MimicLabs

What Matters in Learning from Large-Scale Datasets for Robot Manipulation

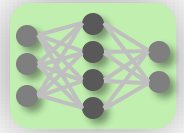
ICLR 2025

Vaibhav Saxena, Matthew Bronars*, Nadun Ranawaka*, Kuancheng Wang, Woo Chul Shin, Soroush Nasiriany, Ajay Mandlekar†, Danfei Xu†

*equal contribution †equal advising

Large-Scale Multi-Task Datasets play a critical role in Robot Learning

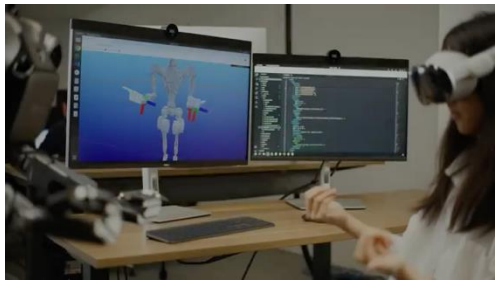
Tesla



Physical Intelligence

1X

NVIDIA Project GR00T



Large-Scale Multi-Task Datasets play a critical role in Robot Learning

But they take months of effort to collect, and we don't even understand **what kind of data matters!**

How can we know…

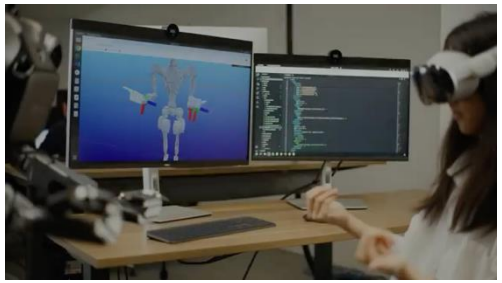
1. where to focus when **collecting** large-scale datasets?
2. how to **use existing large-scale datasets**?

Physical Intelligence

Tesla

1X

NVIDIA Project GR00T



To this end, we present…

MimicLabs

An end-to-end system to
build your own robotics datasets
in simulation

MimicLabs Dataset

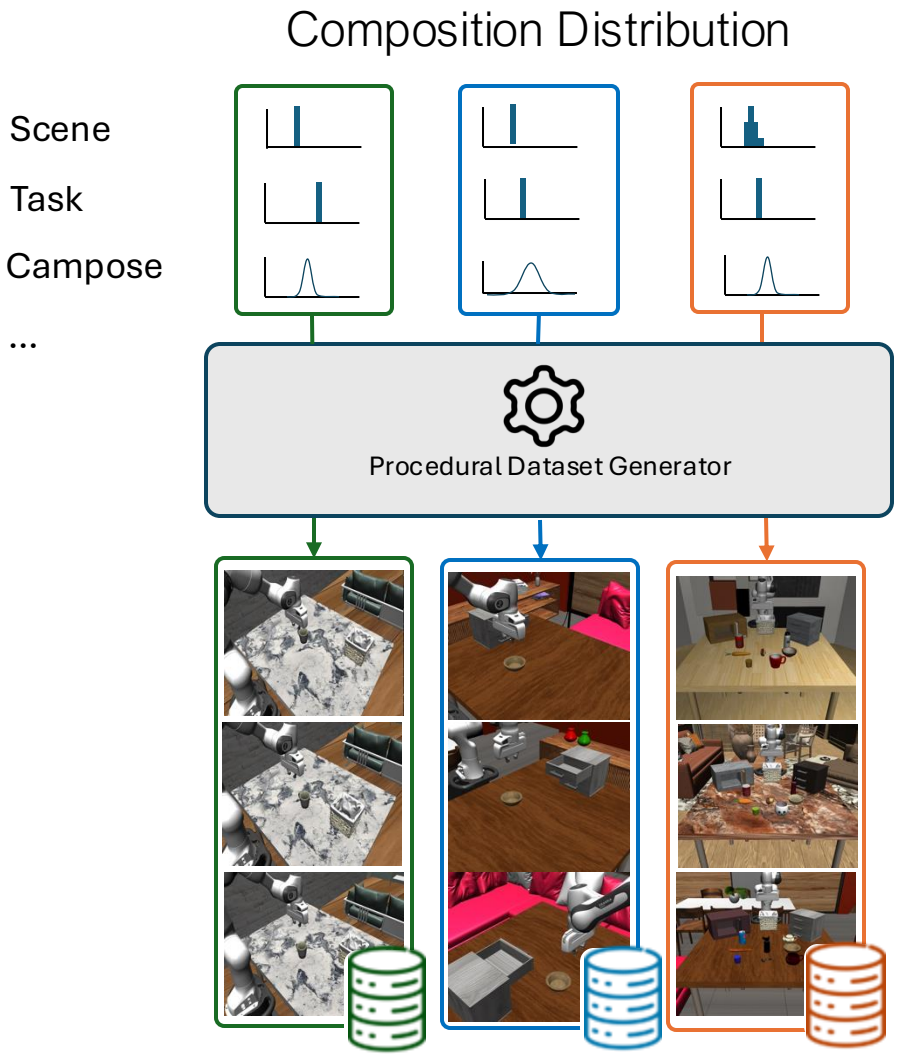
A **large-scale dataset** for doing
systematic studies in robot learning

MimicLabs Study

A **data-composition study** around
large-scale data collection and retrieval


MimicLabs
Build your own robotics datasets.

Synthetic Dataset Generation with Controlled Composition



Describe your task as a BDDL

```
(define (problem MimicLabs_Lab3_Tabletop_Manipulation)
  (:domain robosuite)
  (:language put the mug in the bin)
  (:regions
    (basket_init_region
        (:target table)
        (:ranges (
            (-0.01 0.39 0.01 0.41)
          )
        )
        (:yaw_rotation (
            (3.141592653589793 3.141592653589793)
          )
        )
    )
    (contain_region
        (:target basket_1)
    )
    (object_init_region
        (:target table)
        (:ranges (
            (-0.1 -0.2 0.1 0.0)
            (-0.1 0.0 0.1 0.2)
            (-0.3 0.0 -0.1 0.2)
          )
        )
        (:yaw_rotation (
            (0.0 0.0)
            (0.0 0.0)
            (0.0 0.0)
          )
        )
    )
  )

  (:camera
    (:ranges (
        (1.2 45 -45 1.4 67.5 0)
        (1.2 45 0 1.4 67.5 45)
        (1.2 67.5 0 1.4 90 45)
      )
    )
    (:unit degrees)
    (:jitter_mode uniform)
  )

  (:fixtures
    table - table
  )

  (:objects
    basket_1 - basket
    object_1 - robocasa_mug_3
  )

  (:obj_of_interest
    basket_1
    object_1
  )

  (:init
    (On object_1 table_object_init_region)
    (On basket_1 table_basket_init_region)
  )

  (:goal
    (In object_1 basket_1_contain_region)
  )

  (:demonstration
    (Grasp object_1)
    (In object_1 basket_1_contain_region)
  )

)
```

A single config to describe your task

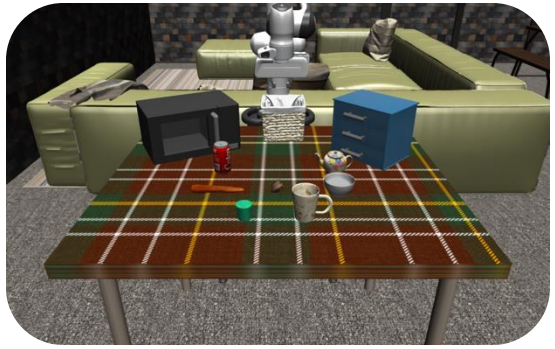
- Scene
- Object and receptacle placements
- Camera poses
- Initial and goal predicates
- Demonstration predicates

Previously used in multiple open-source projects including **BEHAVIOR** and **LIBERO**

Choose among 8 scenes








Control individual dimensions of variation in your task

Base variation

Camera pose

Table texture

Object texture

Spatial Arrangements

Many use cases: Test robustness to geometric changes

Train on manipulating a certain object geometry

Test on novel objects

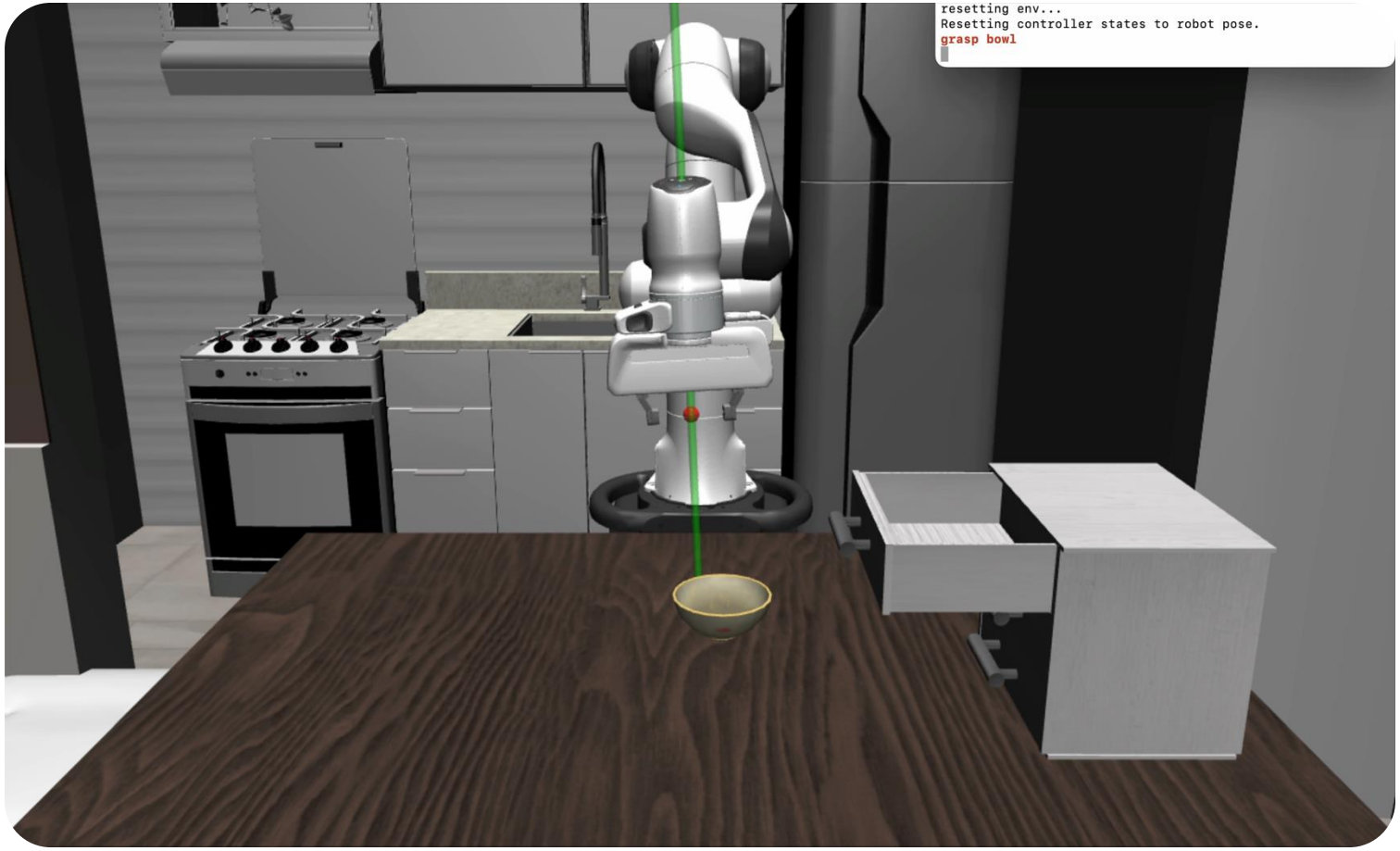
Many use cases: swap out a distractor

Train on certain distractors

Add new distractors during testing

Data Collection with Subgoal Checking

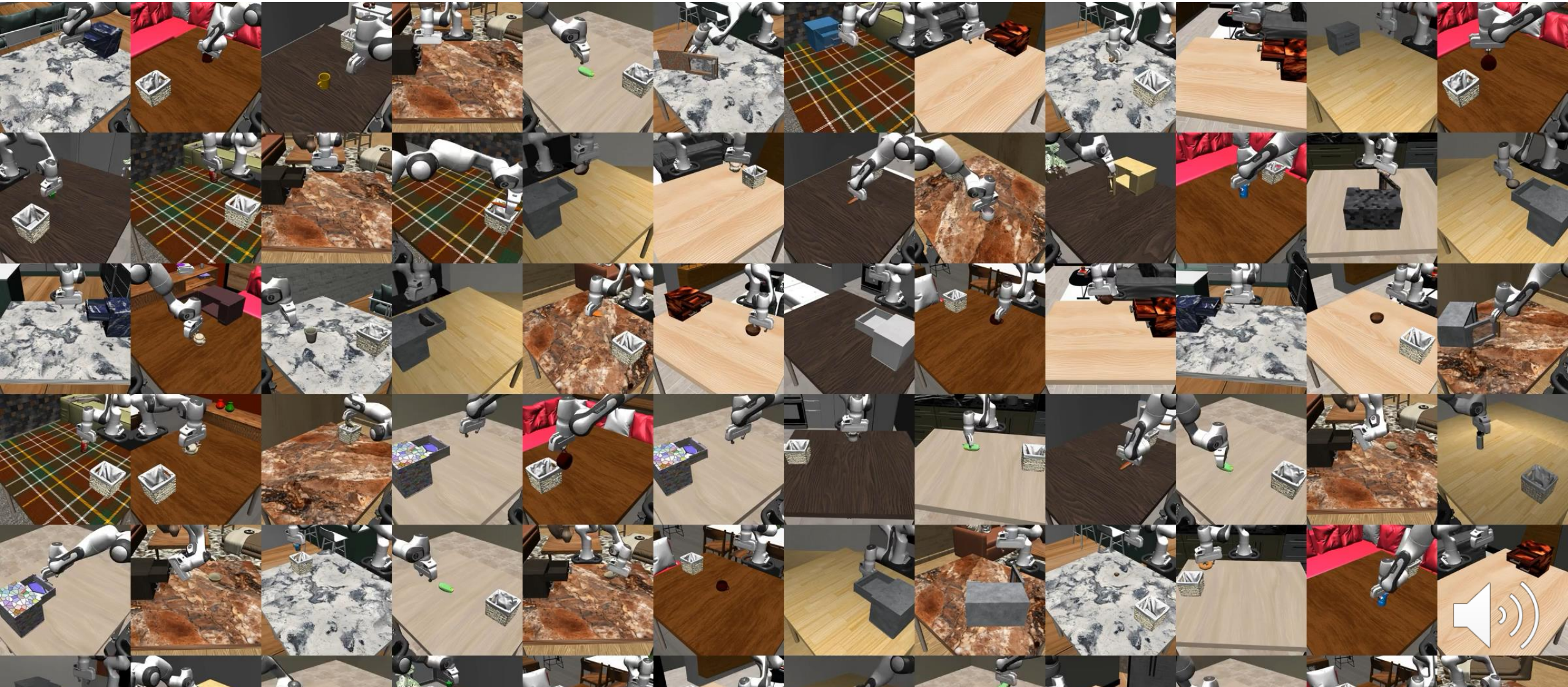
Collect expert data using your device of choice



Meta Quest

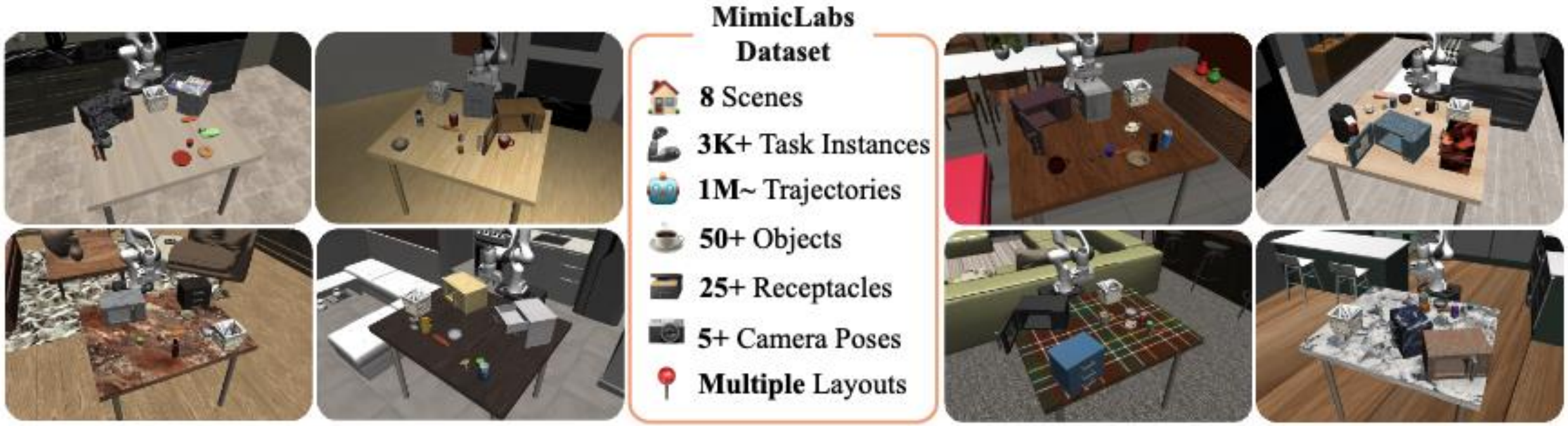
Space Mouse

Integration with MimicGen enables large-scale data generation



MimicLabs Dataset

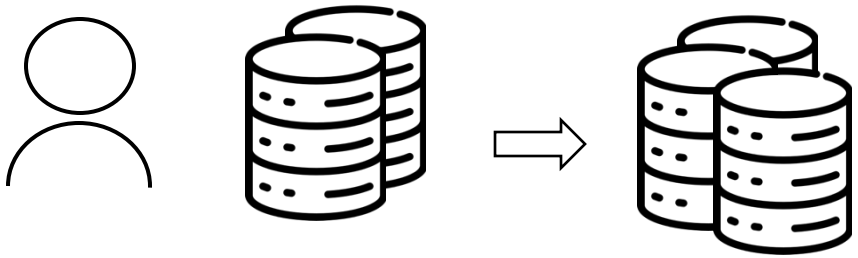
A large simulation dataset constructed to study a variety of problems in robot learning!



MimicLabs Study

Collector's perspective

What data should I **collect** next to improve the dataset utility?



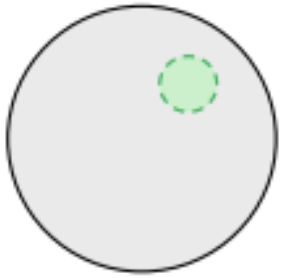
Retriever's perspective

What data should I **retrieve** from a foundation dataset to improve my task performance?

MimicLabs Study

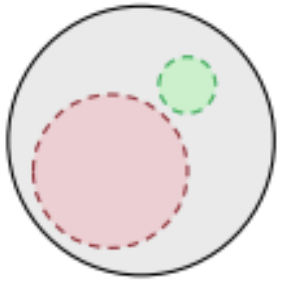
Collector's perspective

Analyzing one dimension at a time



Choose a target distribution

Collect little data *outside* the target distribution → test if co-training helps

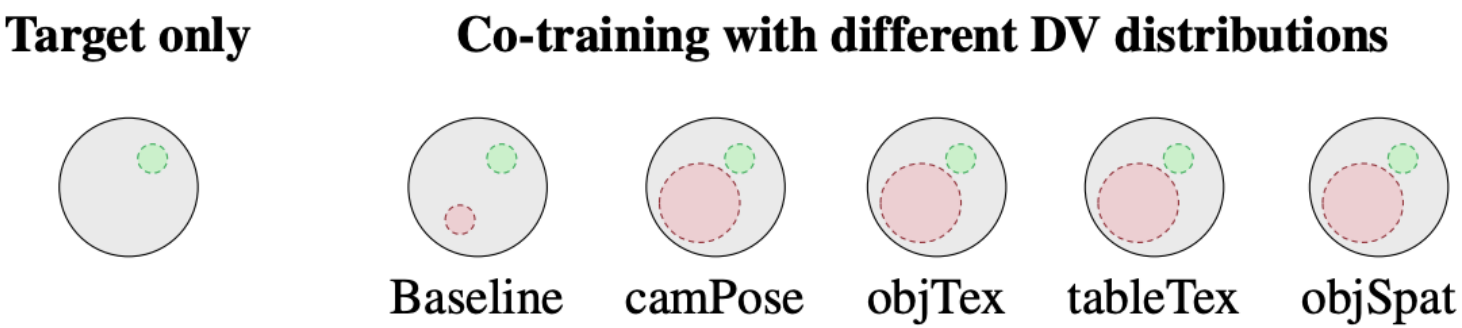


If not, test if increasing variation helps

MimicLabs Study

Collector's perspective

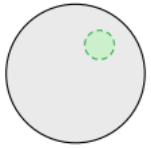
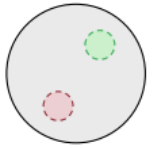
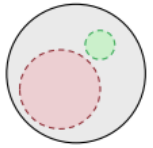
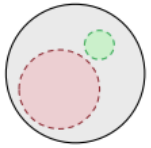
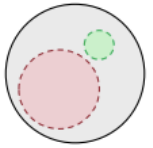
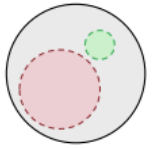
Construct multiple co-training distributions for each target DV



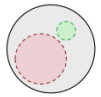
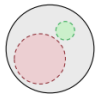
MimicLabs Study

Collector's perspective

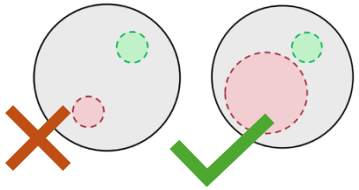
Test which data variation helps
with target-cotrain misalignment

DV w/ target-cotrain misalignment	**Target only**	**Co-training with different DV distributions**				
						
		Baseline	camPose	objTex	tableTex	objSpat
camPose	16.67	43.33	**90**	43.33	43.33	30
objTex	30	**93.33**	**96.67**	**90**	**90**	**93.33**
tableTex	43.33	66.67	**80**	63.33	**83.33**	70
objSpat	10	26.67	46.67	33.33	26.67	**56.67**

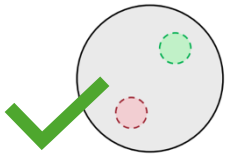
MimicLabs Study

DV w/ target-cotrain misalignment	Target only	Co-training with different DV distributions				
		 Baseline	 camPose	 objTex	 tableTex	 objSpat
camPose	16.67	43.33	**90**	43.33	43.33	30
objTex	30	**93.33**	**96.67**	**90**	**90**	**93.33**
tableTex	43.33	66.67	**80**	63.33	**83.33**	70
objSpat	10	26.67	46.67	33.33	26.67	**56.67**

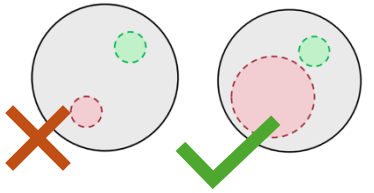
Collector's perspective



Less diverse and misaligned **camera poses** prevent skill transfer

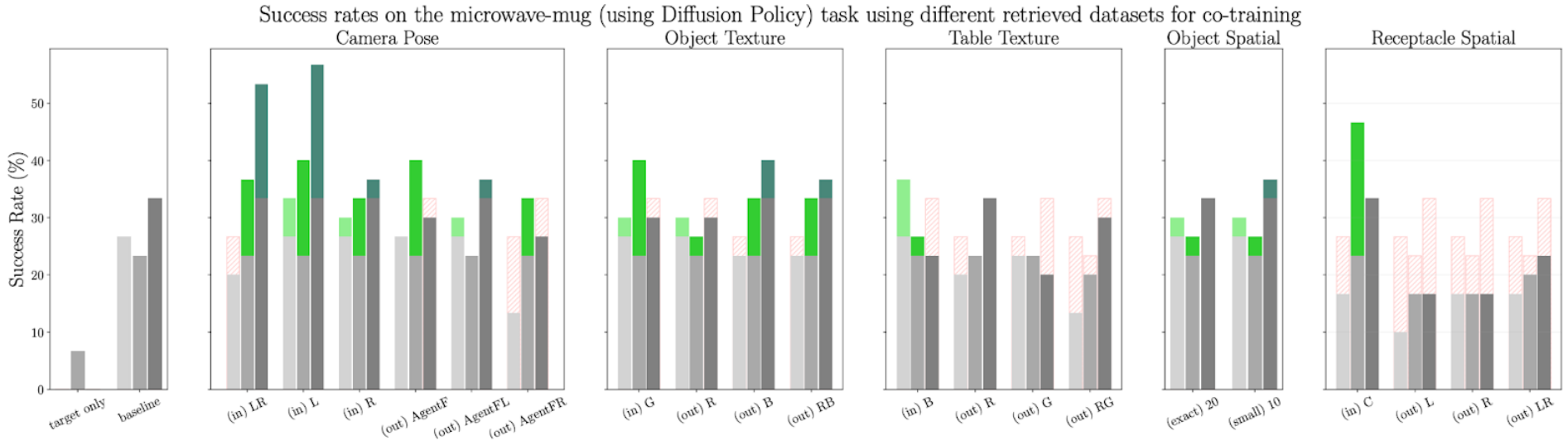


Co-training with **minimal variation in object textures is sufficient** for downstream success

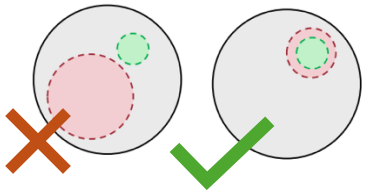


Spatial coverage is critical to downstream success!

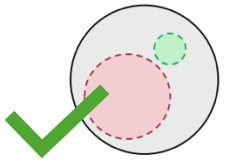
MimicLabs Study



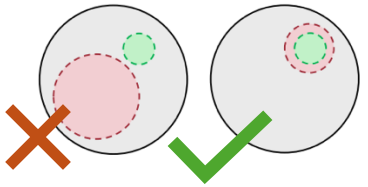
Retriever's perspective



Aligning **camera poses** enables better transfer of skills.



Object texture alignments have limited impact on performance after retrieval.



Spatial alignment is **critical** to downstream success!

Checkout our study paper!

Published as a conference paper at ICLR 2025

WHAT MATTERS IN LEARNING FROM LARGE-SCALE DATASETS FOR ROBOT MANIPULATION

Vaibhav Saxena[1], Matthew Bronars[1,*], Nadun Ranawaka Arachchige[1,*], Kuancheng Wang[1], Woo Chul Shin[1], Soroush Nasiriany[2], Ajay Mandlekar[3], Danfei Xu[1,3]
[1]Georgia Institute of Technology
[2]The University of Texas at Austin
[3]NVIDIA

ABSTRACT

Imitation learning from large multi-task demonstration datasets has emerged as a promising path for building generally-capable robots. As a result, 1000s of hours have been spent on building such large-scale datasets around the globe. Despite the continuous growth of such efforts, we still lack a systematic understanding of what data should be collected to improve the utility of a robotics dataset and facilitate downstream policy learning. In this work, we conduct a large-scale *dataset composition study* to answer this question. We develop a data generation framework to procedurally emulate common sources of diversity in existing datasets (such as sensor placements and object types and arrangements), and use it to generate large-scale robot datasets with controlled compositions, enabling a suite of dataset composition studies that would be prohibitively expensive in the real world. We focus on two practical settings: (1) what types of diversity should be emphasized when future researchers *collect* large-scale datasets for robotics, and (2) how should current practitioners *retrieve* relevant demonstrations from existing datasets to maximize downstream policy performance on tasks of interest. Our study yields several critical insights - for example, we find that camera poses and spatial arrangements are crucial dimensions for both diversity in collection and alignment in retrieval. In real-world robot learning settings, we find that not only do our insights from simulation carry over, but our retrieval strategies on existing datasets such as DROID allow us to consistently outperform existing training strategies by up to 70%. More results at http://robo-mimiclabs.github.io/

1 INTRODUCTION

Imitation learning from offline datasets has emerged as a promising method to teach robots complex real-world manipulation tasks. Importantly, prior works have found that robot performance scales favorably with the dataset size and quality (Mandlekar et al., 2021; Brohan et al., 2022). Consequently, much efforts have been invested in building large-scale robot datasets that cover diverse tasks and environments. Recent large-scale efforts such as DROID (Khazatsky et al., 2024) and the Open X Embodiment datasets (Collaboration et al., 2023) have amassed millions of trajectories for table-top manipulation tasks. These datasets, while still orders of magnitude smaller than their vision and language counterparts, can allow robots trained on this data to generalize to different scenarios.

Despite the continuous growth and promising results of these data collection efforts, we still lack a systematic understanding of *what data should be collected* to improve the utility of a robotics dataset. However, gaining this understanding poses significant challenges. Data collection is extremely costly and time-consuming, often requiring teams of human operators, fleets of robots, and months of manual effort (Brohan et al., 2022), and the cost is compounded by the time and effort it takes to evaluate different robot models trained on these datasets. This makes testing the effectiveness of

[*]Equal contribution, [†]equal advising
Correspondence to vaasena3@gatech.edu

1

What Matters in Learning from Large-Scale Datasets for Robot Manipulation

ICLR 2025

Website: robo-mimiclabs.github.io