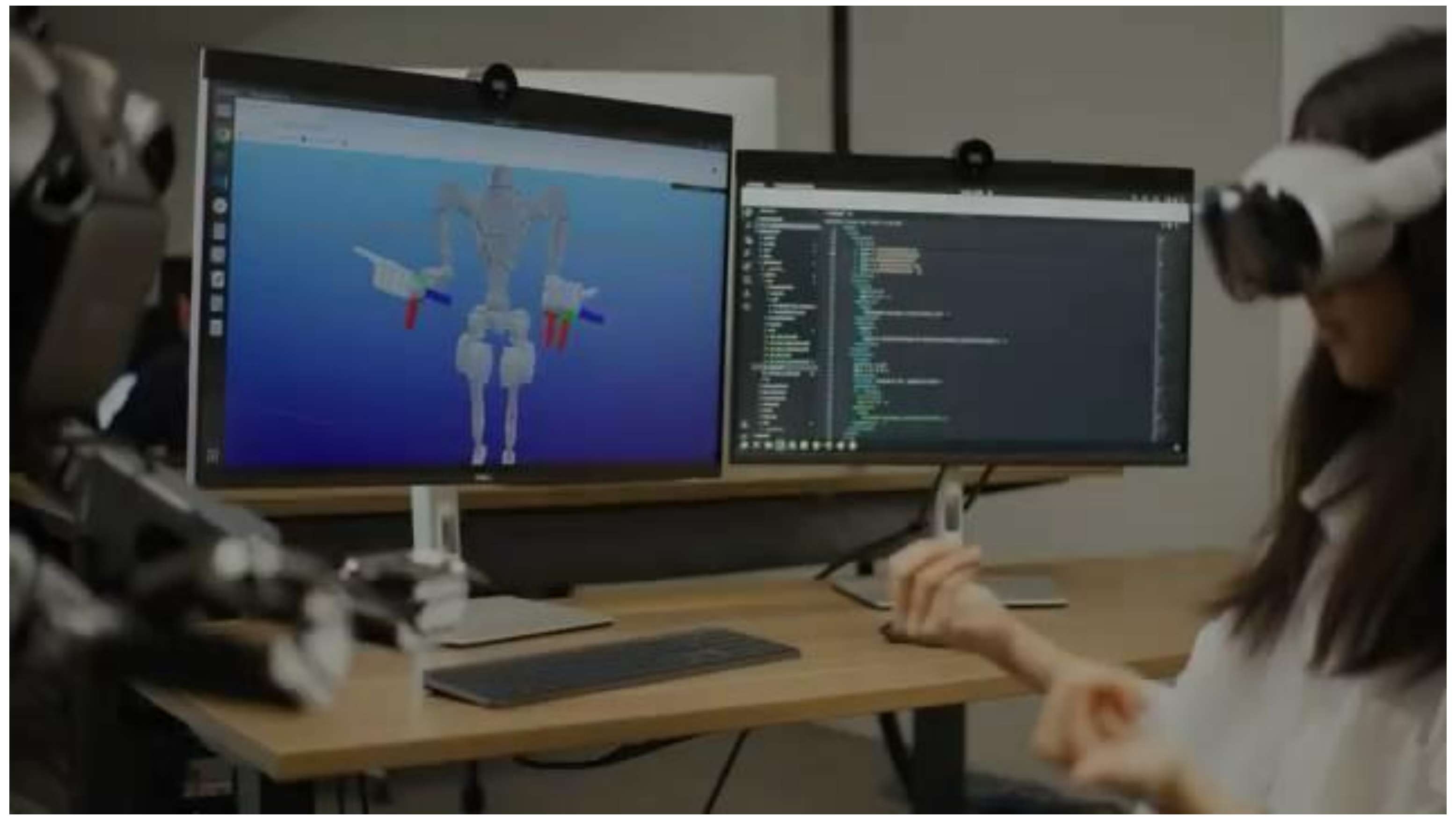
Imitation Learning from Large-Scale Multi-Task Datasets

Tesla

1X

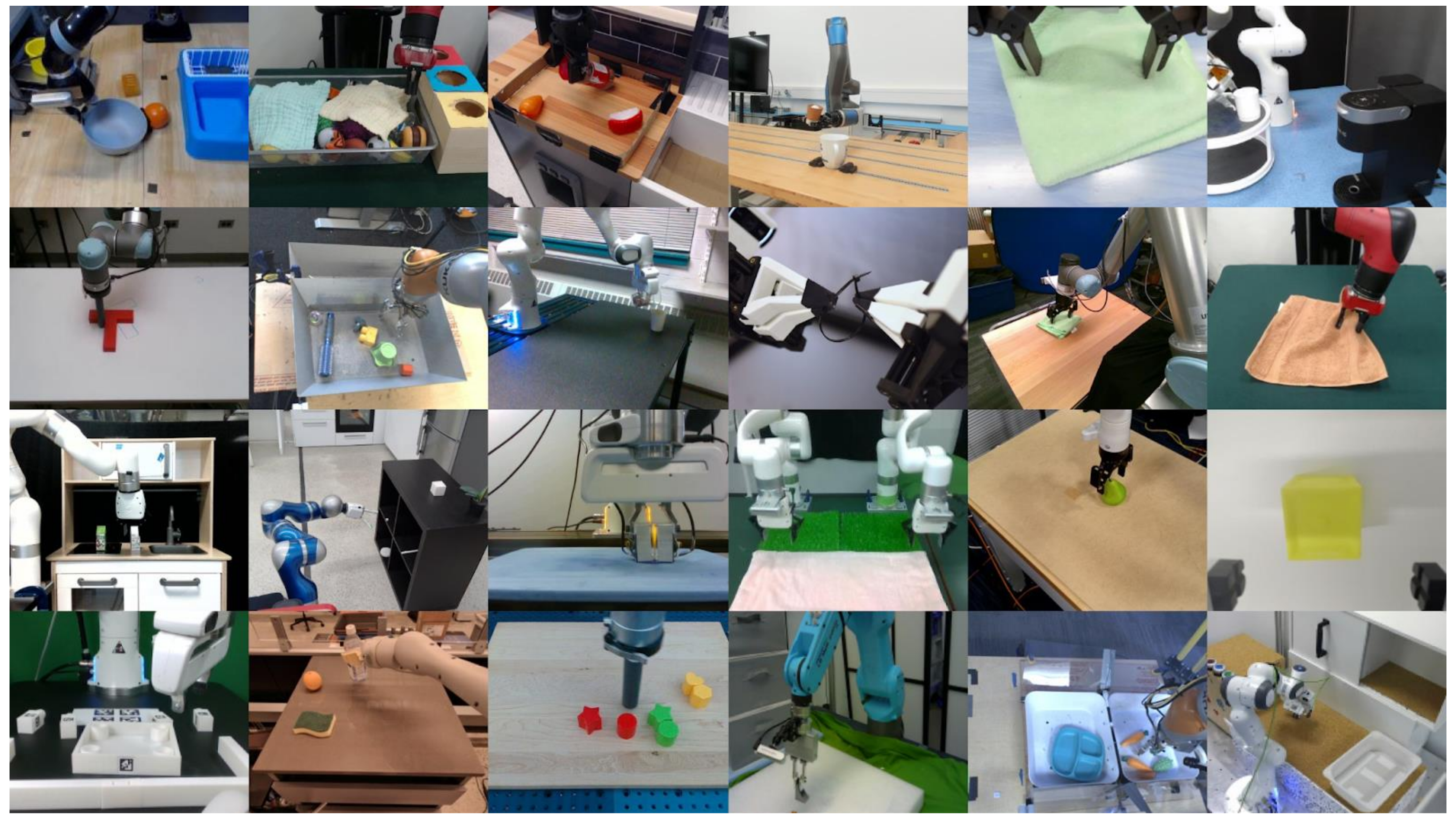
Physical Intelligence

NVIDIA Project GR00T



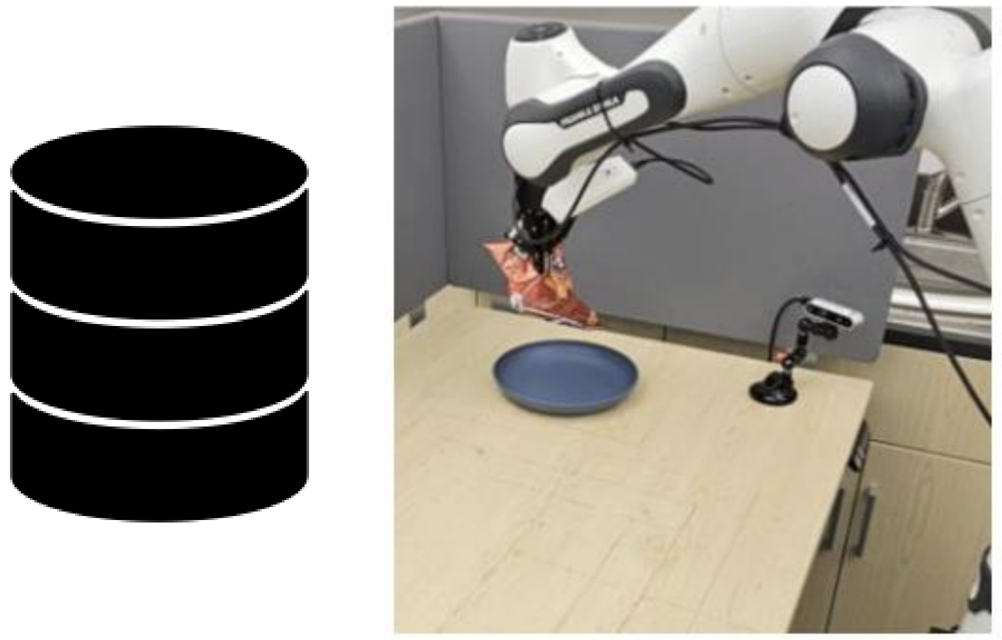
Using Large-Scale Multi-Task Datasets for Downstream Tasks

Large Co-Training Data

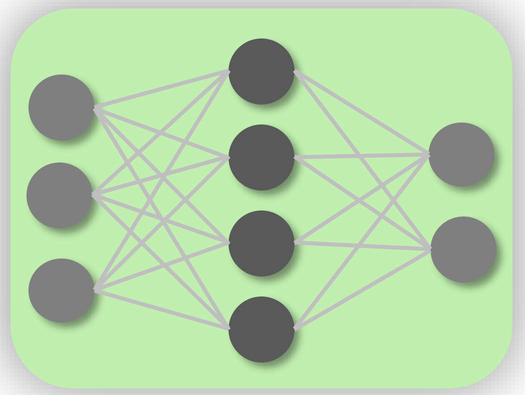


\+

Small Task-Specific Data



Policy π_θ



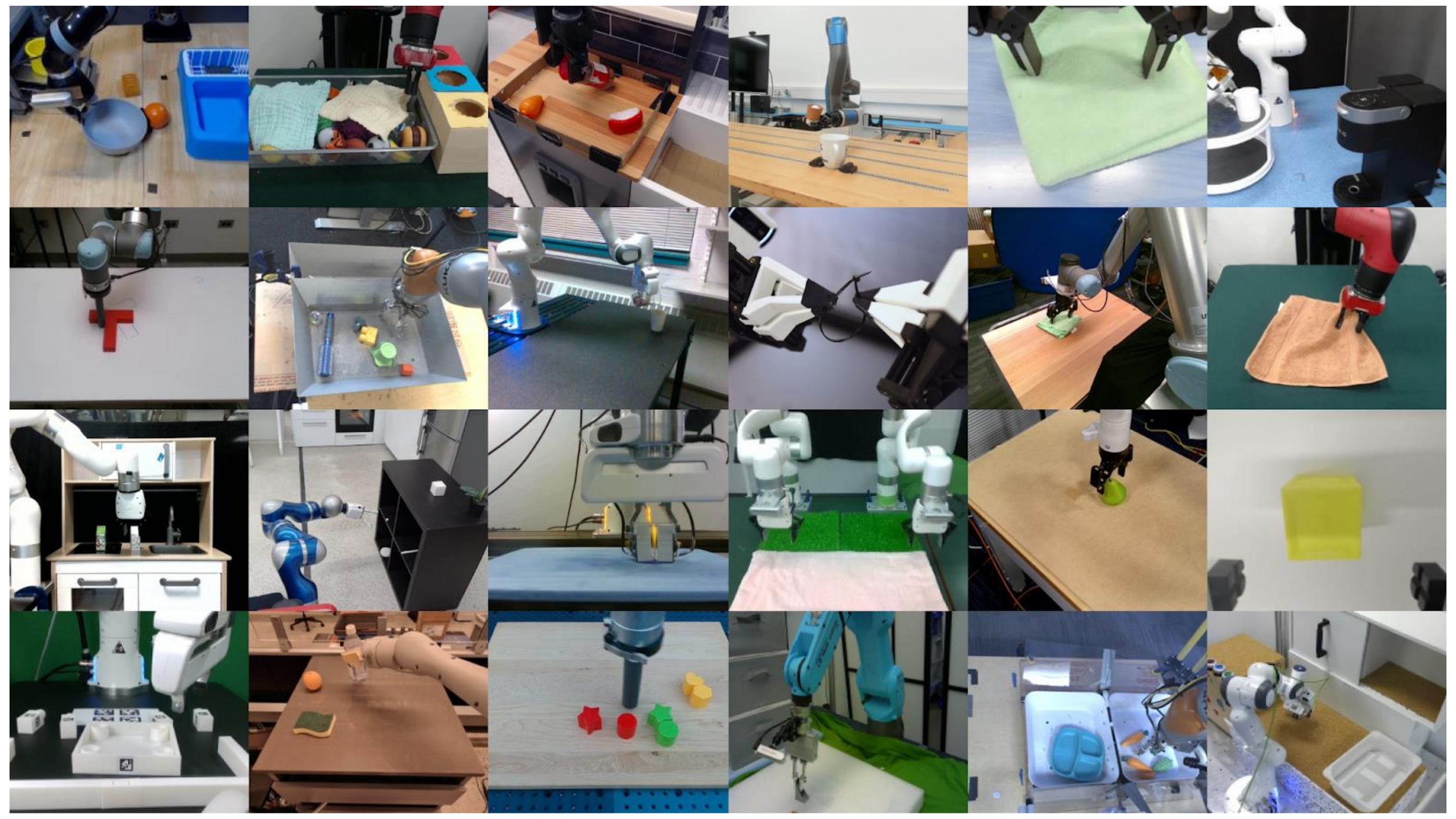
Goal: train closed-loop policy for your robot and task with much less task-specific data by using co-training data

Using Large-Scale Multi-Task Datasets for Downstream Tasks

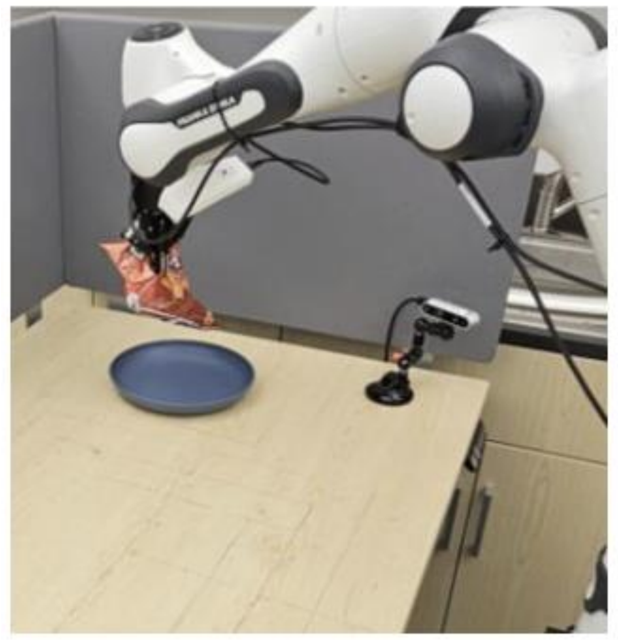
How does co-training data composition affect downstream task performance?

Large Co-Training Data

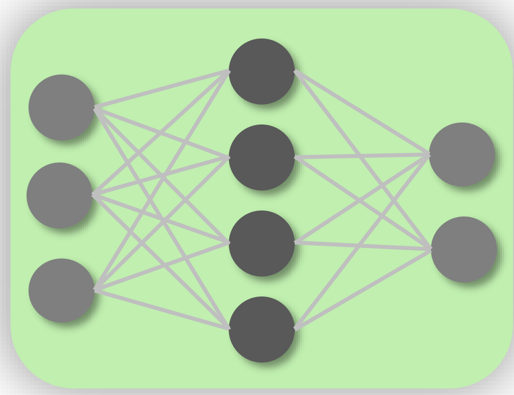
Small Task-Specific Data



+



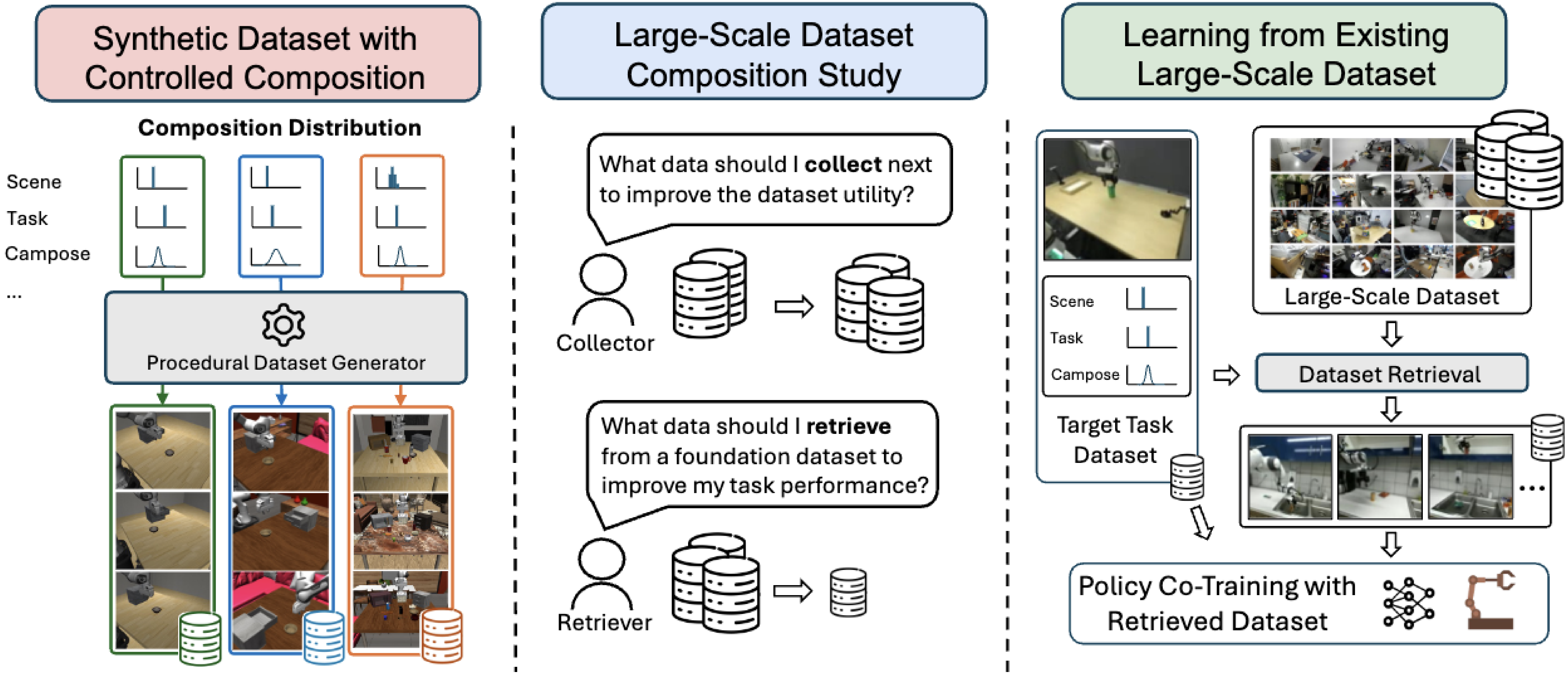
Policy π_θ



We can’t easily test different co-training datasets – they take months to collect!

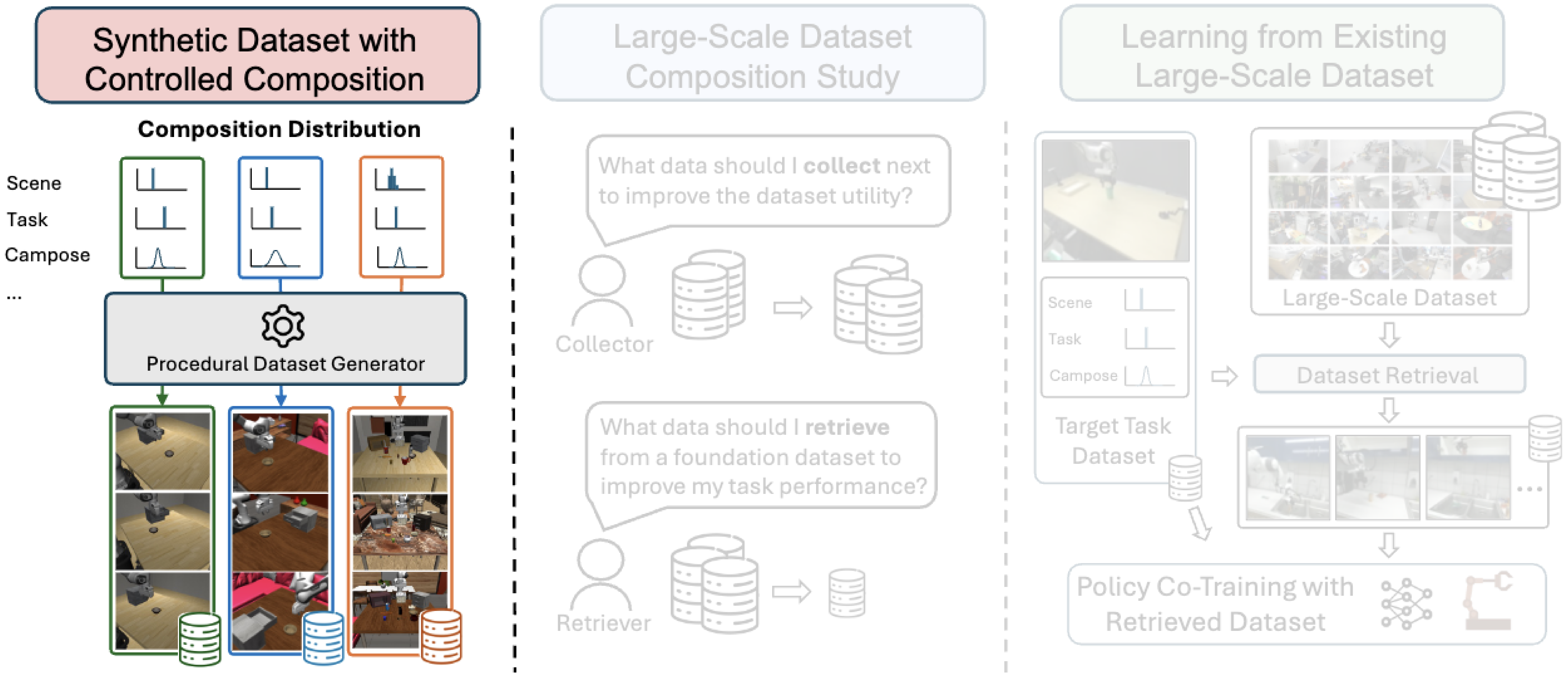
Idea: use synthetic data generation and simulation to conduct a study!

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning

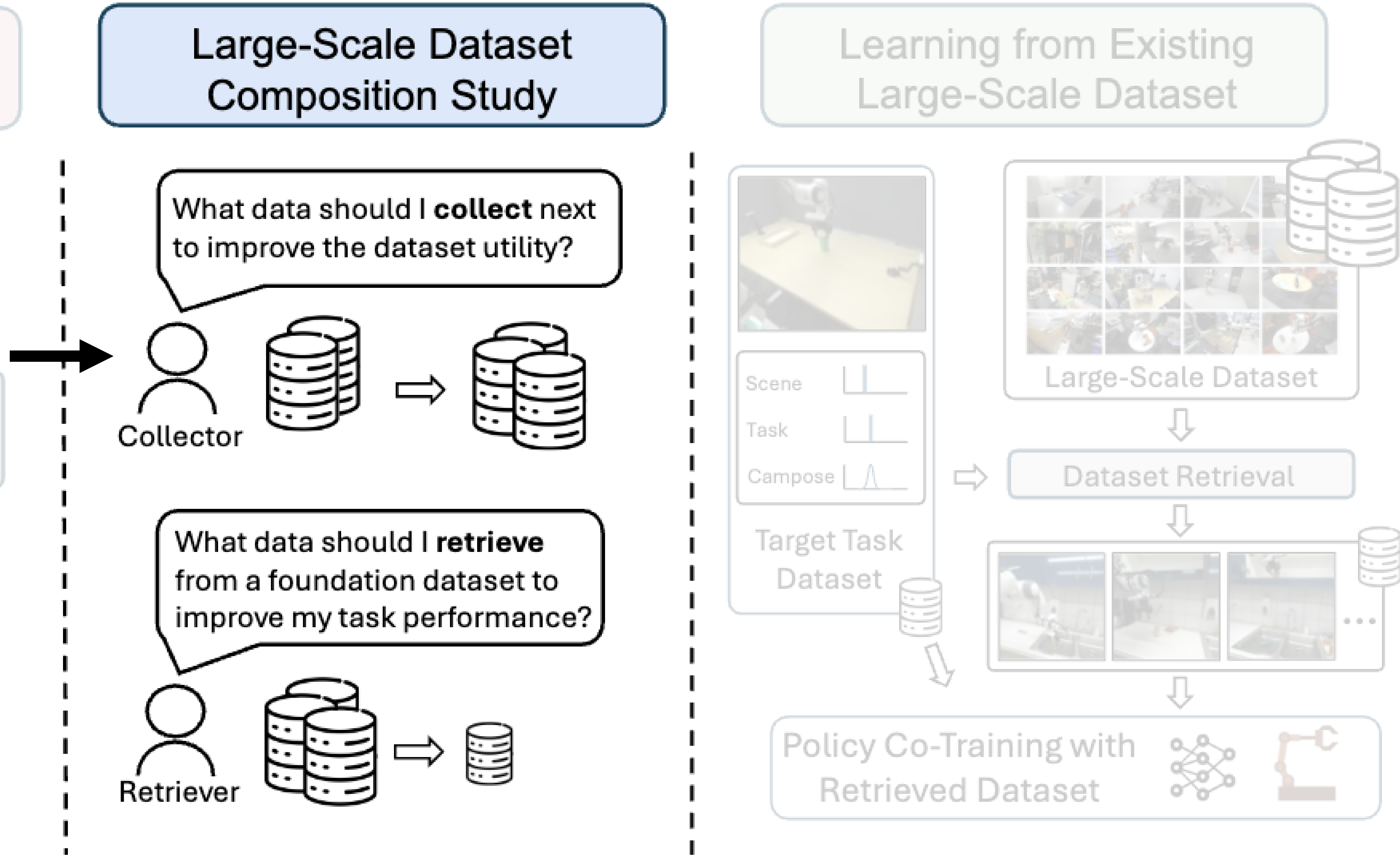


Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning

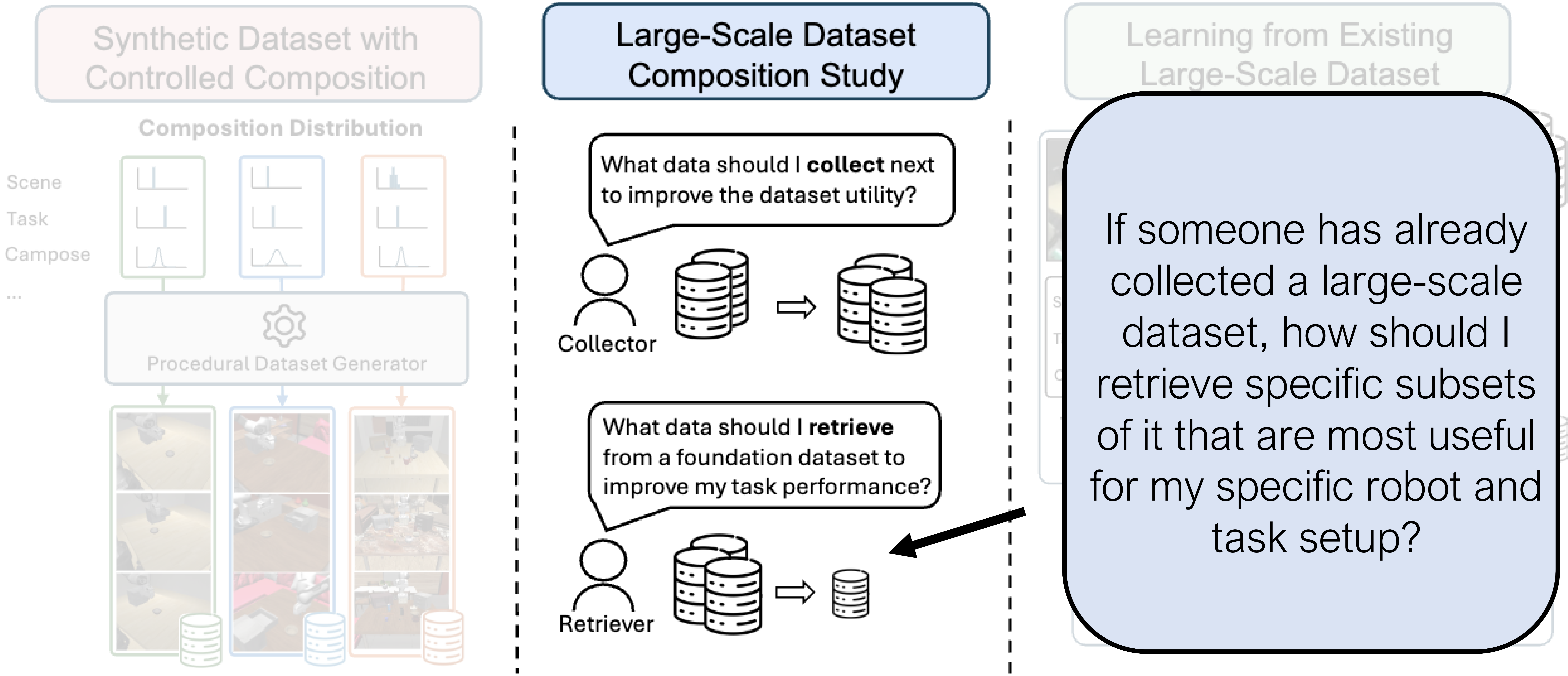
Given a finite data budget, what kinds of data should be collected?

Should I prioritize diversity in tasks, robot motions, objects, object spatial diversity, or camera locations?



Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

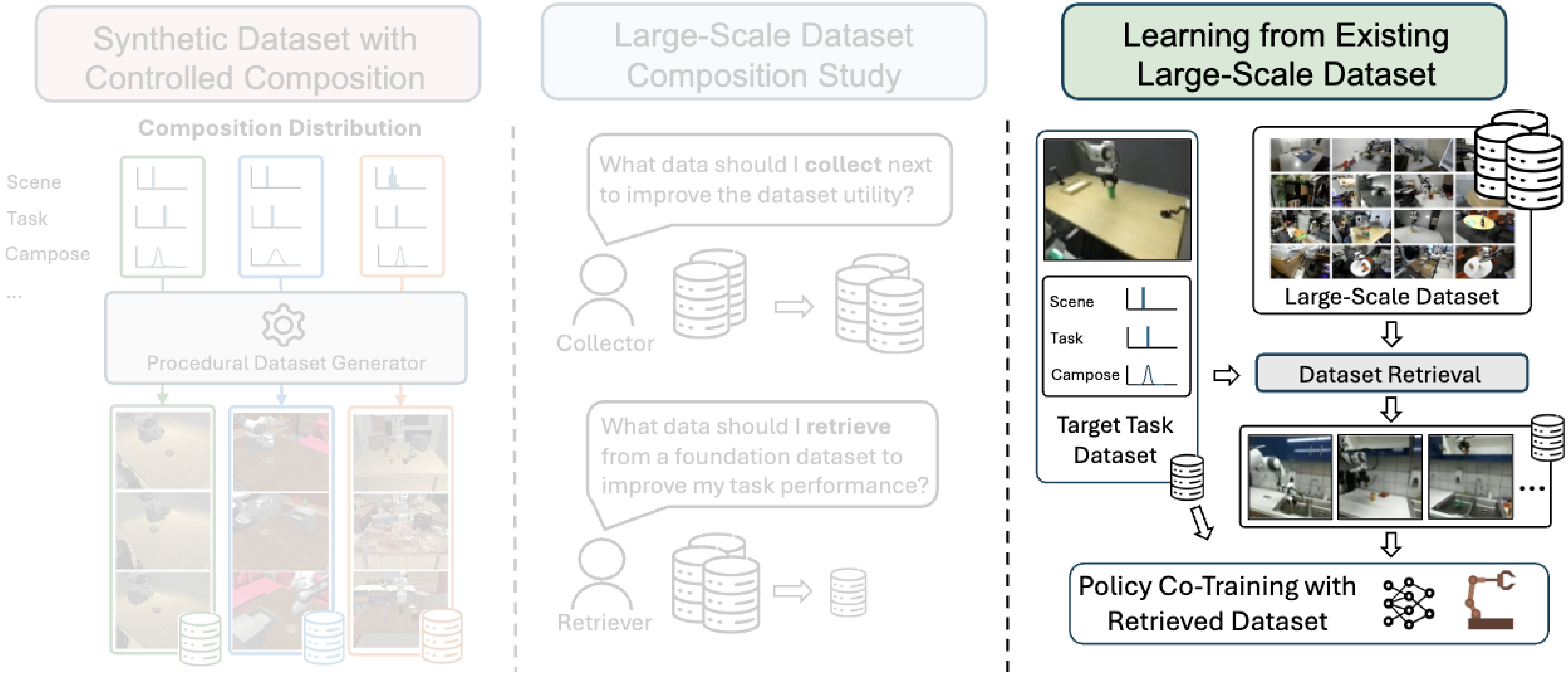
MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



If someone has already collected a large-scale dataset, how should I retrieve specific subsets of it that are most useful for my specific robot and task setup?

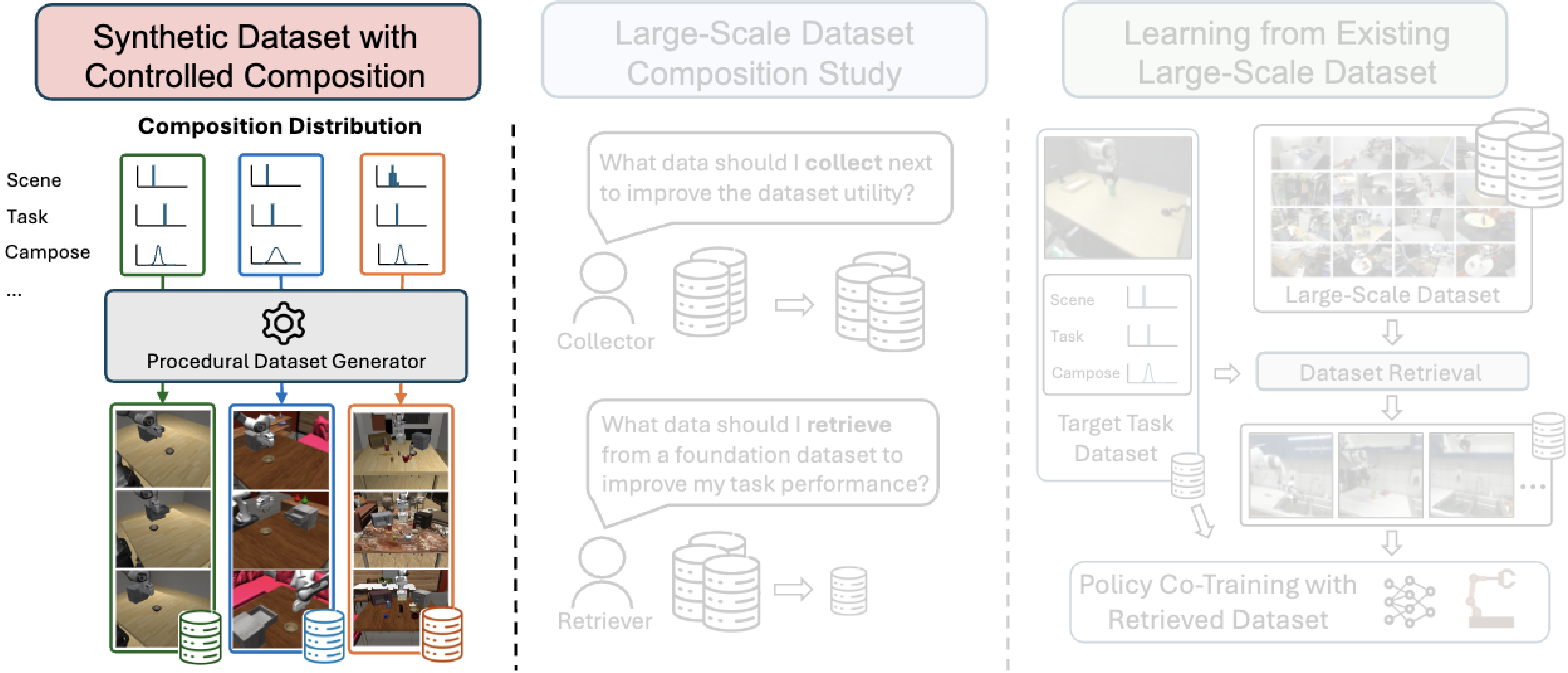
Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



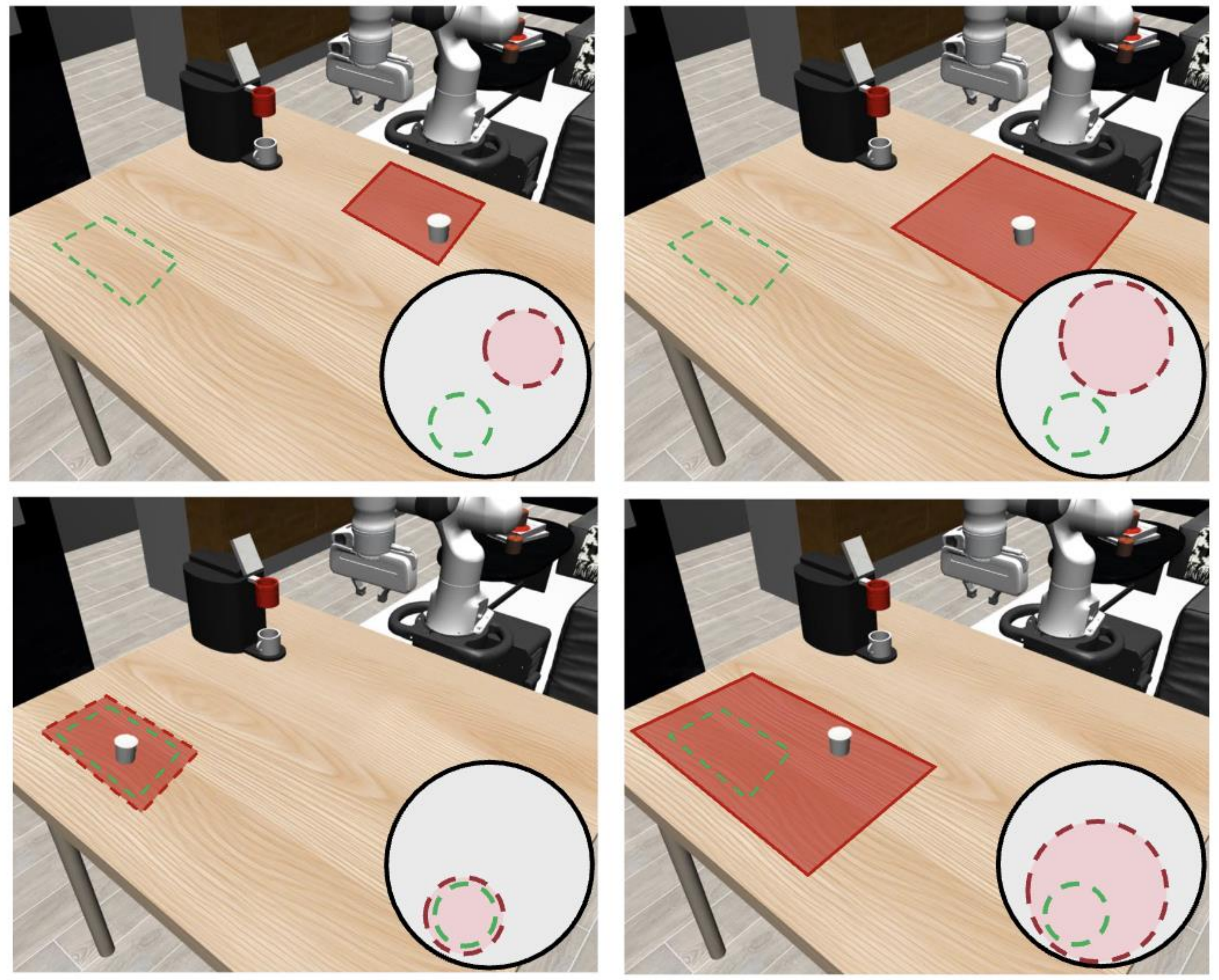
Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

DVs: Parametrizing Dataset Composition Distributions

- Each demonstration can be parameterized by "dimensions of variation (DVs)," some of which can be:
  - Visual – camera poses, object texture, table texture
  - Spatial – object and receptacle placements
  - Motion – skills that comprise a task

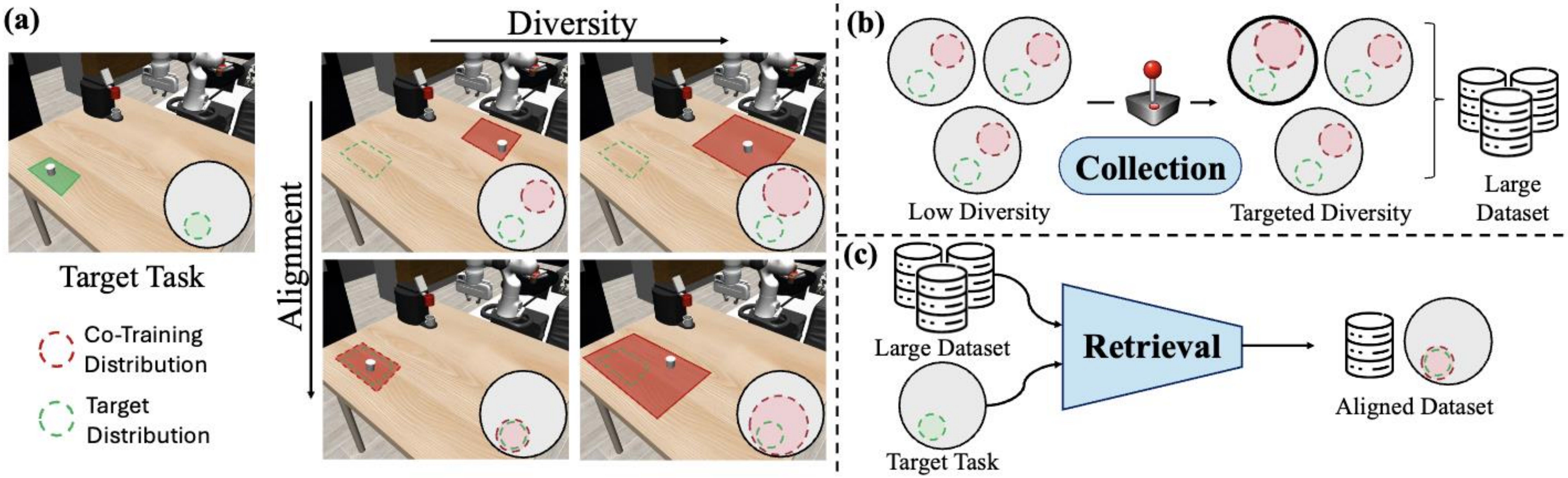
- Target and Co-training Distributions of Variation

Target Task



Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

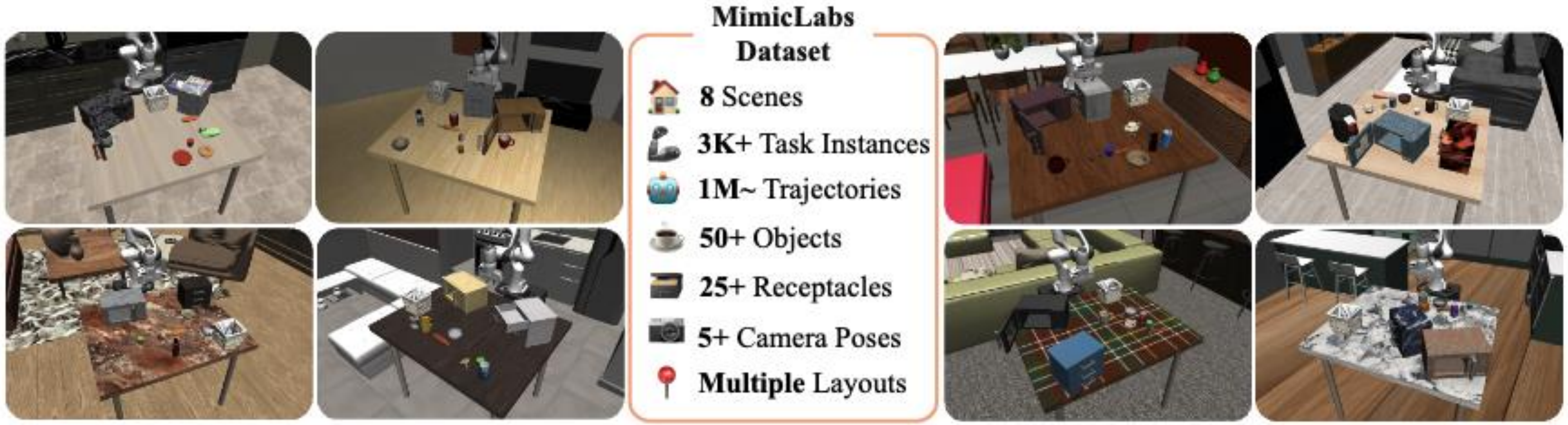
Relating DVs to the Collector and Retrieval Perspectives



- Collector: Given fixed data collection budget, which DVs should be more diverse?
- Retriever: Which DVs should I try to align with my setup?

Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

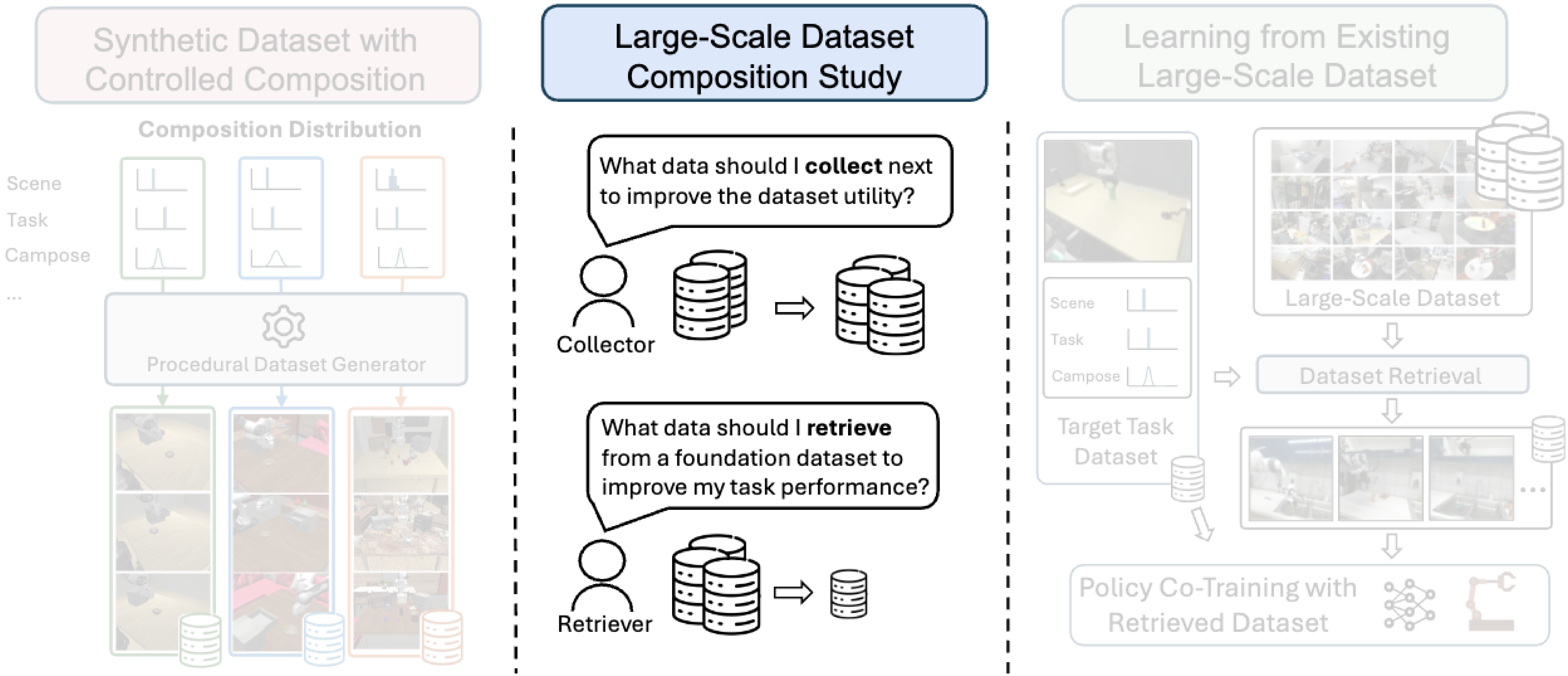
MimicLabs: a Large-Scale Retrievable Demonstration Dataset



- Procedural task specification as a BDDL (Behavior Domain Definition Language) [Srivastava et al. "BEHAVIOR" and Liu et al. "LIBERO"]
- Controllable data synthesis with MimicGen to scale ~500 human demonstrations up to ~1M trajectories

Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

MimicLabs: Testing different co-training dataset compositions

Baseline
(co-training)

Alternative Co-training Distributions

camPose

objTex

tableTex

objSpat

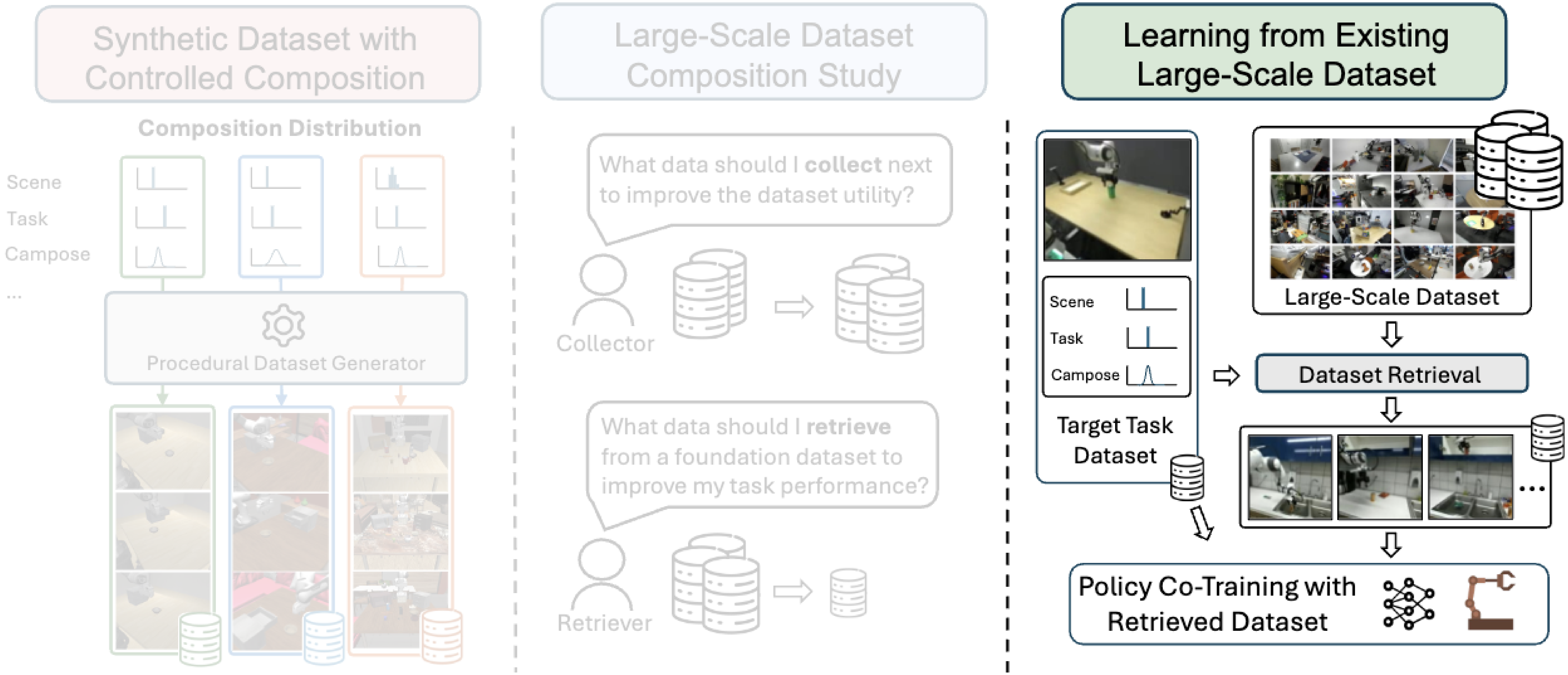
MimicLabs: Takeaways from Sim Retriever Experiment

Target task	#demos	Target only	Retrieving relevant object/skill					Missing relevant object/skill				
			obj/skill	*+ camPose*	*+ objSpat*	*+ recepSpat*	*+ all*	*no-obj/skill*	*+ camPose*	*+ objSpat*	*+ recepSpat*	*+ all*
bin carrot	10	50	70	**96.67**	86.67	83.33	90	30	40	53.33	43.33	**56.67**
bin bowl	10	33.33	50	70	53.33	63.33	**73.33**	36.67	**60**	50	40	46.67
clear table	10	23.33	20	20	20	20	23.33	20	23.33	20	20	16.67
	20	36.67	43.33	**46.67**	43.33	43.33	40	33.33	33.33	23.33	**43.33**	30
microwave teapot	10	23.33	20	23.33	16.67	13.33	16.67	10	10	6.67	13.33	**20**
	20	30	33.33	**50**	33.33	36.67	33.33	20	**36.67**	26.67	33.33	23.33
make coffee	10	13.33	10	**23.33**	6.67	13.33	6.67	3.33	13.33	6.67	6.67	6.67
	50	33.33	33.33	36.67	30	36.67	33.33	30	30	30	30	**40**
top-3 labs	50	33.33	30	**53.33**	40	40	40	36.67	36.67	**43.33**	30	40

- Retrieving skills for co-training, significantly boosts performance
- Aligning camera poses enables better transfer of skills
- Quality often matters over quantity in co-training data

Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

Case Study: Apply Study Insights to Retrieval from DROID

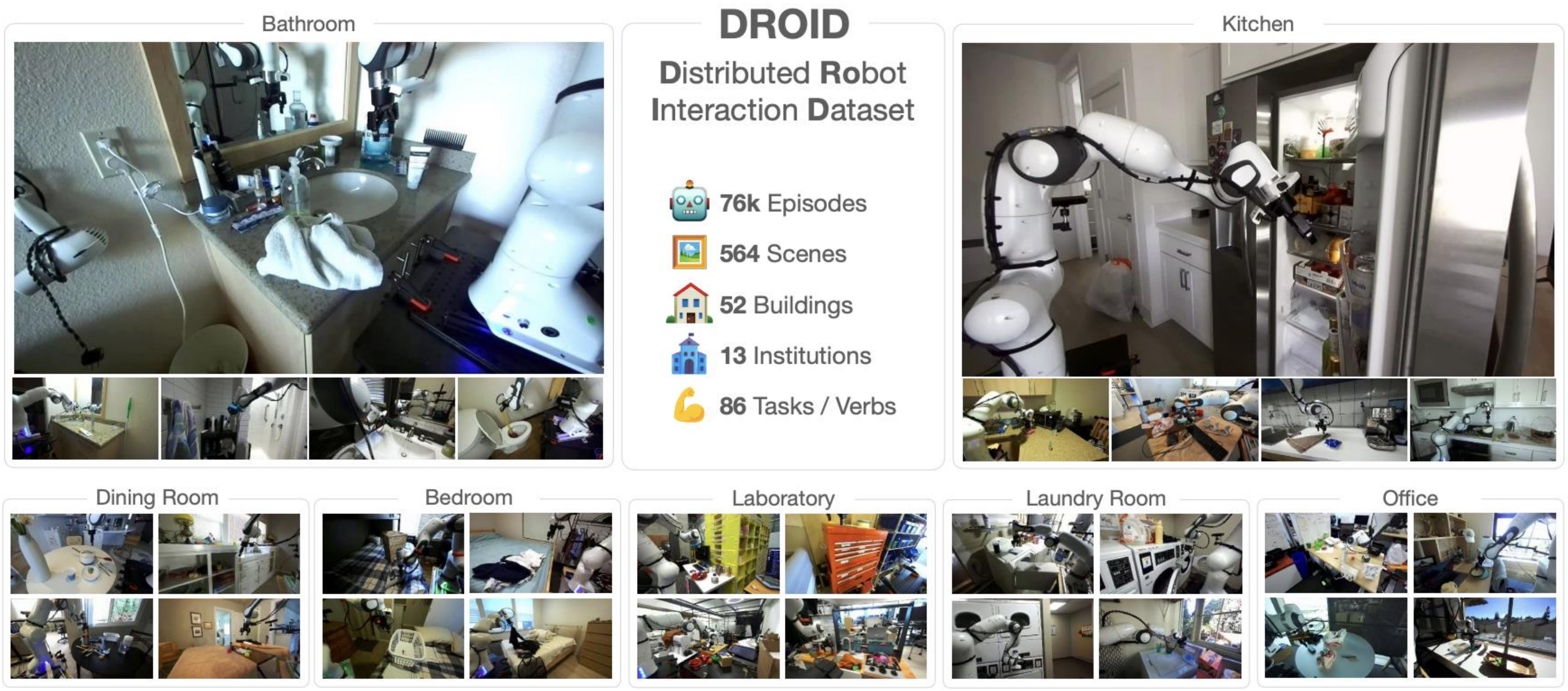
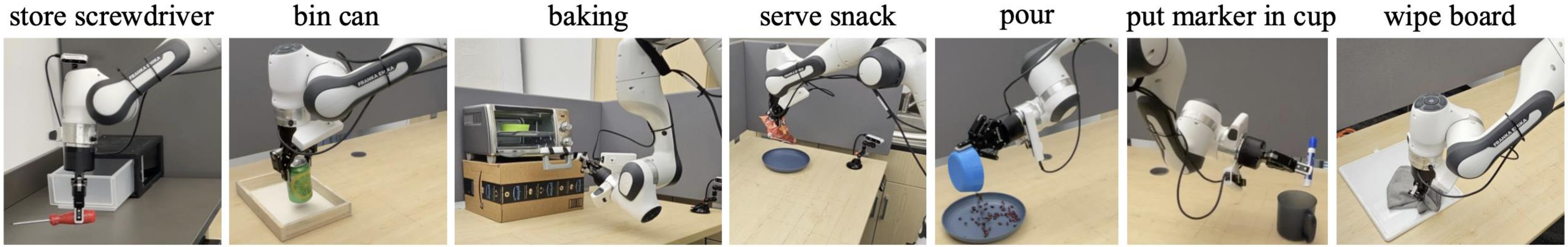


Fig. 1: We introduce DROID (**D**istributed **Ro**bot **I**nteraction **D**ataset), an "in-the-wild" robot manipulation dataset with 76k trajectories or 350 hours of interaction data, collected across 564 scenes, 86 tasks, and 52 buildings over the course of 12 months. Each DROID episode contains three synchronized RGB camera streams, camera calibration, depth information, and natural language instructions. We demonstrate that training with DROID leads to policies with higher performance, greater robustness, and improved generalization ability. We open source the full dataset, pre-trained model checkpoints, and a detailed guide for reproducing our robot setup.

Case Study: Real Robot Tasks



Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

Case Study: Qualitative Examples of DROID Retrieval

Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

Retrieval from DROID based on simulation takeaways result in large performance boost

Target task	**#demos**	**Target only**	**DROID**	*obj/skill*	*+camPose*	*+objTex*	*+objSpat*
serve snack	20	5	0	65	70	35	**85**
bin can	20	60	0	15	**85**	65	**85**
pour	20	50	0	35	**75**	60	65
wipe board	20	55	0	45	55	55	**65**
baking	20	40	0	40	**55**	40	35
put marker in cup	50	30	0	20	**35**	15	20

- Aligned retrieval can significantly improve performance
- Random co-training with a large dataset can hurt performance
- Retrieval is a promising direction and standardization of dataset metadata will be beneficial

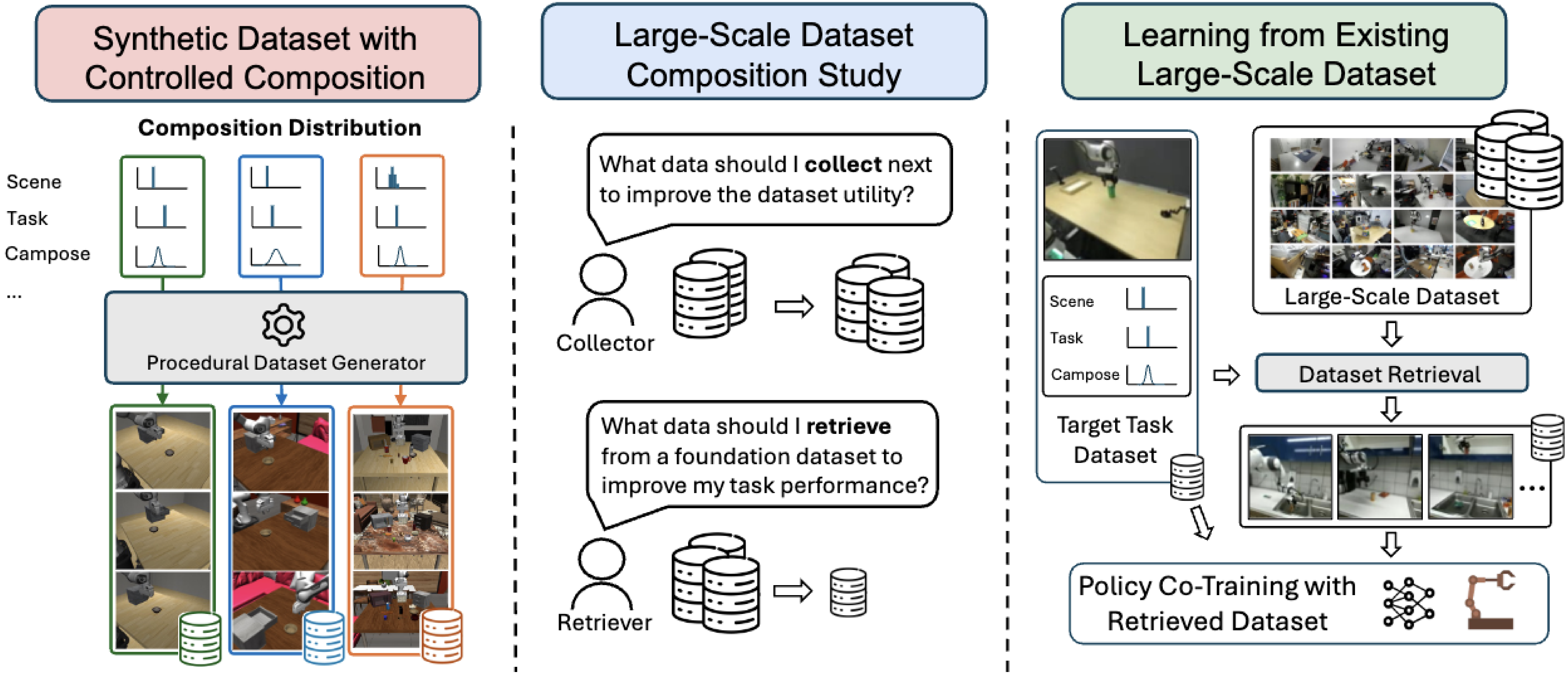
Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

Retrieval from DROID based on simulation takeaways result in large performance boost

Retrieval from DROID based on simulation takeaways result in large performance boost

Retrieval from DROID based on simulation takeaways result in large performance boost

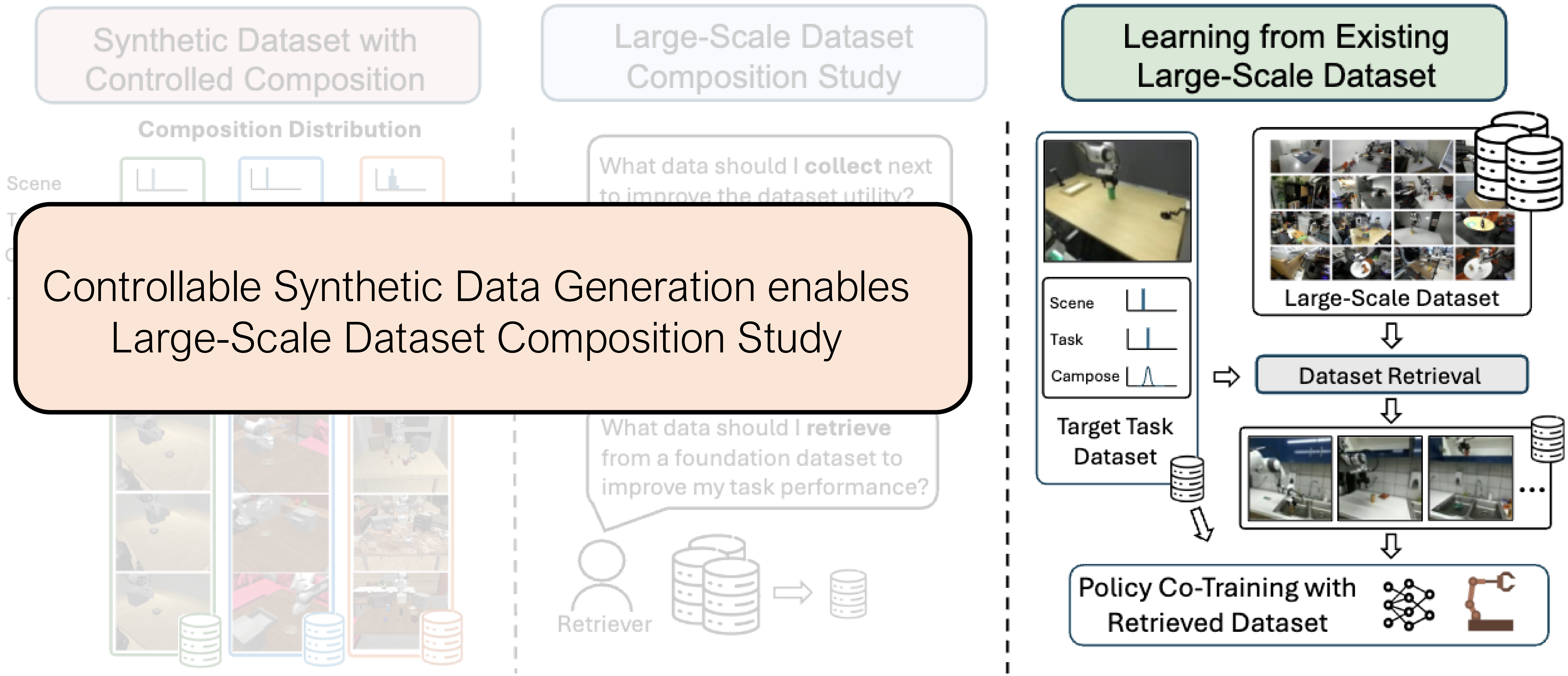
MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

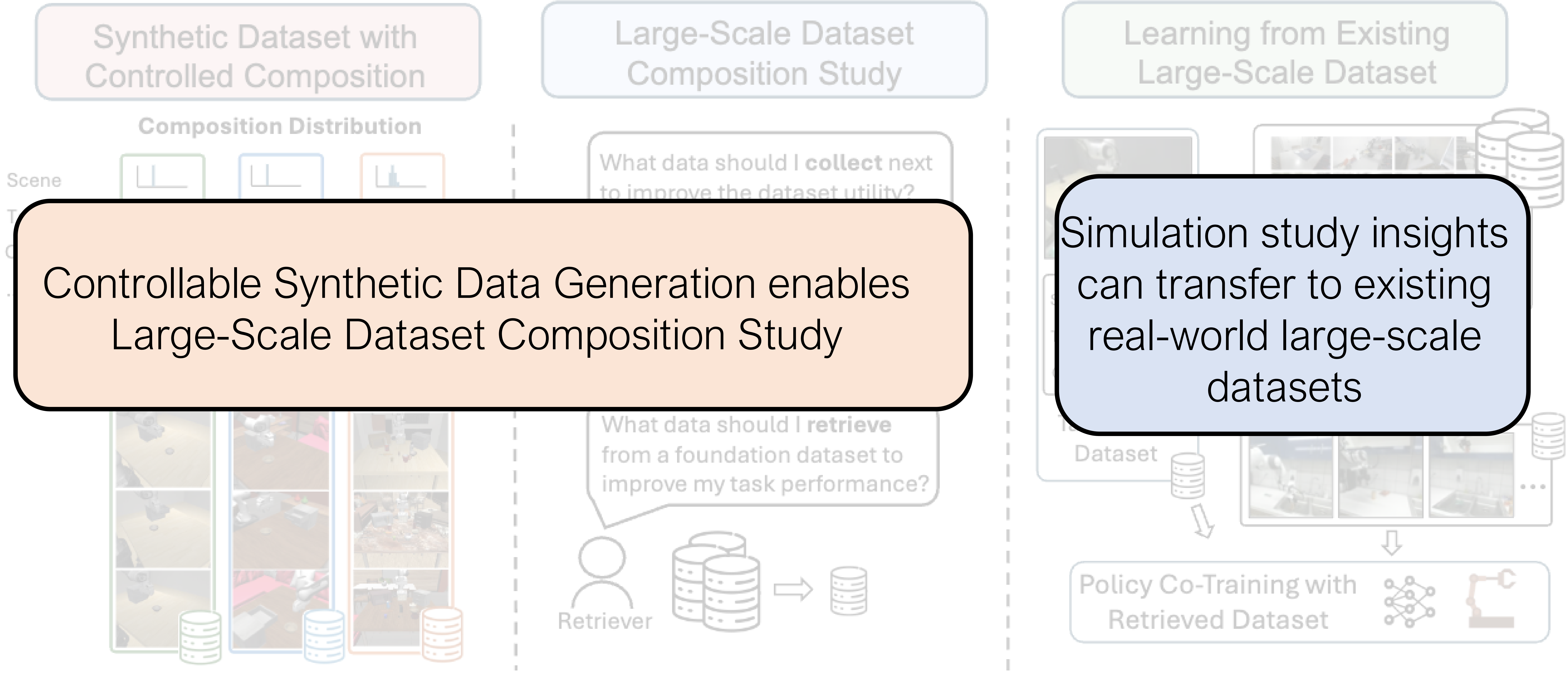
MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning

Controllable Synthetic Data Generation enables Large-Scale Dataset Composition Study



Saxena et al. "What Matters in Learning from Large-Scale Datasets for Robot Manipulation", ICLR 2025

MimicLabs Study: How Large-Scale Dataset Composition Influences Policy Learning



Controllable Synthetic Data Generation enables Large-Scale Dataset Composition Study

Simulation study insights can transfer to existing real-world large-scale datasets

Saxena et al. “What Matters in Learning from Large-Scale Datasets for Robot Manipulation”, ICLR 2025

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations

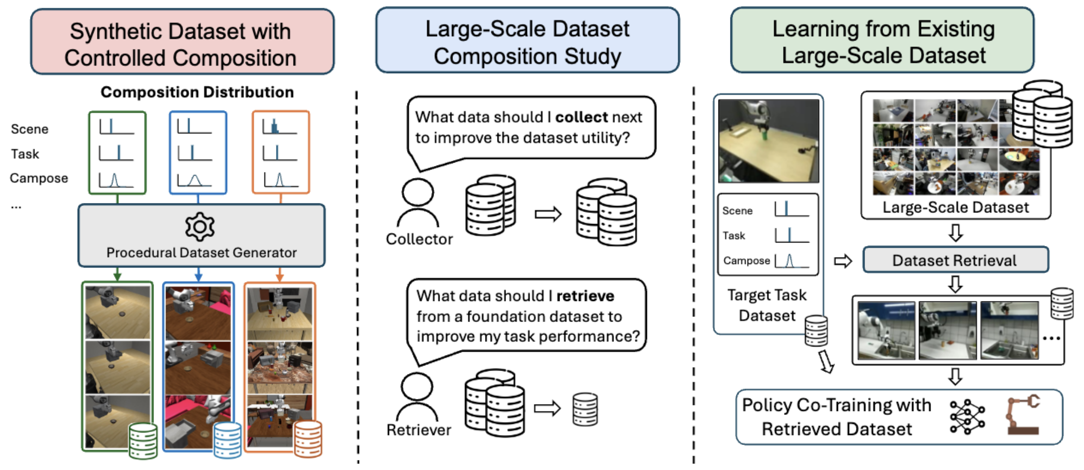
OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)



MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen (ICRA 2025)

SkillMimicGen (CoRL 2024)

AHA (ICLR 2025)

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations

OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)

MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen (ICRA 2025)

SkillMimicGen (CoRL 2024)

AHA (ICLR 2025)

Scaling data collection is expensive, especially for humanoids

Tesla Optimus Robot Demo

Requirements:

- Must be able to walk 7+ hours a day while carrying up to 30 lbs.
- Ability to wear and operate a motion capture suit and VR headset for extended periods of time.
- Continuous hand/eye coordination and fine manipulation, body coordination, and kinesthetic awareness and ability to walk up/downstairs.
- Must have the ability to stand, sit, walk, stoop, bend, reach, crouch, and twist throughout the day.
- Ability to work a flexible schedule: day/night shift and 1 weekend day + overtime when needed.

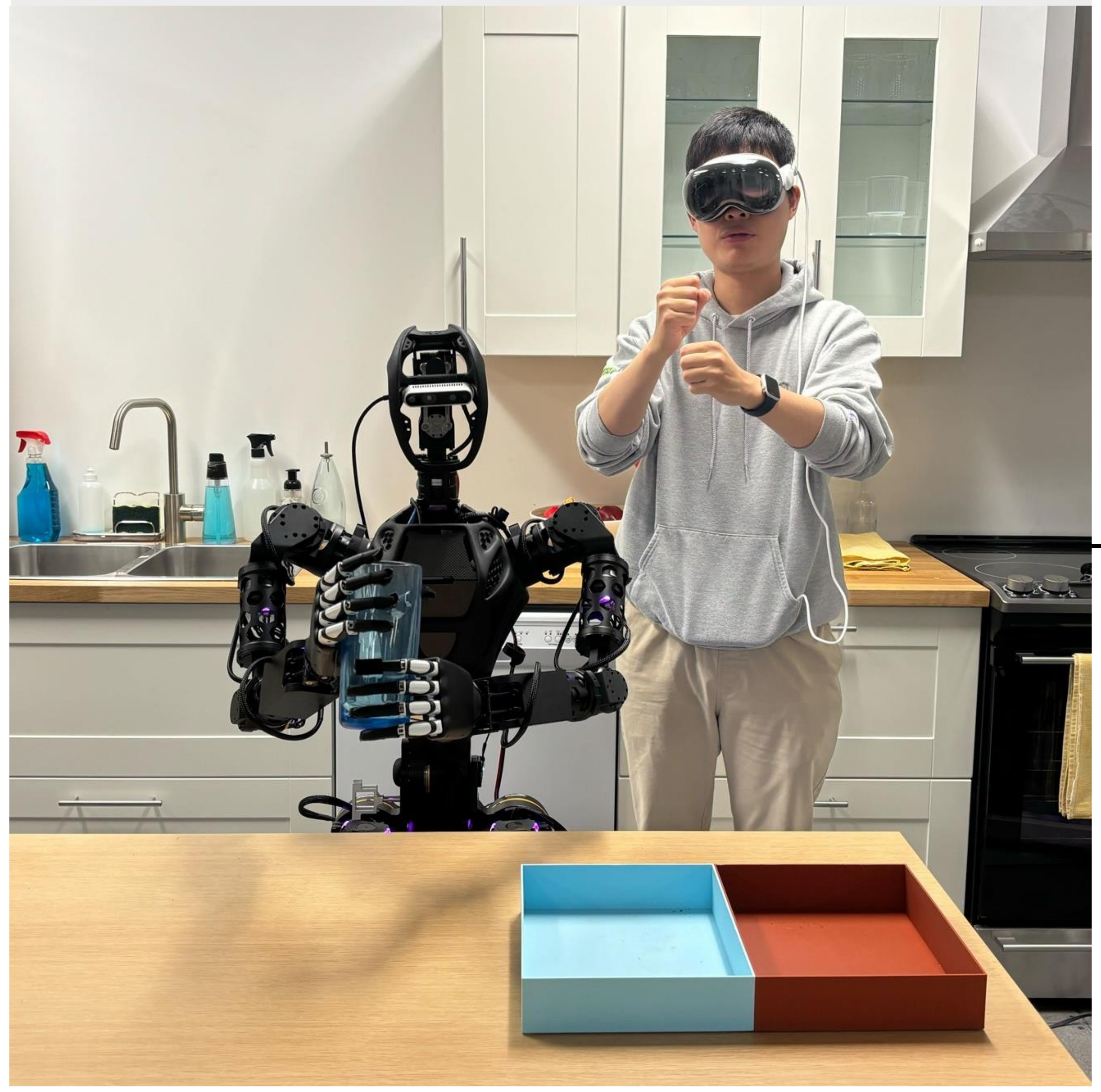
Source: Tesla Data Collection Operator job posting

DexMimicGen

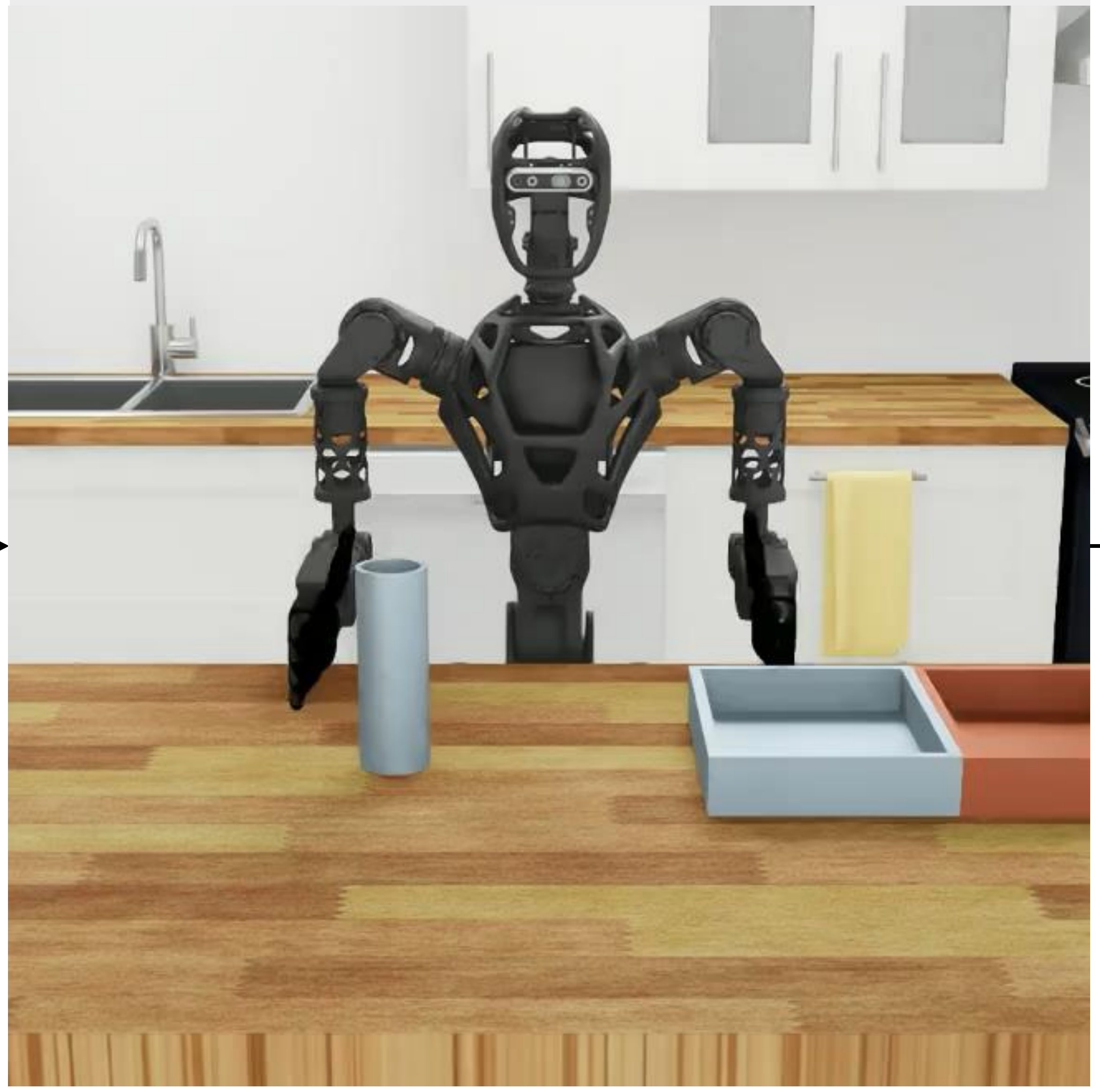
Automated Data Generation for Bimanual Dexterous Manipulation

DexMimicGen generates large-scale data for bimanual dexterous manipulation

Human teleoperation

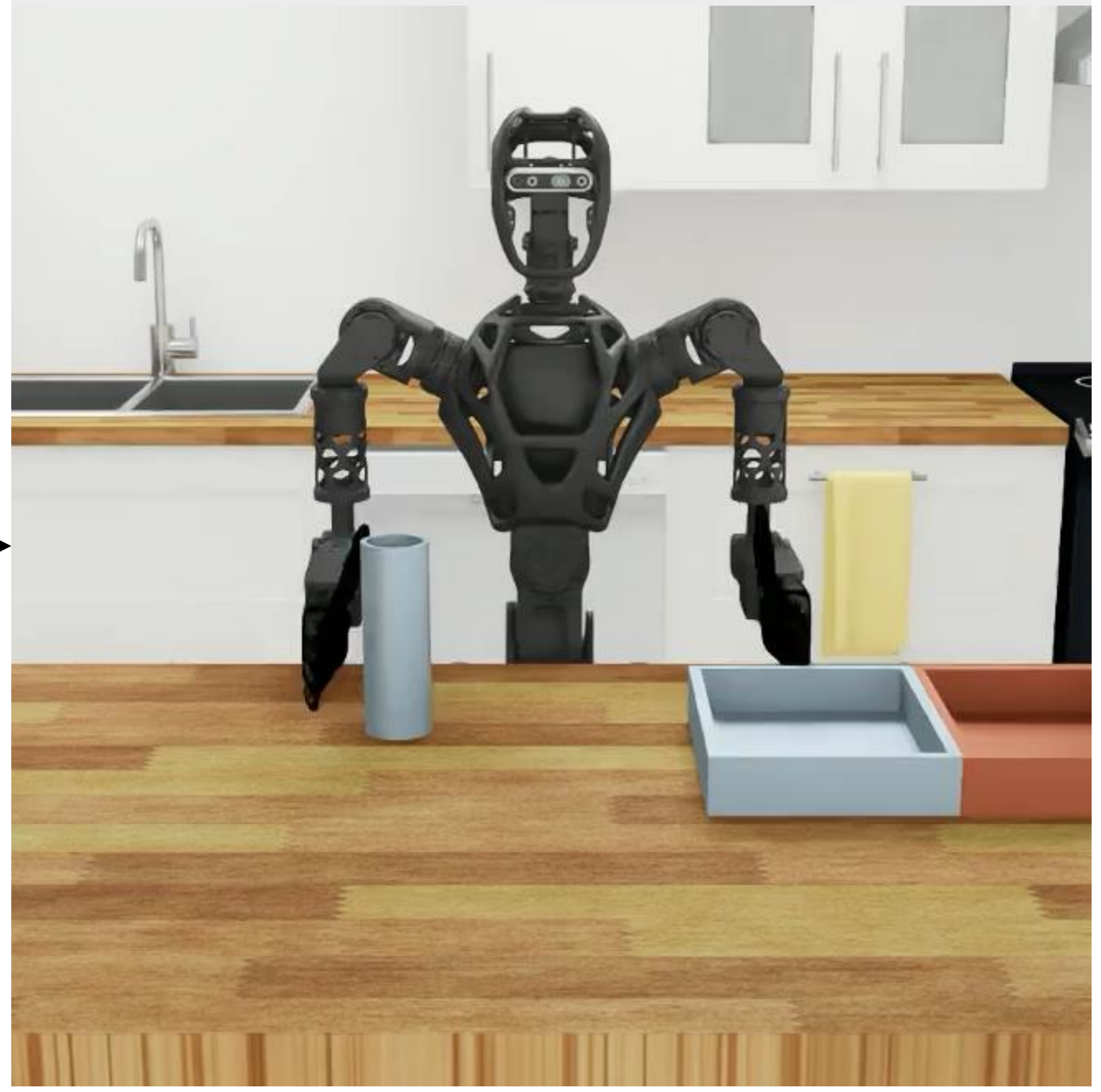


5 source demos



DexMimicGen

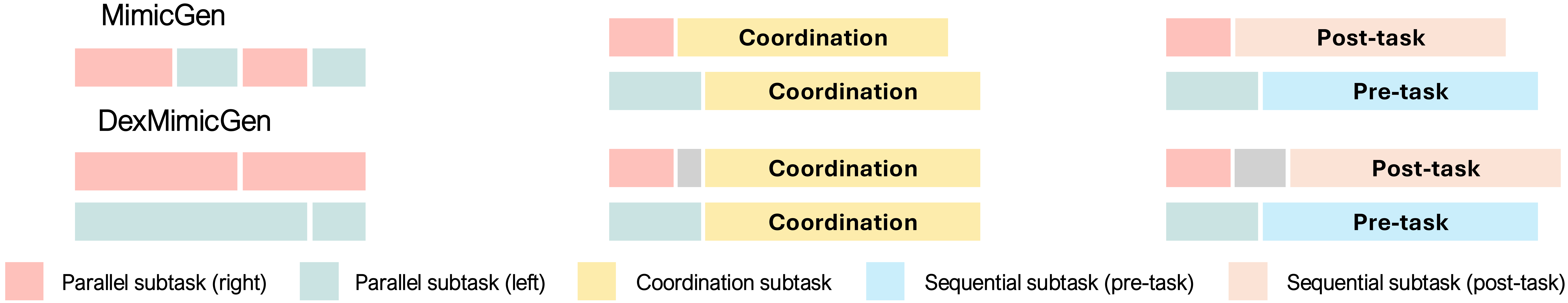
1000 generated demos



Jiang et al. “DexMimicGen: Automated Data Generation for Bimanual Dexterous Manipulation via Imitation Learning”, ICRA 2025

Challenge: Bimanual Coordination

Parallel subtasks | Coordination subtasks | Sequential subtasks



DexMimicGen generates data for a large range of tasks

Contact-rich tasks

DexMimicGen generates data for a large range of tasks

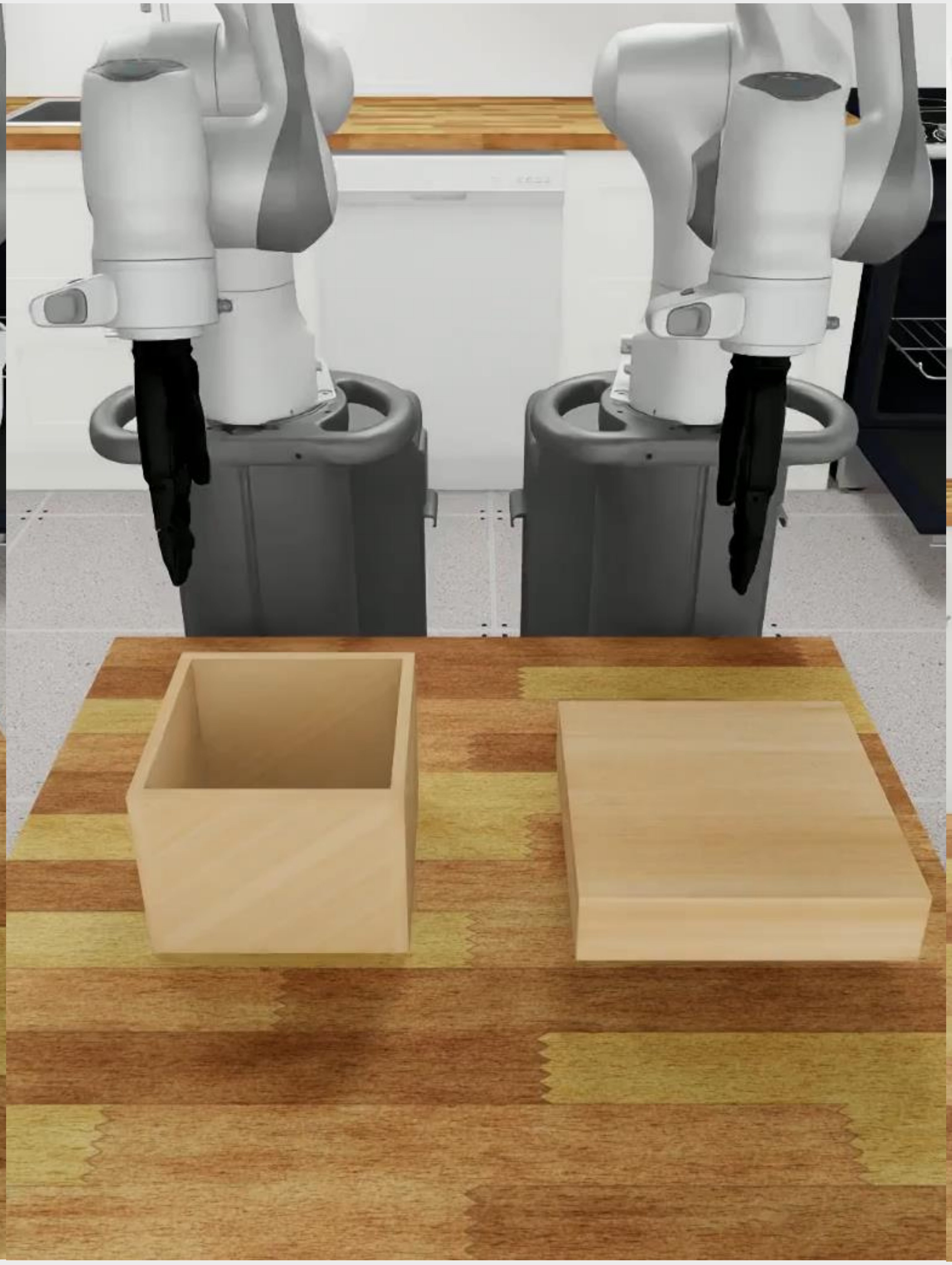
Long-horizon tasks

DexMimicGen generates data for a large range of tasks

Articulated object manipulation

DexMimicGen generates data for tasks with different coordination complexities

Parallel subtasks

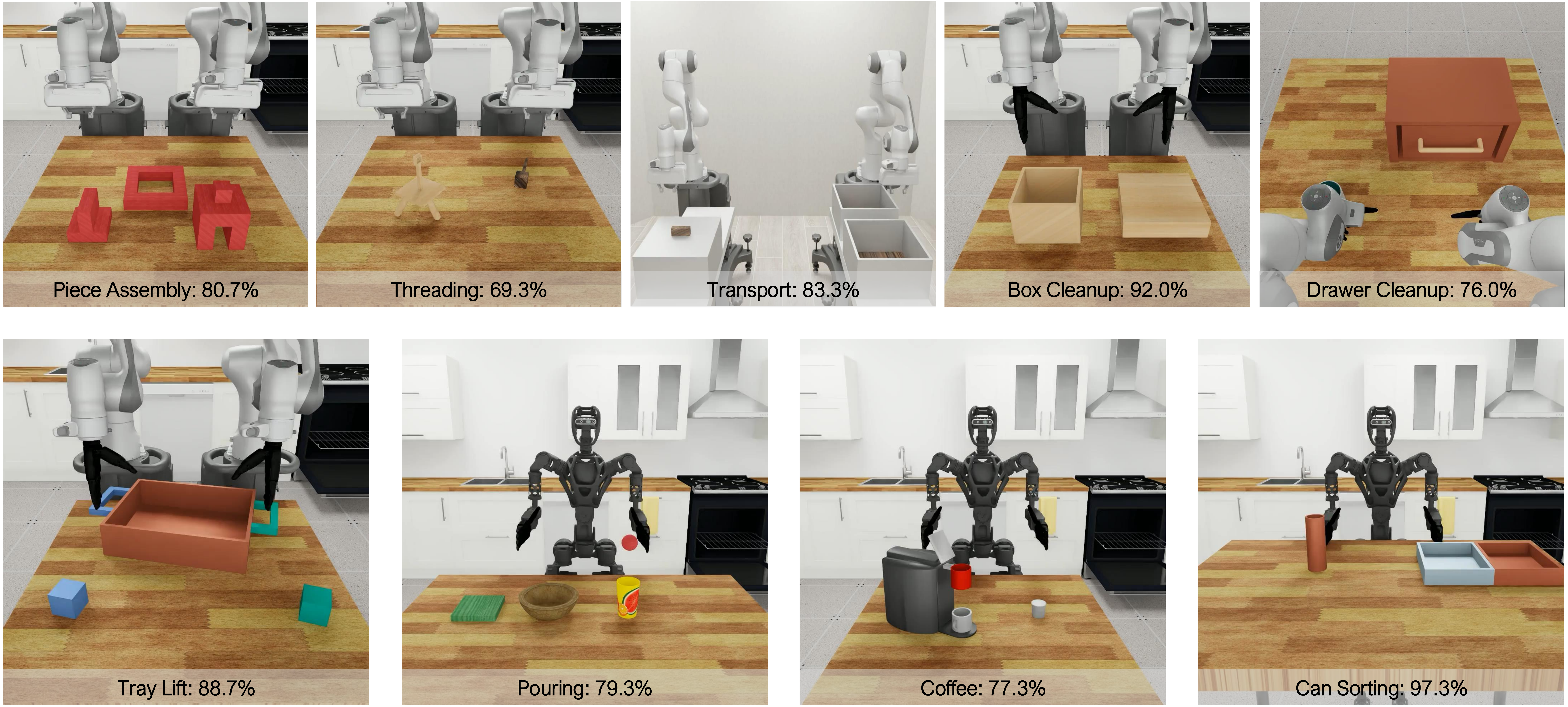


Coordination subtasks

Sequential subtasks

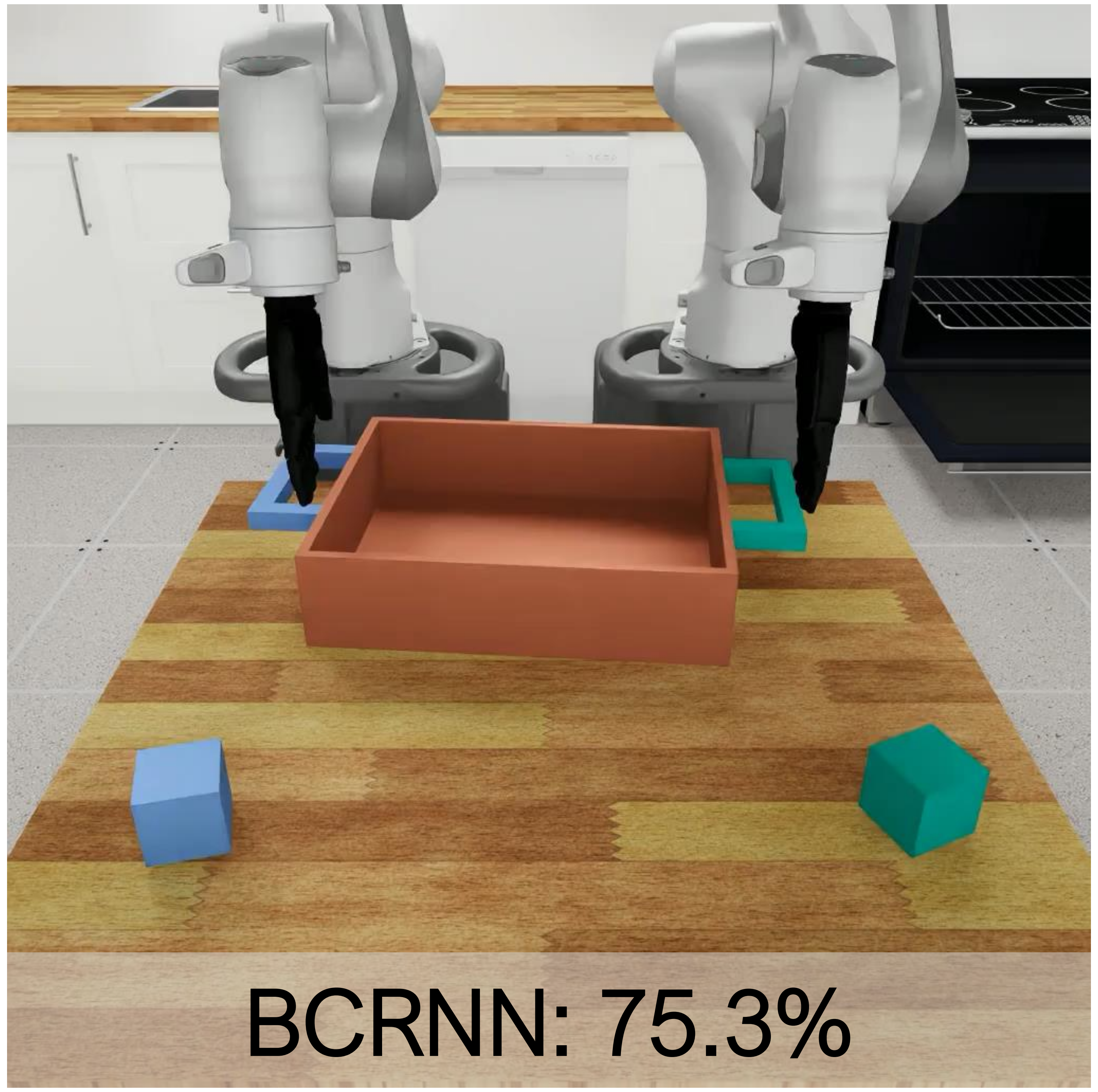

DexMimicGen generated 21k demos from 60
source human demos.

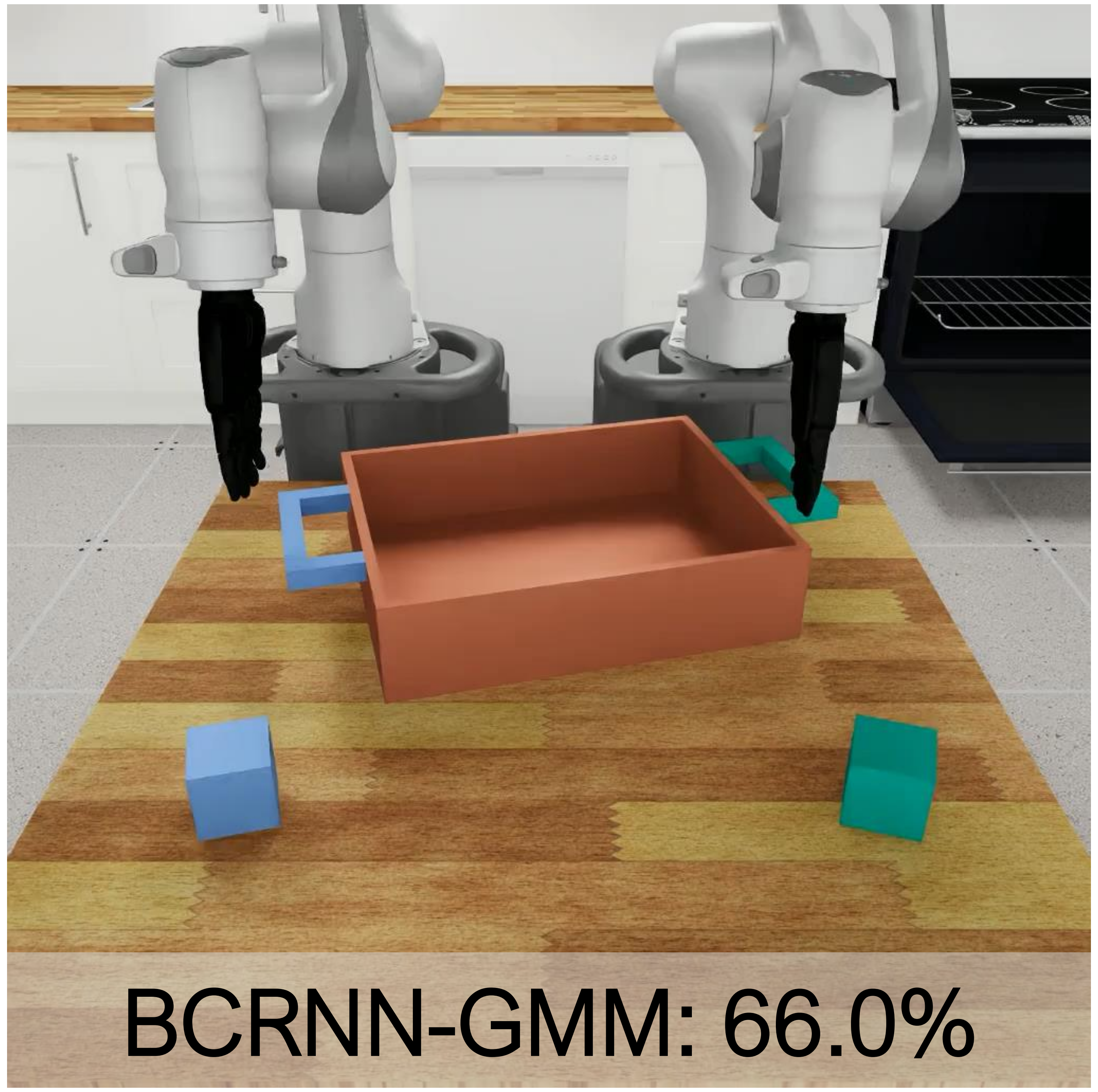
Behavioral cloning on DexMimicGen data trains performant policies

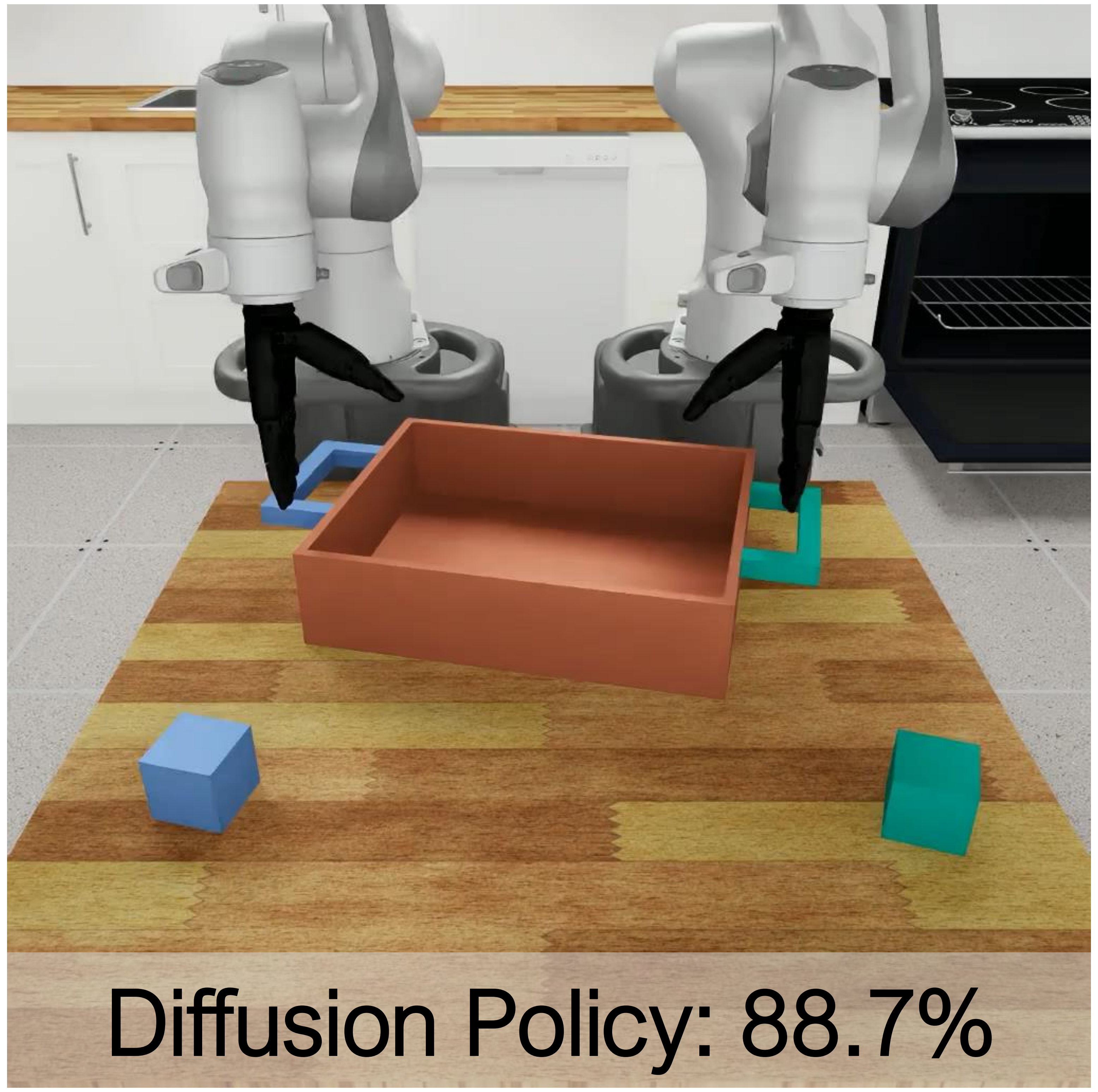


Success rates of Diffusion Policy

DexMimicGen data can be used to investigate the effectiveness of different policy learning methods for bimanual dexterous manipulation

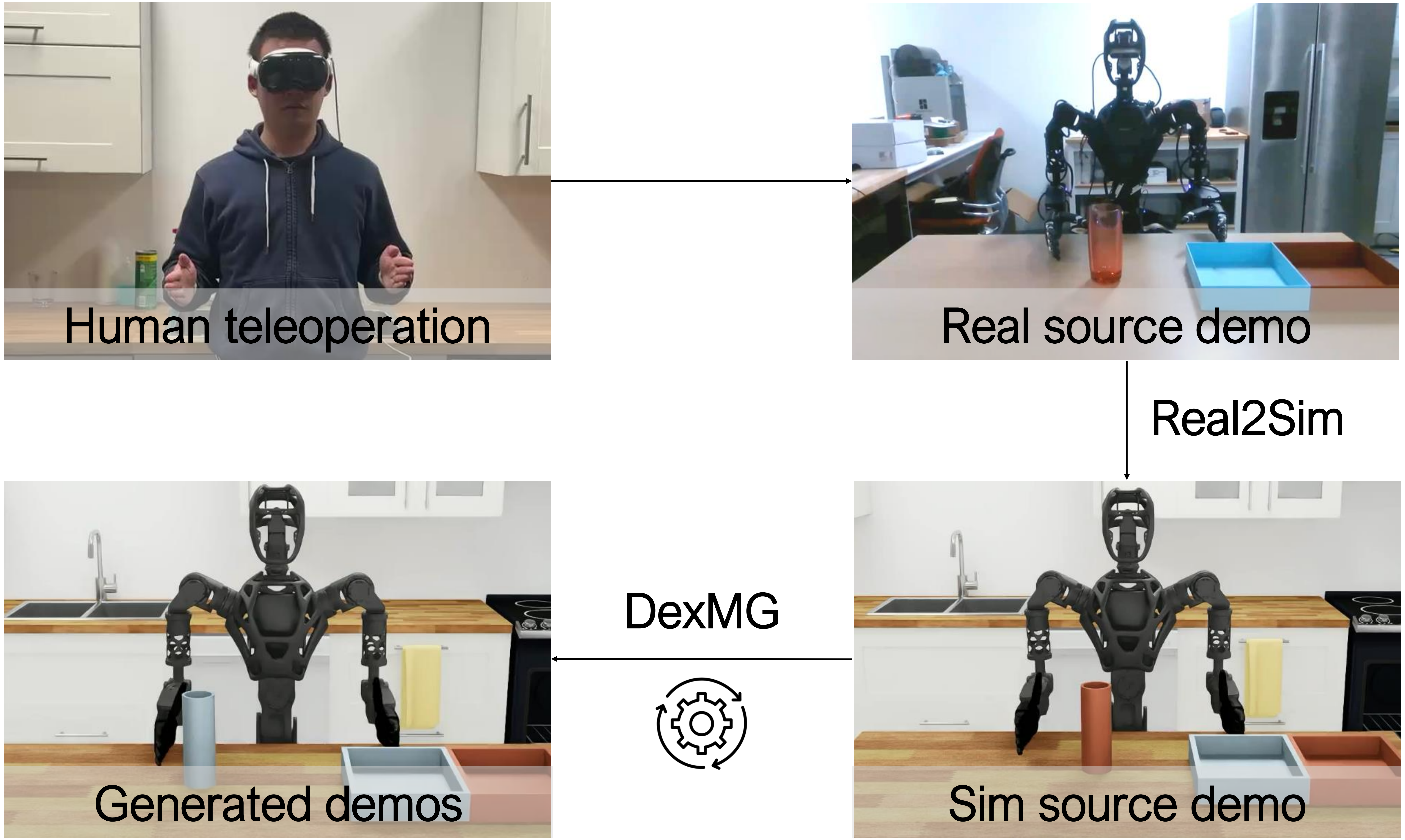






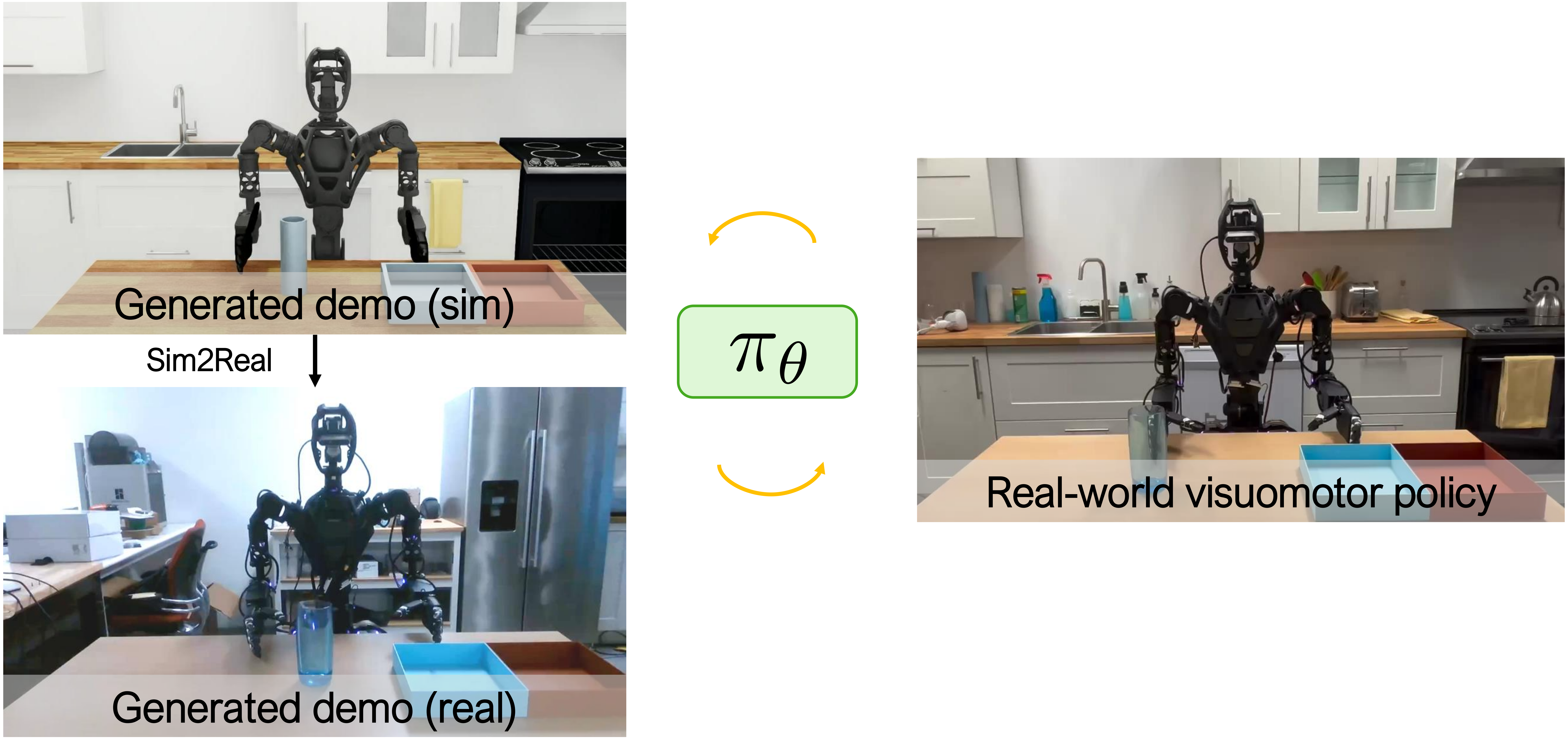
Policy performance improves by generating more DexMimicGen data

DexMimicGen can be used to train real-world visuomotor policies



Transfer real demo to sim using digital twin to ensure the sim demos are valid in real

DexMimicGen can be used to train real-world visuomotor policies



Transfer only successful generated demos from sim to real to train a visuomotor policy

DexMimicGen can be used to train real-world visuomotor policies

Real-world visuomotor policy rollouts (10X)

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations

OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)

MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen
(ICRA 2025)

SkillMimicGen
(CoRL 2024)

AHA
(ICLR 2025)

Simple Recipe for Skill Learning: Scale Data to Scale Performance

Brohan et al. “RT-1: Robotics Transformer for Real-World Control at Scale”, 2022

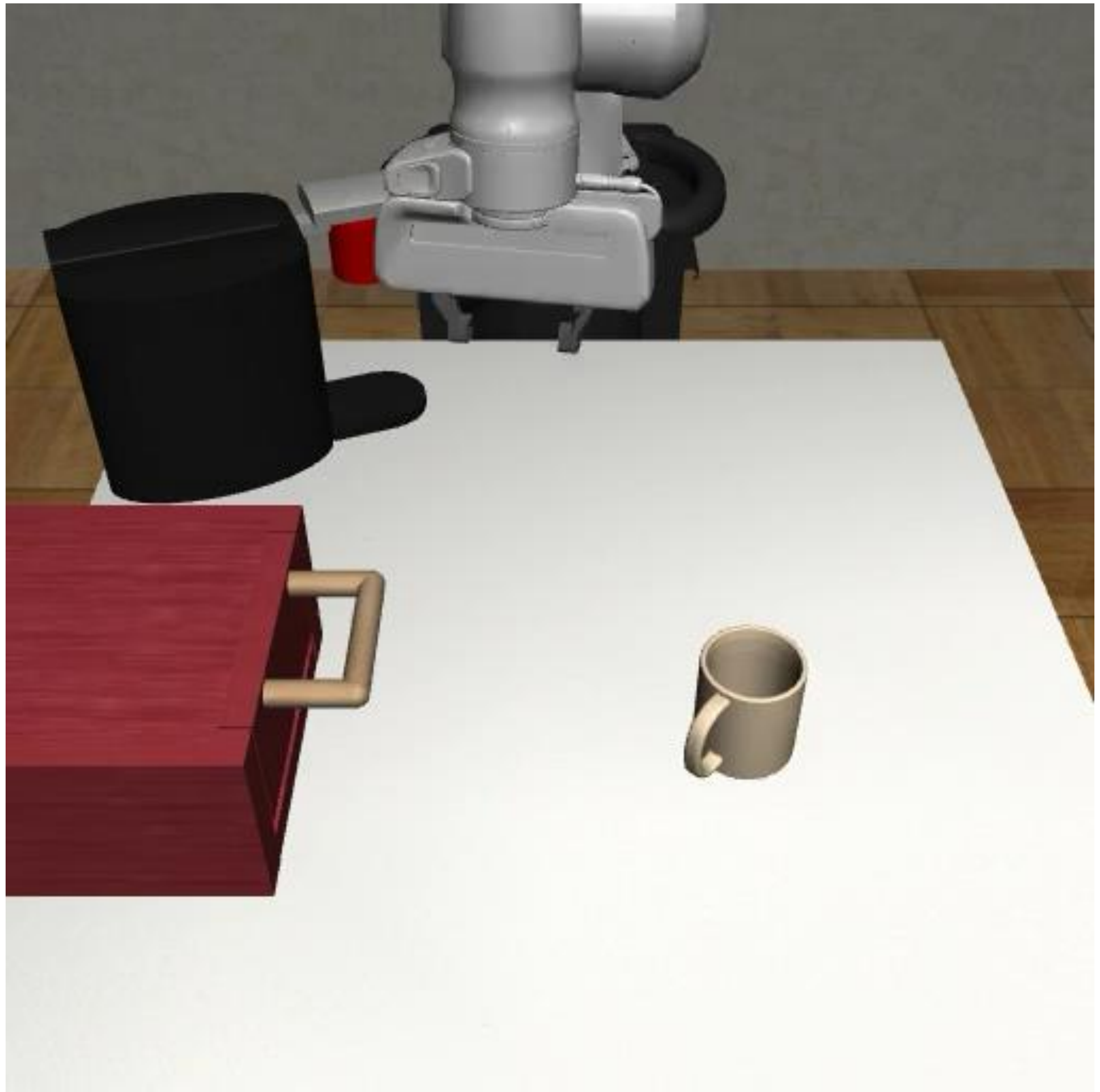
18 months of data with a large team of human contractors and robots to achieve 97% success on rearrangement tasks

Simple Recipe for Skill Learning: Scale Data to Scale Performance

Brohan et al. “RT-1: Robotics Transformer for Real-World Control at Scale”, 2022

18 months of data with a large team of human contractors and robots to achieve 97% success on rearrangement tasks

MimicGen can struggle with Long-Horizon Manipulation



- Data Generation Problems
  - Errors from replay increase over time
  - Naïve linear interpolation for stitching segments together can cause issues
- Imitation Learning Problems
  - Policy must learn full long-horizon task, easy to fall off distribution
  - Interpolation segments can be hard to imitate

Task and Motion Planning excels at Long-Horizon Manipulation

Use world models and optimization to solve long-term objectives

Idea: automate most parts of long-horizon manipulation with planning, and focus learning effort on local contact-rich manipulation

HITL-TAMP: Long-Horizon Manipulation with Planning and Imitation

Mandlekar and Garrett et al. “Human-in-the-Loop Task and Motion Planning for Imitation Learning”, CoRL 2023

HITL-TAMP: Long-Horizon Manipulation with Planning and Imitation

50 demos + HITL-TAMP → 64%

200 demos → 3%

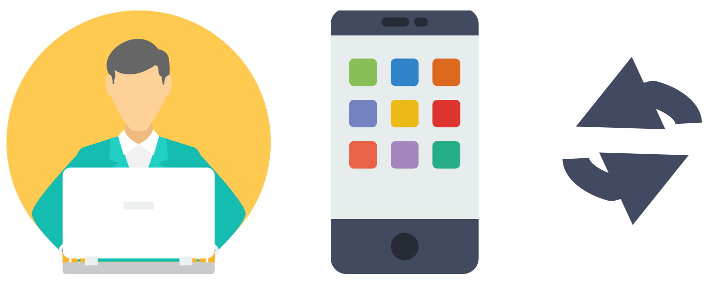
Mandlekar and Garrett et al. “Human-in-the-Loop Task and Motion Planning for Imitation Learning”, CoRL 2023

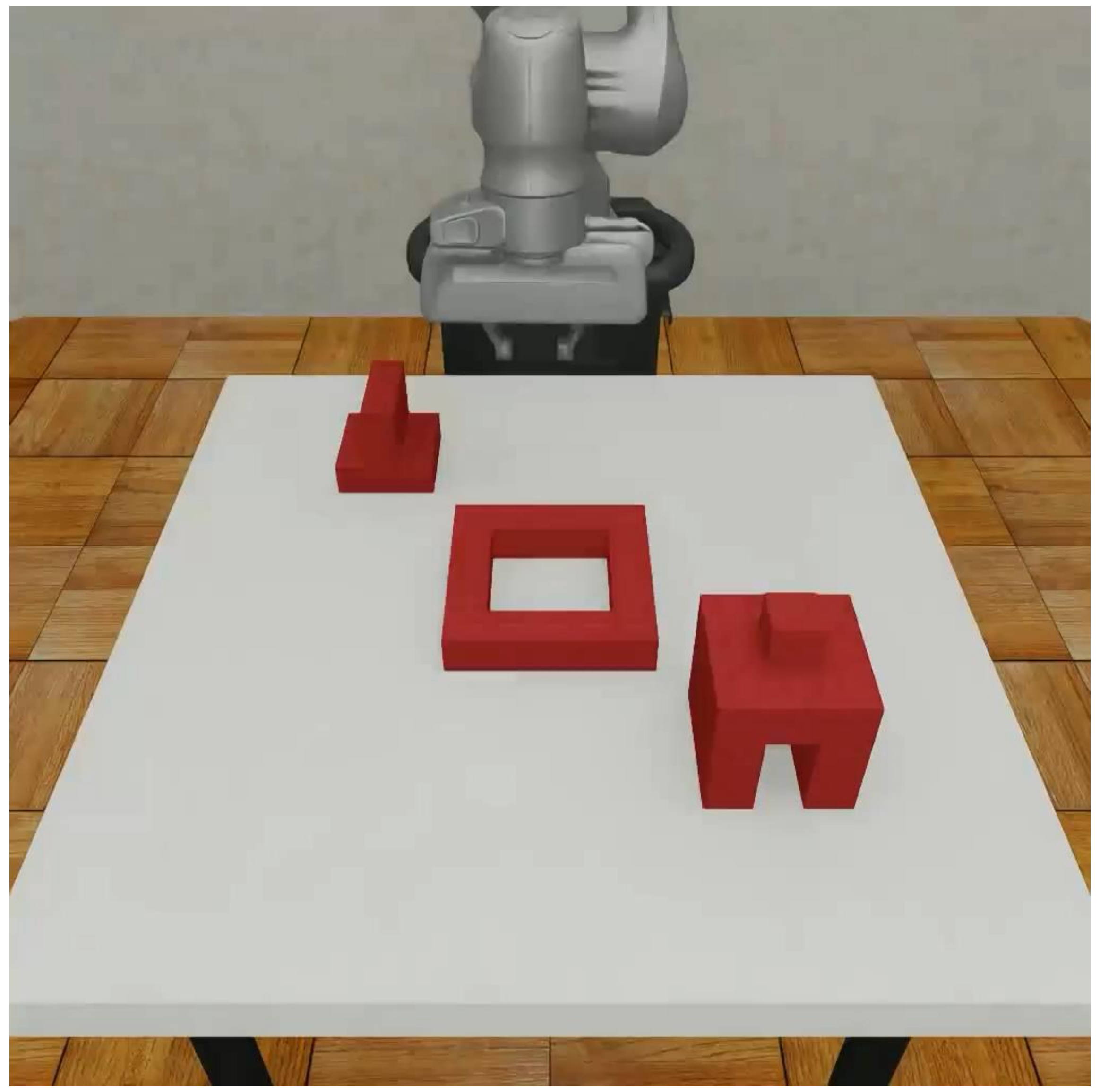
SkillMimicGen

a skill-based data generation system for
robot learning from demonstration

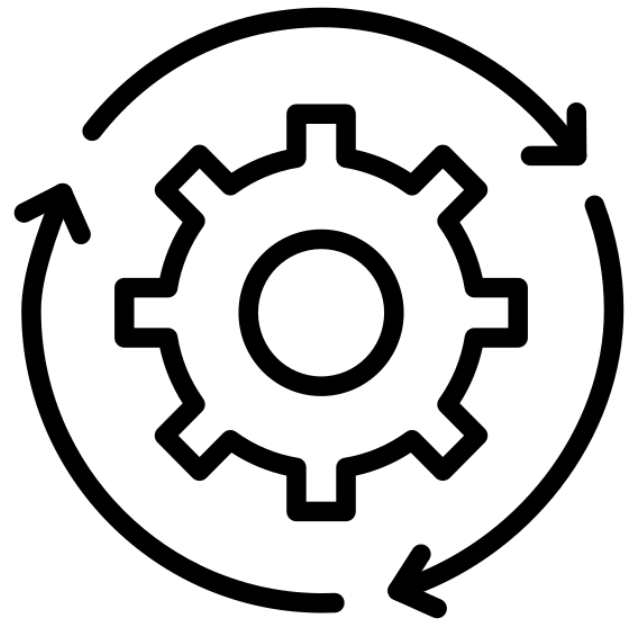
SkillMimicGen (SkillGen) automatically generates large imitation learning datasets from a few human demos

10 human demos

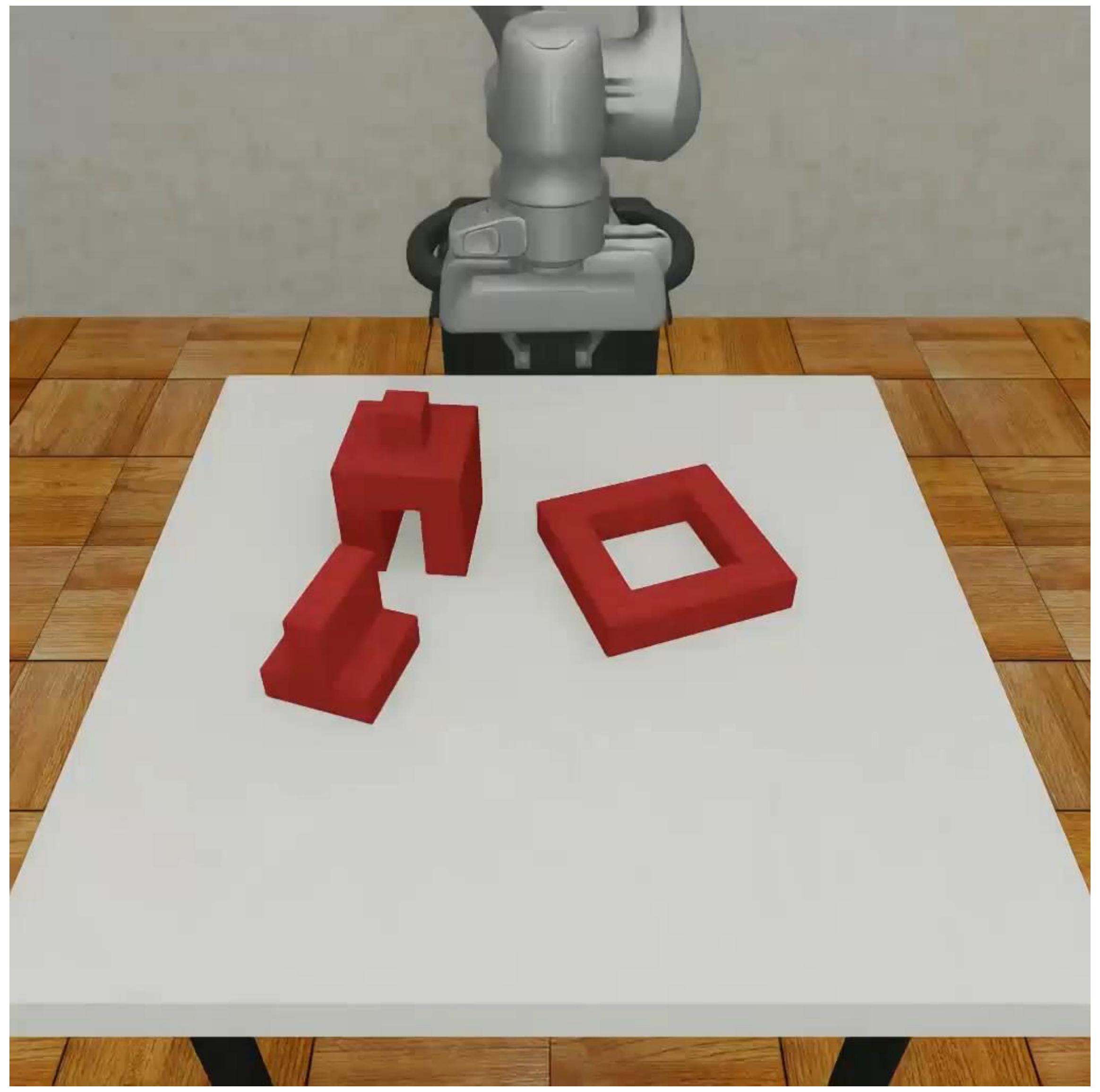




SkillMimicGen



>1000 generated demos



Human collects a handful of teleoperated demos

SkillGen generates more demos through skill adaptation and motion planning

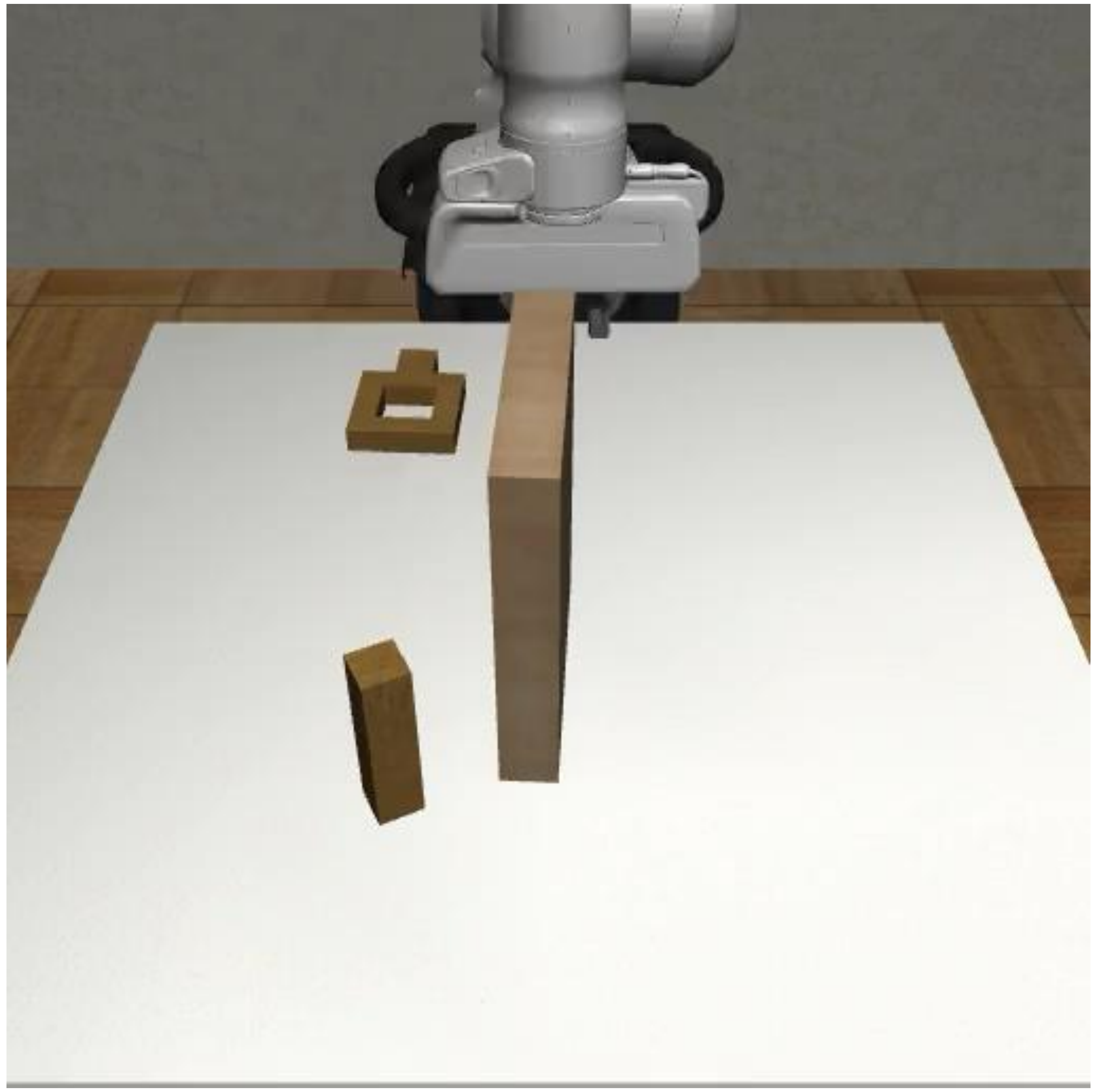
Garrett and Mandlekar et al. “SkillMimicGen: Automated Data Generation for Efficient Skill Learning and Deployment”, CoRL 2024

SkillGen automates dataset generation by interleaving motion planning and contact-rich demonstration adaptation

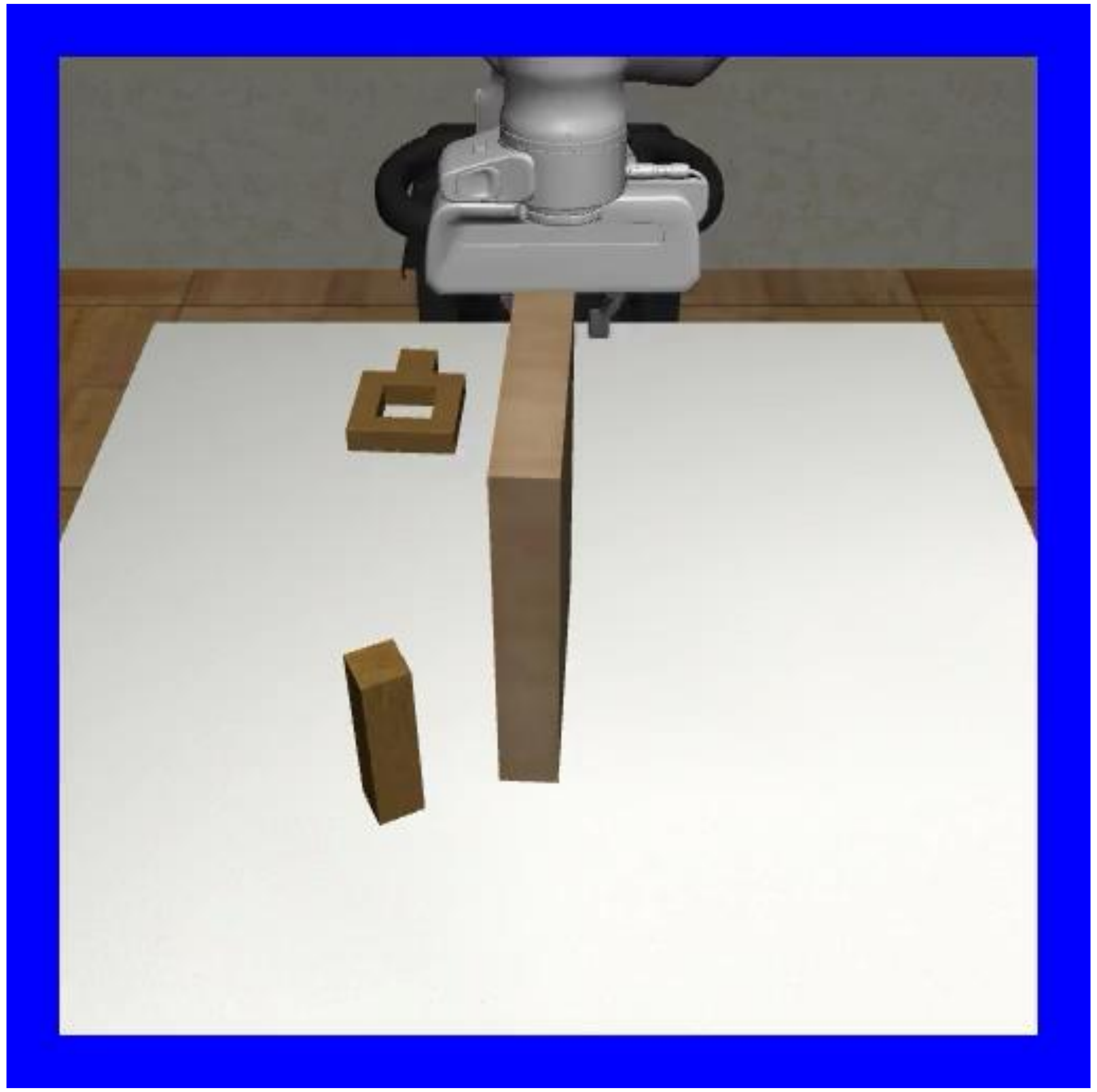
SkillGen generates diverse datasets with large pose variation and is robust to clutter

SkillGen greatly outperforms MimicGen

MimicGen (DGR: 14.5%)

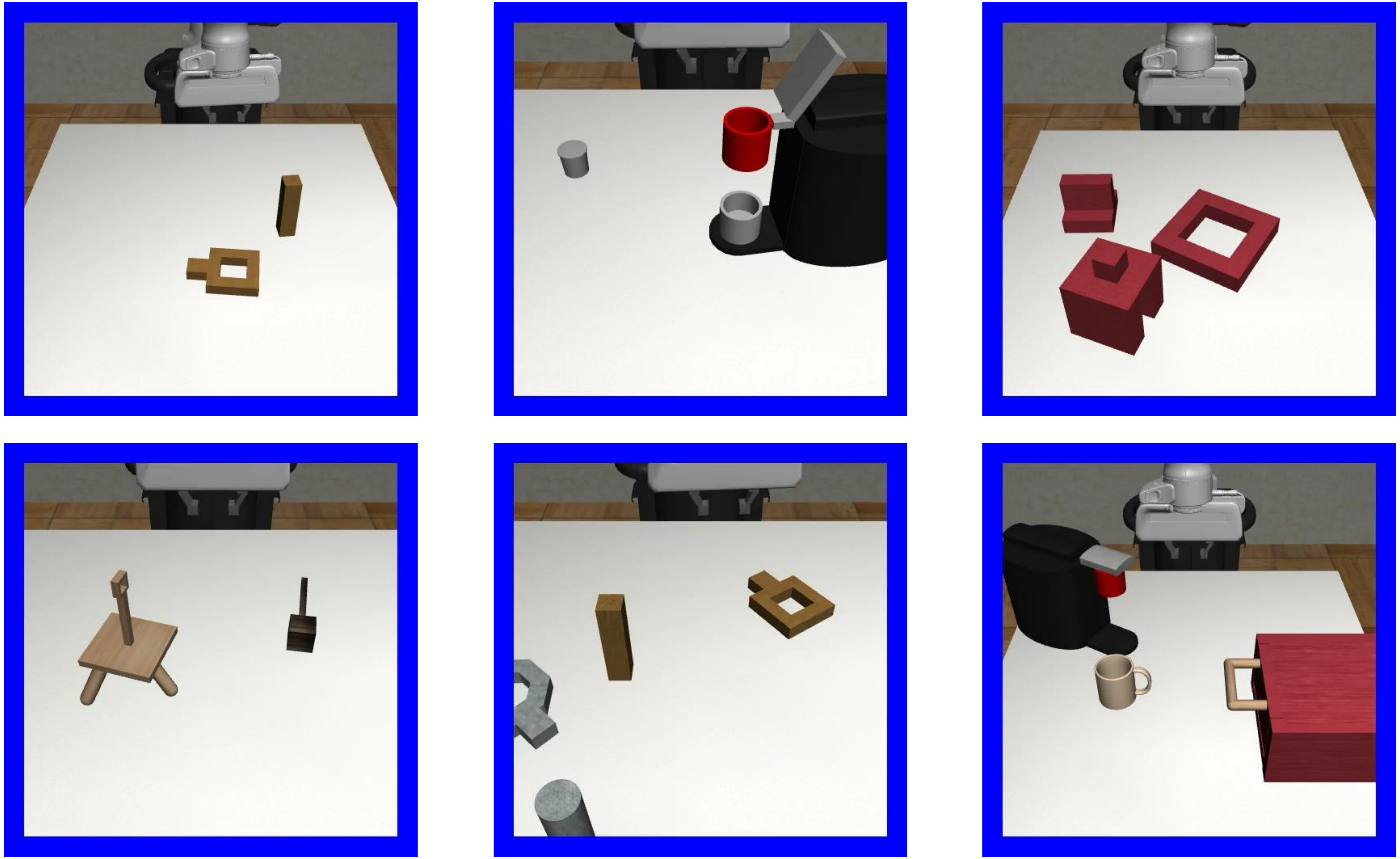


SkillGen (DGR: 72%)



Garrett and Mandlekar et al. “SkillMimicGen: Automated Data Generation for Efficient Skill Learning and Deployment”, CoRL 2024

SkillGen data trains high-performing behavior cloning agents



SkillGen agents are 46% more successful than the SOTA across tasks with broad initializations

SkillGen generation is robust to real-world cluttered environments

SkillGen policies outperform MimicGen in real-world deployment

SkillGen agents are robust to large pose variation

SkillGen enables zero-shot sim-to-real transfer using as little as 1 human demo provided in simulation

Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS: Classical robot planners as data generators
- MimicGen: Data generation using a few human demonstrations

OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCasa: Large-scale simulation framework for mobile manipulation with diverse scenes and tasks
- MimicLabs: A study of how large-scale dataset composition affects imitation learning

RoboCasa (RSS 2024)

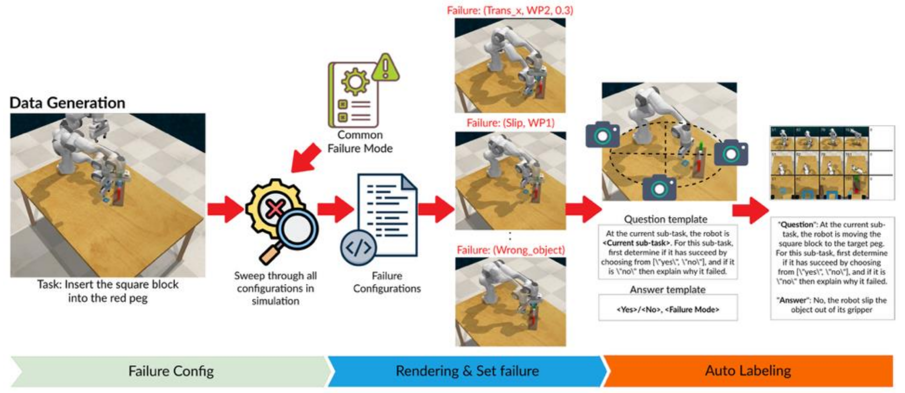
MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and dexterous control
- SkillMimicGen: Combining planning and human demonstrations for data generation
- AHA: A data generator for learning from failures

DexMimicGen
(ICRA 2025)

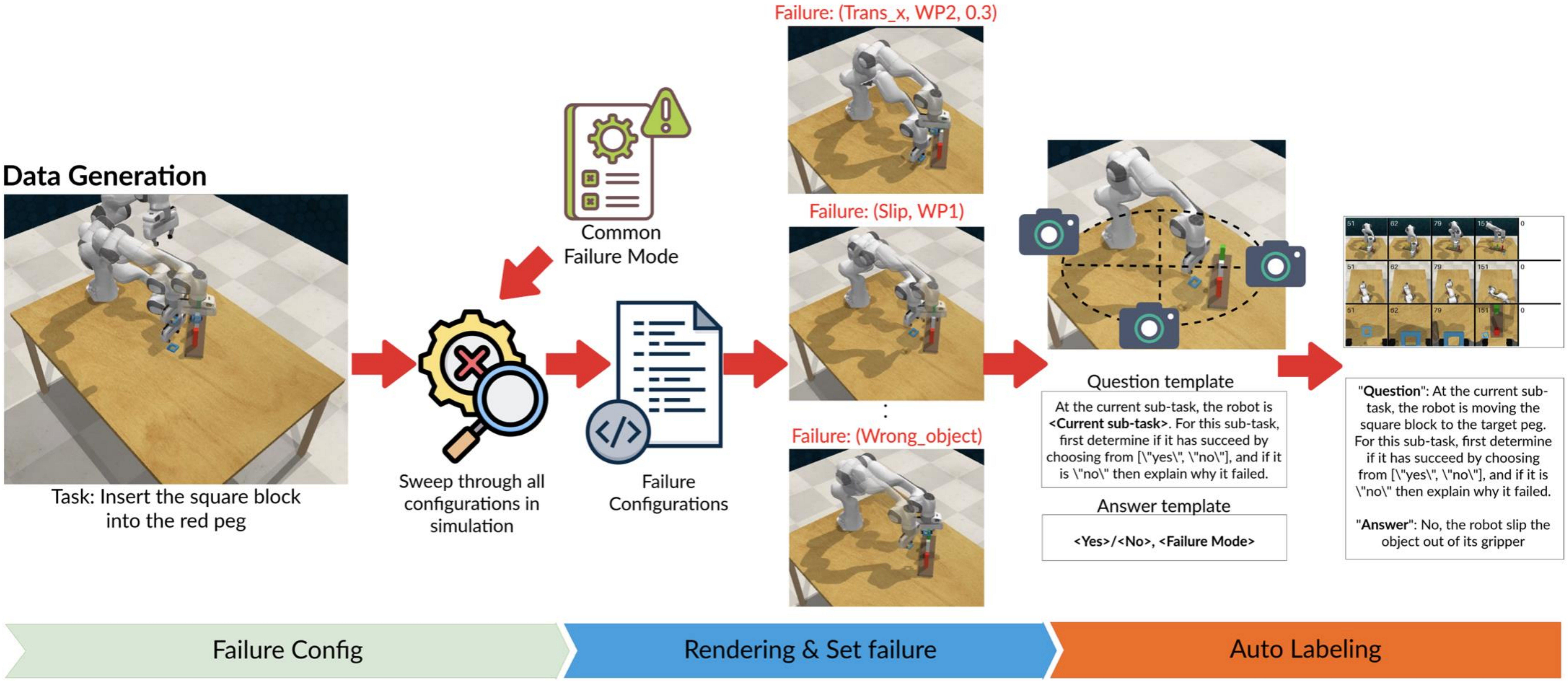
SkillMimicGen
(CoRL 2024)



AHA
(ICLR 2025)

AHA: Synthetic data generation to help robots understand failures

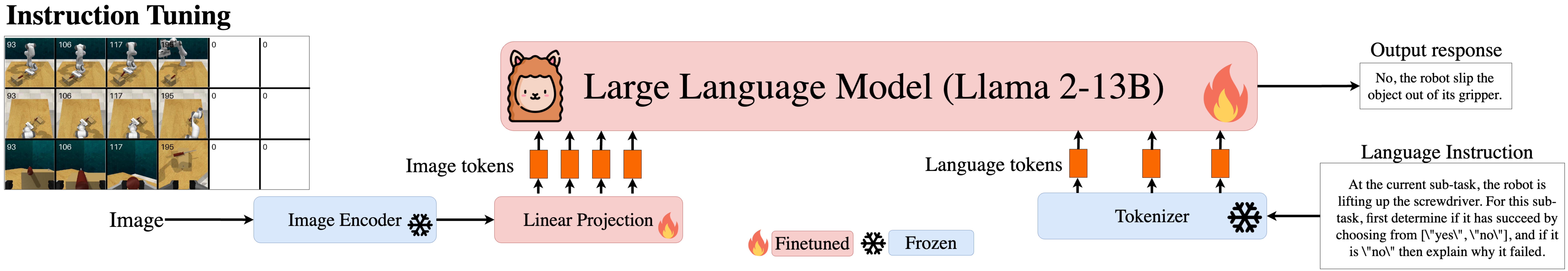
Introduce targeted failures to robot trajectories and auto-label failure explanation



Duan et al. "AHA: A Vision-Language-Model for Detecting and Reasoning over Failures in Robotic Manipulation", ICLR 2025

AHA: Synthetic data generation to help robots understand failures

Finetune an existing VLM on generated robot failure data



Duan et al. “AHA: A Vision-Language-Model for Detecting and Reasoning over Failures in Robotic Manipulation”, ICLR 2025

AHA: Synthetic data generation to help robots understand failures

Apply finetuned VLM to downstream robotic applications

Duan et al. “AHA: A Vision-Language-Model for Detecting and Reasoning over Failures in Robotic Manipulation”, ICLR 2025

Recap: Moving from data collection to data generation

Autonomous Data Generation Tools

- OPTIMUS
- MimicGen
  demonstr

Synthetic data generation is a powerful tool

OPTIMUS (CoRL 2023)

MimicGen (CoRL 2023)

Data Generation Applications

- RoboCas
  manipula
- MimicLab
  composition affects imitation learning

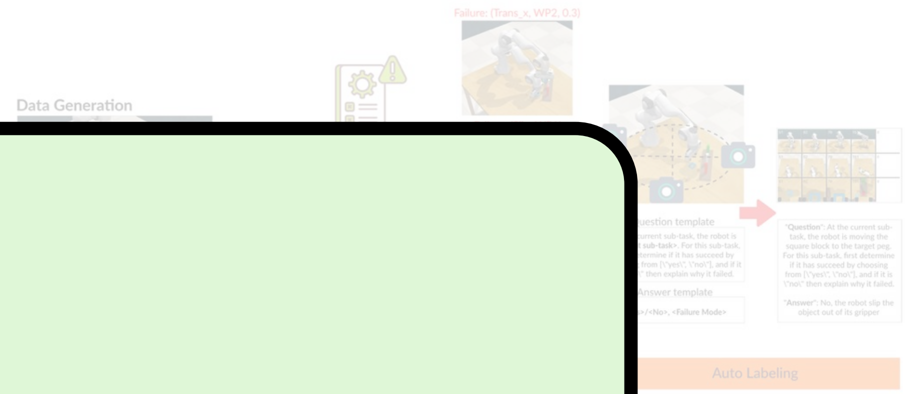
Synthetic data generation tools can enable new applications
that are intractable for human teleoperation-based data collection

RoboCasa (RSS 2024)

MimicLabs (ICLR 2025)

Building More Powerful Data Generators

- DexMimicGen: Data generation for bimanual and
  dexterous
- SkillMimic
  demonstrations for data generation
- AHA: A data generator for learning from failures



Improving synthetic data generation is worthwhile

DexMimicGen
(ICRA 2025)

SkillMimicGen
(CoRL 2024)

AHA
(ICLR 2025)

Thank You!


Dieter Fox


Yashraj Narang


Caelan Garrett


Danfei Xu


Yuke Zhu


Jim Fan


Murtaza Dalal


Ankur Handa


Soroush Nasiriany


Vaibhav Saxena


Matthew Bronars


Nadun Arachchige


Zhenyu Jiang


Yuqi Xie


Kevin Lin


Zhenjia Xu


Weikang Wan


Bowen Wen


Iretiayo Akinola


Jiafei Duan


Ranjay Krishna

Moving from Data Collection to Data Generation: Addressing the Need for Data in Robotics

Ajay Mandlekar

NVIDIA AI Research Scientist

amandlekar@nvidia.com