# Generalization in Dexterous Manipulation via Geometry-Aware Multi-Task Learning

Wenlong Huang[1], Igor Mordatch[2], Pieter Abbeel[1], Deepak Pathak[3]
[1]UC Berkeley, [2]Google Brain, [3]Carnegie Mellon University



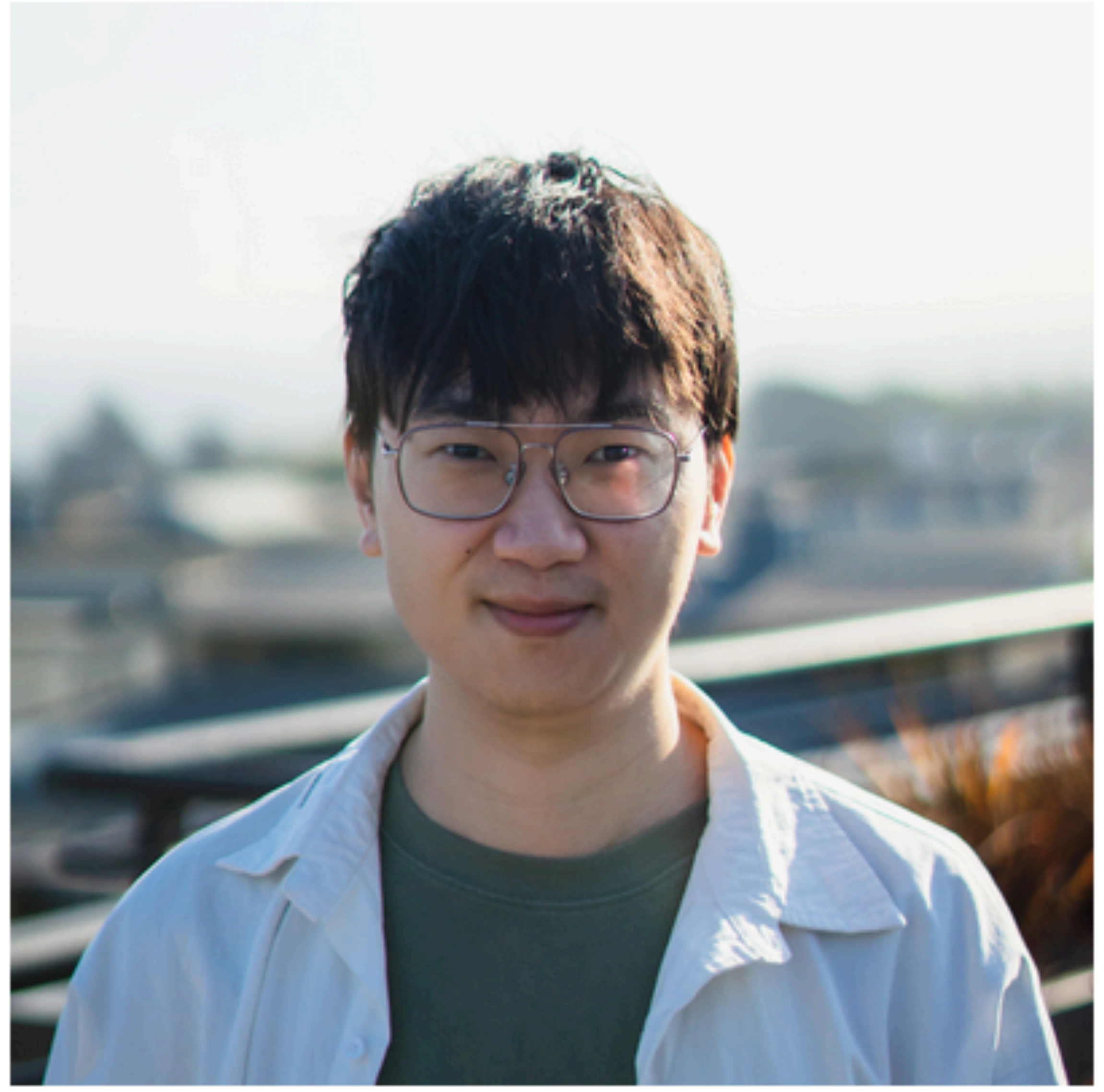
***Abstract*— Dexterous manipulation of arbitrary objects, a fundamental daily task for humans, has been a grand challenge for autonomous robotic systems. Although data-driven approaches using reinforcement learning can develop *specialist* policies that discover behaviors to control a single object, they often exhibit poor generalization to unseen ones. In this work, we show that policies learned by existing reinforcement learning algorithms can in fact be *generalist* when combined with multi-task learning and a well-chosen object representation. We show that a single generalist policy can perform in-hand manipulation of over 100 geometrically-diverse real-world objects and generalize to new objects with unseen shape or size. Interestingly, we find that multi-task learning with object point cloud representations not only generalizes better but even outperforms the single-object specialist policies on both training as well as held-out test objects. Video results at https://huangwl18.github.io/geometry-dex.**

Fig. 1: Our goal in this work is to train a single policy that can perform in-hand manipulation on a large number of objects. We show surprising results that simple multi-task learning combined with appropriate representation not only achieves the aforementioned goal but also outperforms the single-task oracles, on both training and unseen objects.

## I. Introduction

Hand dexterity is fundamental to daily human activities, requiring complex and precise control of finger movements. While most animals exhibit fine-grained motor controls for movement and many show elementary skills for manipula-

# Wenlong Huang



I am a student researcher at Google Brain, co-hosted by Brian Ichter and Karol Hausman.

I recently received my B.A. in Computer Science from UC Berkeley in December 2021, where I was fortunate to be advised by Pieter Abbeel, Deepak Pathak, and Igor Mordatch as part of Berkeley Artificial Intelligence Research (BAIR). Previously, I was also fortunate to work with Zhuowen Tu as a high school researcher at UC San Diego.

In my free time, I love taking photos. You can view some of them here.

Email / Google Scholar / Twitter / GitHub

## News

- [Apr 2022] I will be joining Stanford as a PhD student in fall 2022! Excited for what's lying ahead!

Igor Mordatch
173 Tweets

Follow

**Igor Mordatch**
@IMordatch

Research scientist at Google Brain. Interested in AI/education/visual art/nature. Previously @OpenAI and UC Berkeley.

Oakland, CA Joined April 2020

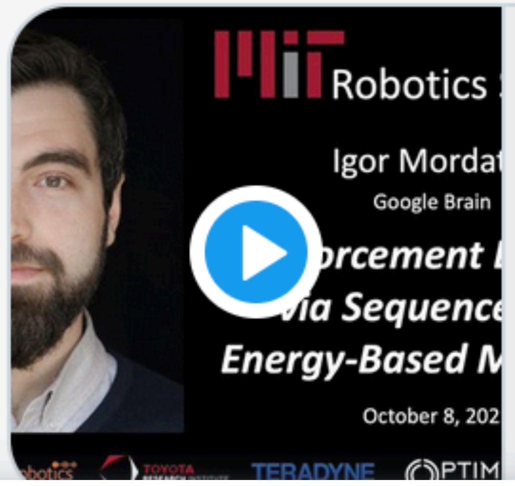
**208** Following **2,607** Followers

**Tweets** Tweets & replies Media Likes

Pinned Tweet

**Igor Mordatch** @IMordatch · Oct 27, 2021

Really enjoyed giving an @MIT Robotics seminar in person last month! Outlines recent work, motivations, and future excitement at the intersection of large sequence models, energy-based models, and robotics




youtube.com
MIT Robotics - Igor Mordatch - Reinforcement Lea...
MIT - October 8, 2021Igor Mordatch"Reinforcement Learning via Sequence and Energy-Based ...

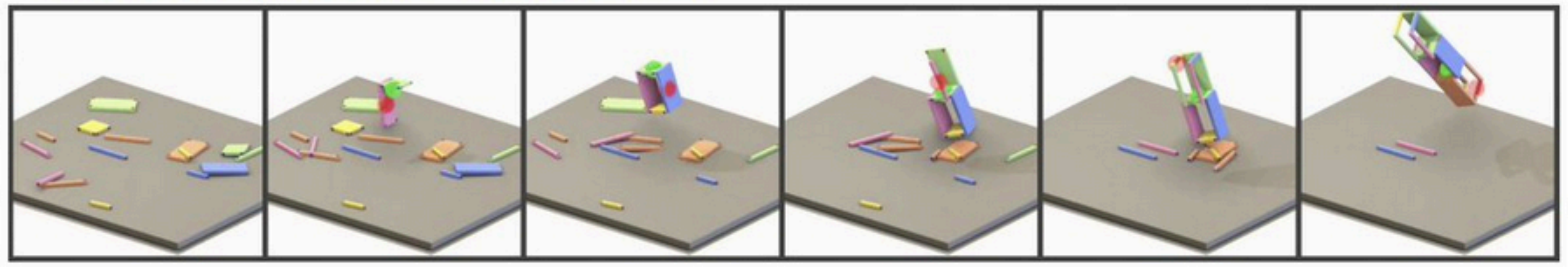
Igor Mordatch Retweeted

**AK** @ak92501 · Mar 27

Blocks Assemble! Learning to Assemble with Large-Scale Structured Reinforcement Learning
abs: arxiv.org/abs/2203.13733

**Abstract**

Assembly of multi-part physical structures is both a valuable end product for autonomous robotics, as well as a valuable diagnostic task for open-ended training of embodied intelligent agents. We introduce a naturalistic physics-based environment with a set of connectable magnet blocks inspired by children's toy kits. The objective is to assemble blocks into a succession of target blueprints. Despite the simplicity of this objective, the compositional nature of building diverse blueprints from a set of blocks leads to an explosion of complexity in structures that agents encounter. Furthermore, assembly stresses agents' multi-step planning, physical reasoning, and bimanual coordination. We find that the combination of large-scale reinforcement learning and graph-based policies – surprisingly without any additional complexity – is an effective recipe for training agents that not only generalize to complex unseen blueprints in a zero-shot manner, but even operate in a reset-free setting without being trained to do so. Through extensive experiments, we highlight the importance of large-scale training, structured representations, contributions of multi-task vs. single-task learning, as well as the effects of curriculums, and discuss qualitative behaviors of trained agents. Our accompanying project webpage can be found at: sites.google.com/view/learning-direct-assembly



15 57

Igor Mordatch Retweeted

**sim2real** @sim2realAIorg · Mar 30

Bi-Manual Manipulation and Attachment via Sim-to-Real Reinforcement Learning

arxiv.org/abs/2203.08277

Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2012)
P. Kry and J. Lee (Editors)

# Contact-Invariant Optimization for Hand Manipulation

Igor Mordatch Zoran Popović Emanuel Todorov

University of Washington

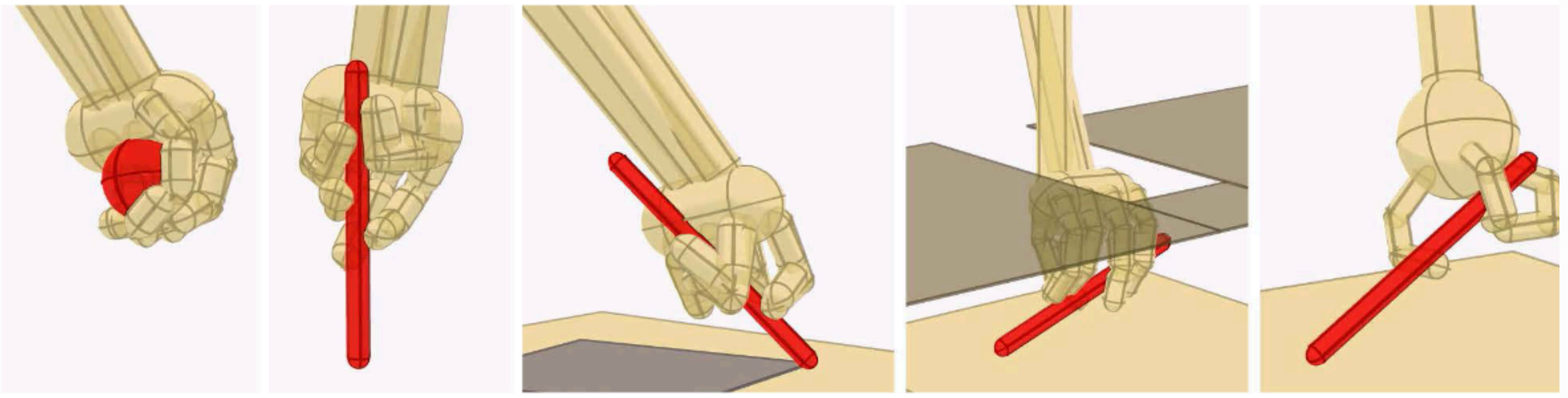
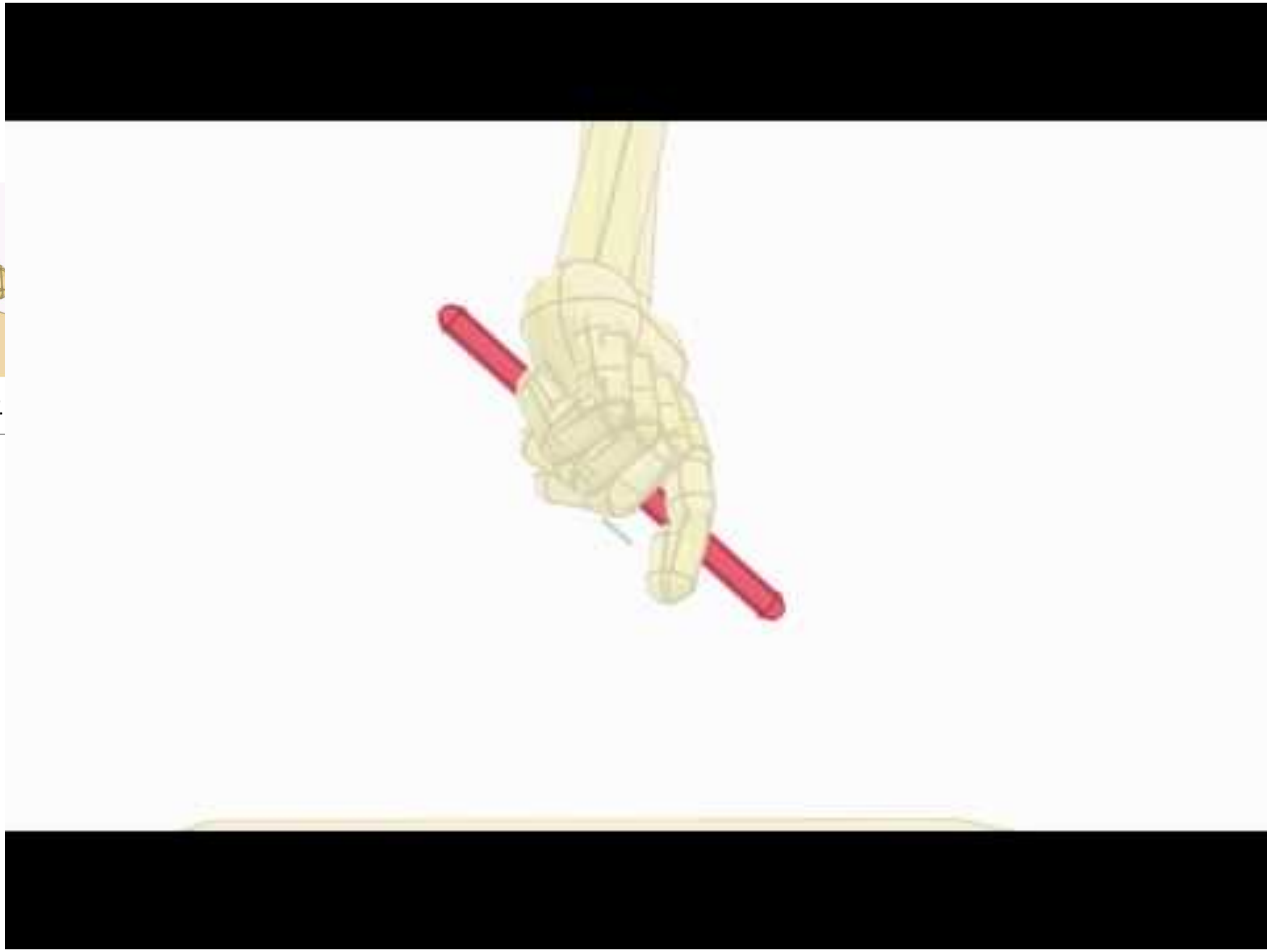


**Figure 1:** *A selection of grasps and motions synthesized by our method.*





# Home

---

**Positions:**
Professor, UC Berkeley
Director of the Berkeley Robot Learning Lab
Co-Director of the Berkeley Artificial Intelligence Research (BAIR) lab
Co-Founder, President, and Chief Scientist Covariant (2017- )
Co-Founder Gradescope (2014-2018: Acquired by TurnItIn)
Host of The Robot Brains Podcast
Founding Investment Partner at AIX Ventures

**Some Personal Background:**
Pretty well covered in this interview by my undergrad school KU Leuven: here

**Contact:**
pabbeel AT cs.berkeley.edu

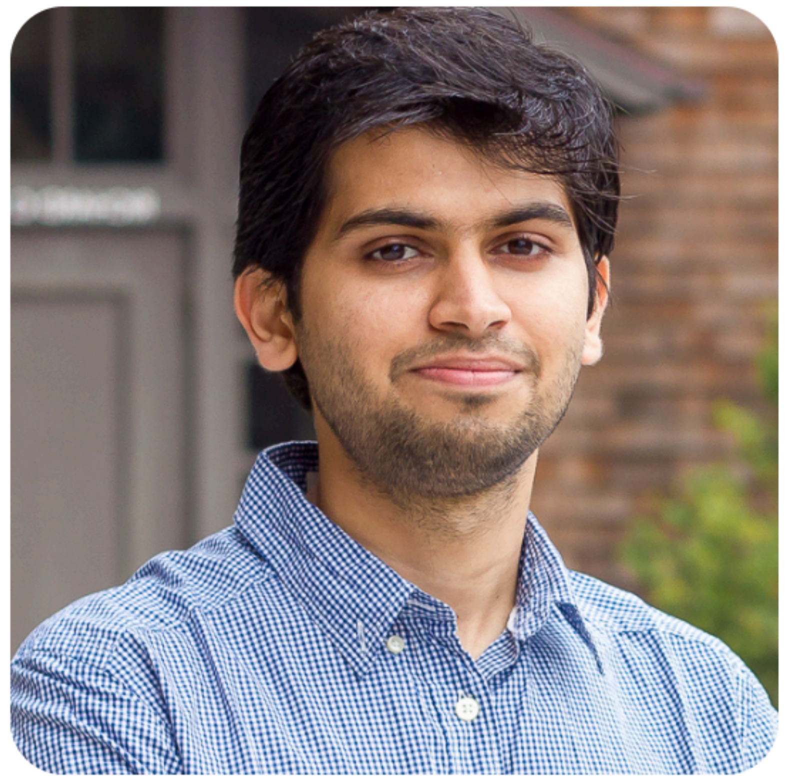
**Quick picks:**

ICML 2021 Tutorial on Unsupervised Reinforcement Learning

CS294-190 Advanced Topics in Learning and Decision Making (co-taught with Stuart Russell)

# Deepak Pathak

**email:** du.sackcaeudm@p.th unscramble



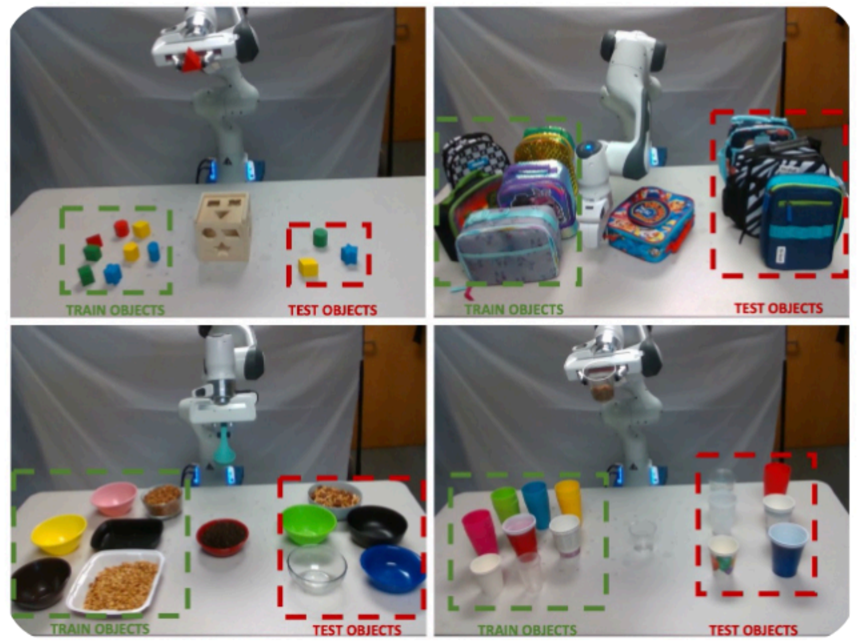
I am an Assistant Professor at Carnegie Mellon University in the School of Computer Science. I am a member of the Robotics Institute and affiliated to Machine Learning Department. I work in Artificial Intelligence at the intersection of Computer Vision, Machine Learning & Robotics.

Previously, I was a researcher at Facebook AI Research and visiting researcher at UC Berkeley with Pieter Abbeel. I received my PhD from UC Berkeley advised by Alyosha Efros and Trevor Darrell. Before that, I completed my Bachelors in Computer Science from IIT Kanpur.

**Prospective students:** Please read this excellent set of FAQs by Ruslan before emailing me as they pretty much apply to joining our group as well.

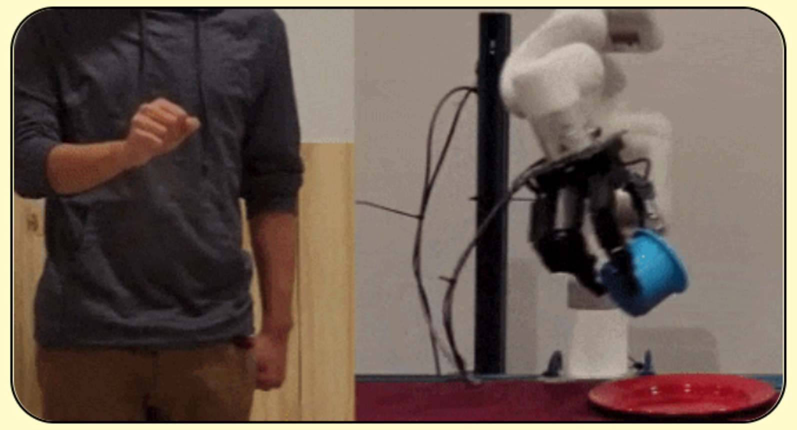
CV | Bio | Google Scholar
Phd Thesis | Github | Twitter

Follow @pathak2206



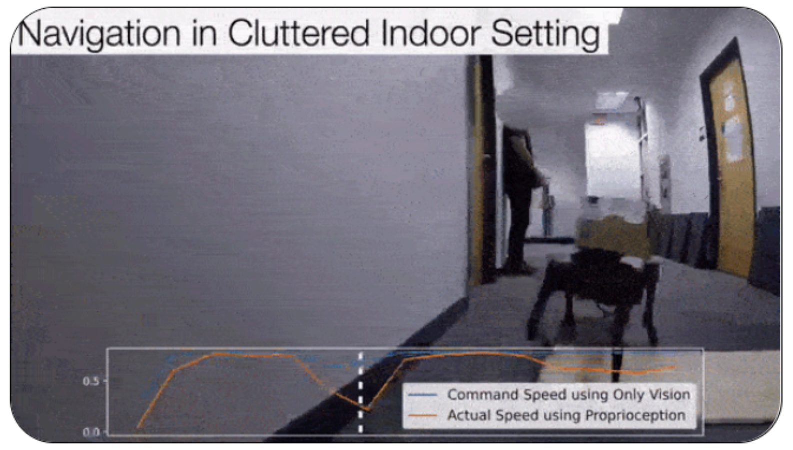
**RB2: Robotic Manipulation Benchmarking with a Twist**
Sudeep Dasari, Jianren Wang, Joyce Hong, Shikhar Bahl, Abitha Thankaraj, Karanbir Chahal, Berk Calli, Saurabh Gupta, David Held, Lerrel Pinto, Deepak Pathak, Vikash Kumar, Abhinav Gupta
NeurIPS 2021
(Datasets and Benchmark)

webpage | pdf | abstract | bibtex | code



**Robotic Telekinesis: Learning a Robotic Hand Imitator by Watching Humans on Youtube**
Aravind Sivakumar, Kenneth Shaw, Deepak Pathak
arXiv 2022

webpage | abstract | bibtex | arXiv | demo video

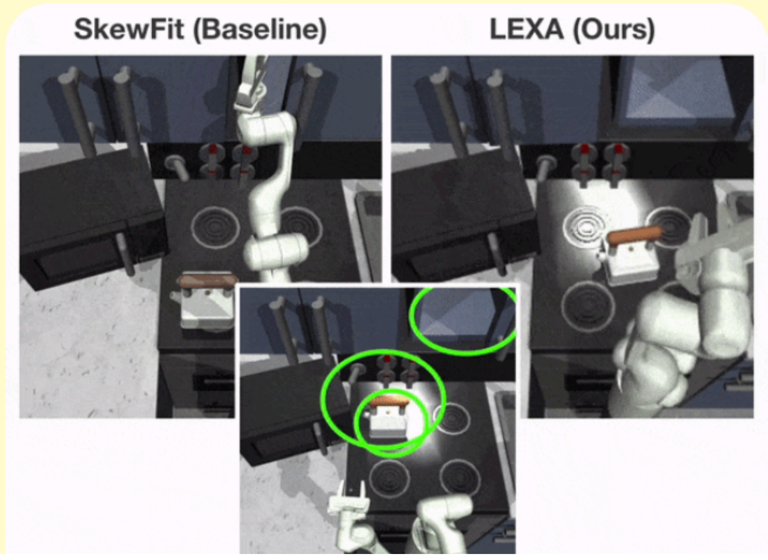



**Coupling Vision and Proprioception for Navigation of Legged Robots**
Zipeng Fu*, Ashish Kumar*, Ananye Agarwal, Haozhi Qi, Jitendra Malik, Deepak Pathak
CVPR 2022

webpage | pdf | abstract | bibtex | arXiv | code | video

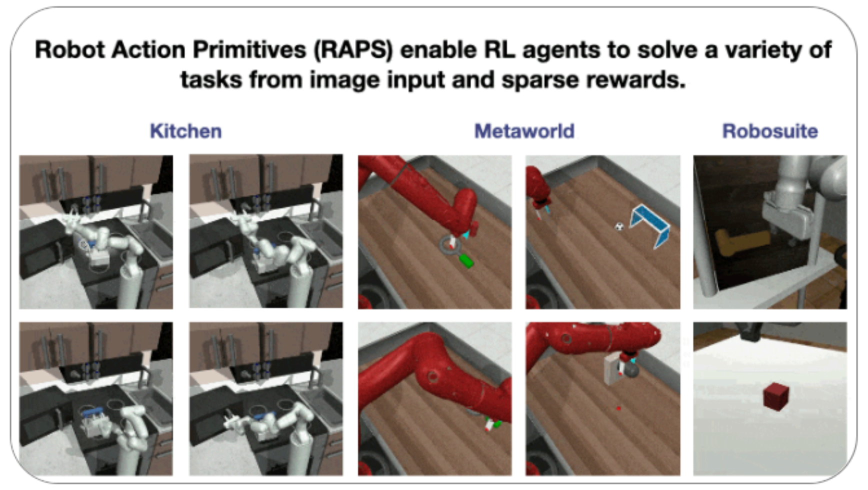
**Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents**
Wenlong Huang, Pieter Abbeel, Deepak Pathak*, Igor Mordatch*
arXiv 2022

webpage | pdf | abstract | bibtex | arXiv | code | demo video




**Discovering and Achieving Goals via World Models**
Russell Mendonca*, Oleh Rybkin*, Kostas Daniilidis, Danijar Hafner, Deepak Pathak
NeurIPS 2021

webpage | pdf | abstract | bibtex | code | benchmark | talk video




**Accelerating Robotic Reinforcement Learning via Parameterized Action Primitives**
Murtaza Dalal, Deepak Pathak*, Ruslan Salakhutdinov*
NeurIPS 2021

webpage | pdf | abstract | bibtex | arXiv |

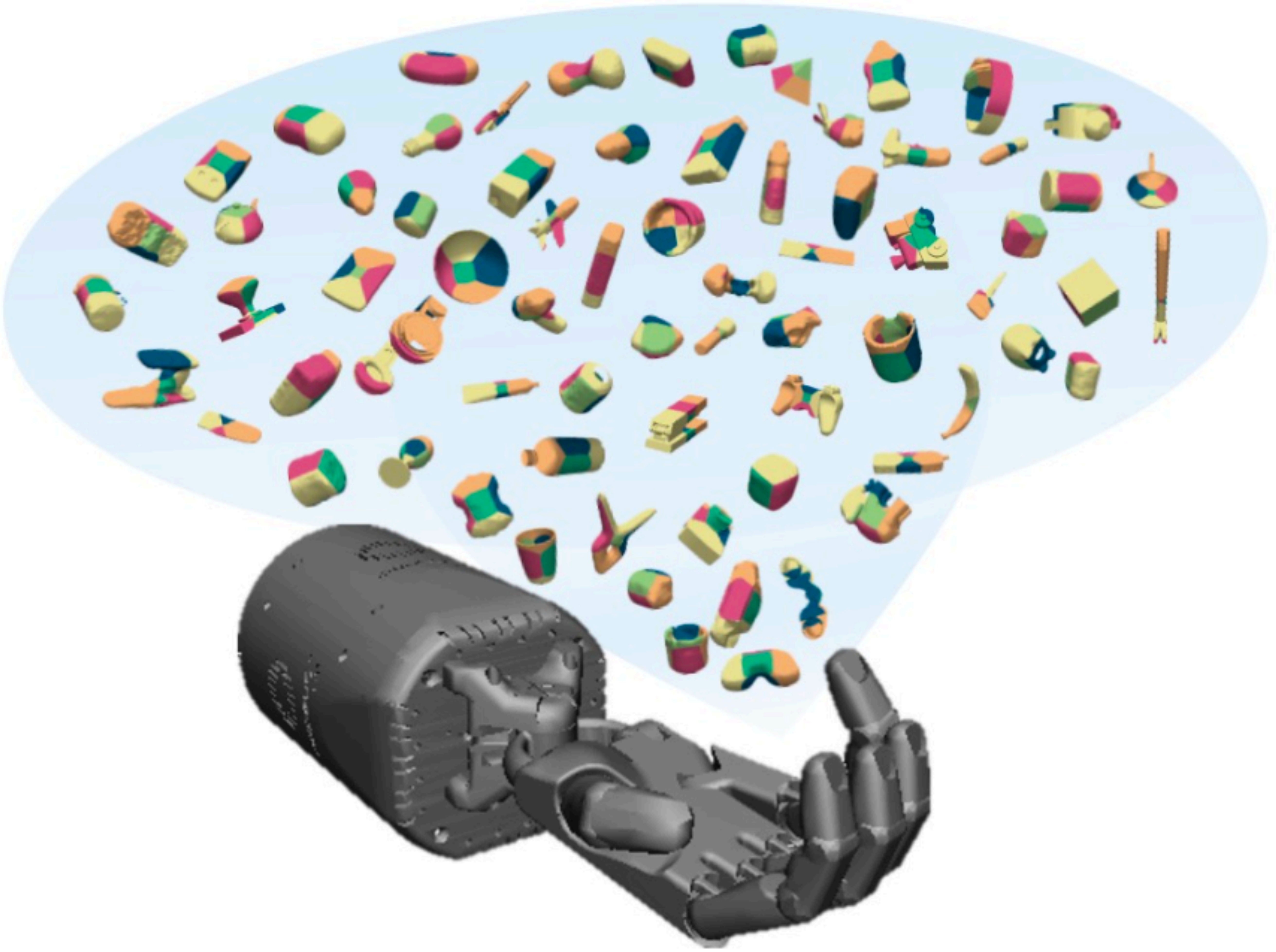


Fig. 1: Our goal in this work is to train a single policy that can perform in-hand manipulation on a large number of objects. We show surprising results that simple multi-task learning combined with appropriate representation not only achieves the aforementioned goal but also outperforms the single-task oracles, on both training and unseen objects.

*"However, generalization to completely unseen and geometrically-diverse objects for dexterous manipulation policies has been under-explored in the community, mostly due to the specious belief that such generalization is out of reach for current RL algorithms."*

*"In fact, we show that in the context of dexterous manipulation, a multi-task policy can be a generalist that can match the performance of those single- task specialist policies."*

# Multi-task Learning

# Objective for the Vanilla Multi-Task Policy

number of tasks

policy

$$\max_{\theta_p} \mathbb{E}_{\pi_{\theta_p}} \sum_{i=1}^{N}\sum_{t=0}^{\infty}\left[\gamma^t\ r_t^i(s_t,\ \underline{\pi_{\theta_p}(s_t)})\right]$$

action obtained by executing
the policy at state s

discount
factor

reward
for task i
at time t

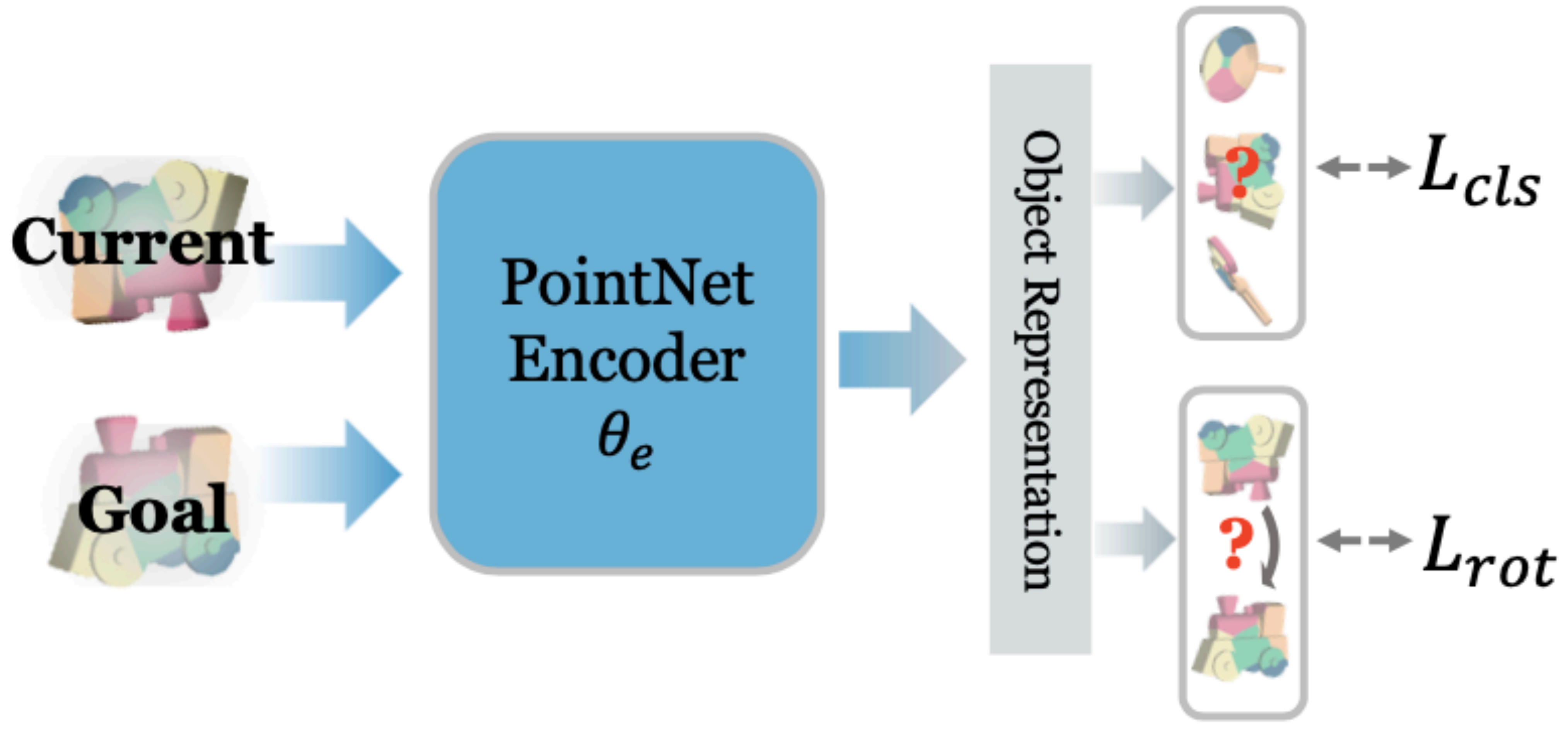
state

**Optimize the above objective function**
**using the DDPG (Deep Deterministic Policy Gradient) algorithm**
**with HER (Hindsight Experience Replay)**

# Multi-task Learning with Geometric Information

*"To explicitly model the object geometries, we propose to learn an object representation encoder based on object point clouds."*

*"However, in the presence of many different objects, we lack a canonical coordinate frame for them, making rotation matrix prediction based on a single point cloud an ill-defined problem. We resolve this by making the encoder module take as input two copies of the same object point cloud. The first describes the current orientation at time t and the second describes the desired orientation. The encoder is tasked with predicting the correct class of the object and the relative rotation matrix between the two point clouds."*




Representation Encoder Pre-Training

# Multi-task Learning
# for the Geometry Aware Multi-Task Policy

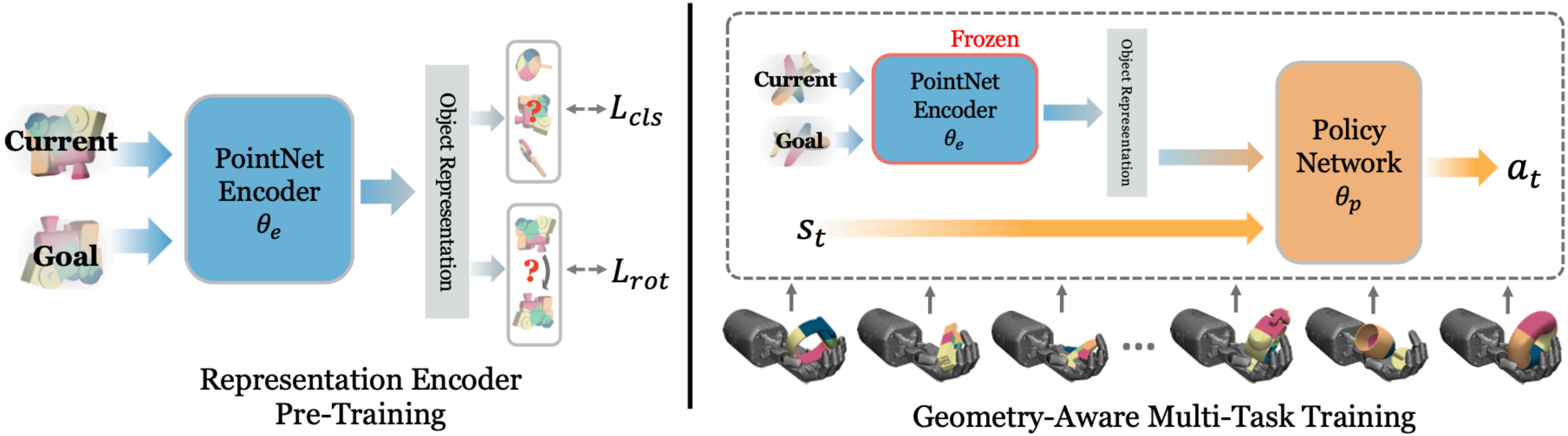



Fig. 2: We show that simple extensions to existing RL algorithms can produce geometry-aware dexterous manipulation policies that are robust to over 100 diverse objects. We first train an object representation encoder using object point clouds (left). Then we perform multi-task RL training on a large number of objects leveraging the encoded object representation (right).

# Implementation Details

- Start with OpenAI Gym

- Extend it with YCB and ContactDB datasets
- Proportionally scale each object so it can fit into the palm and be touched by fingers
- Filter out similar objects — resulting set has 114 hand-sized objects

- Train a policy for each object individually using single-task RL

- Split objects into 85-object training set / 29-object testing set such that training and test sets have similar success rates on average in the single-task policy

# OpenAI Gym Robot Environments


FetchSlide-v1
Slide a puck to a goal position.

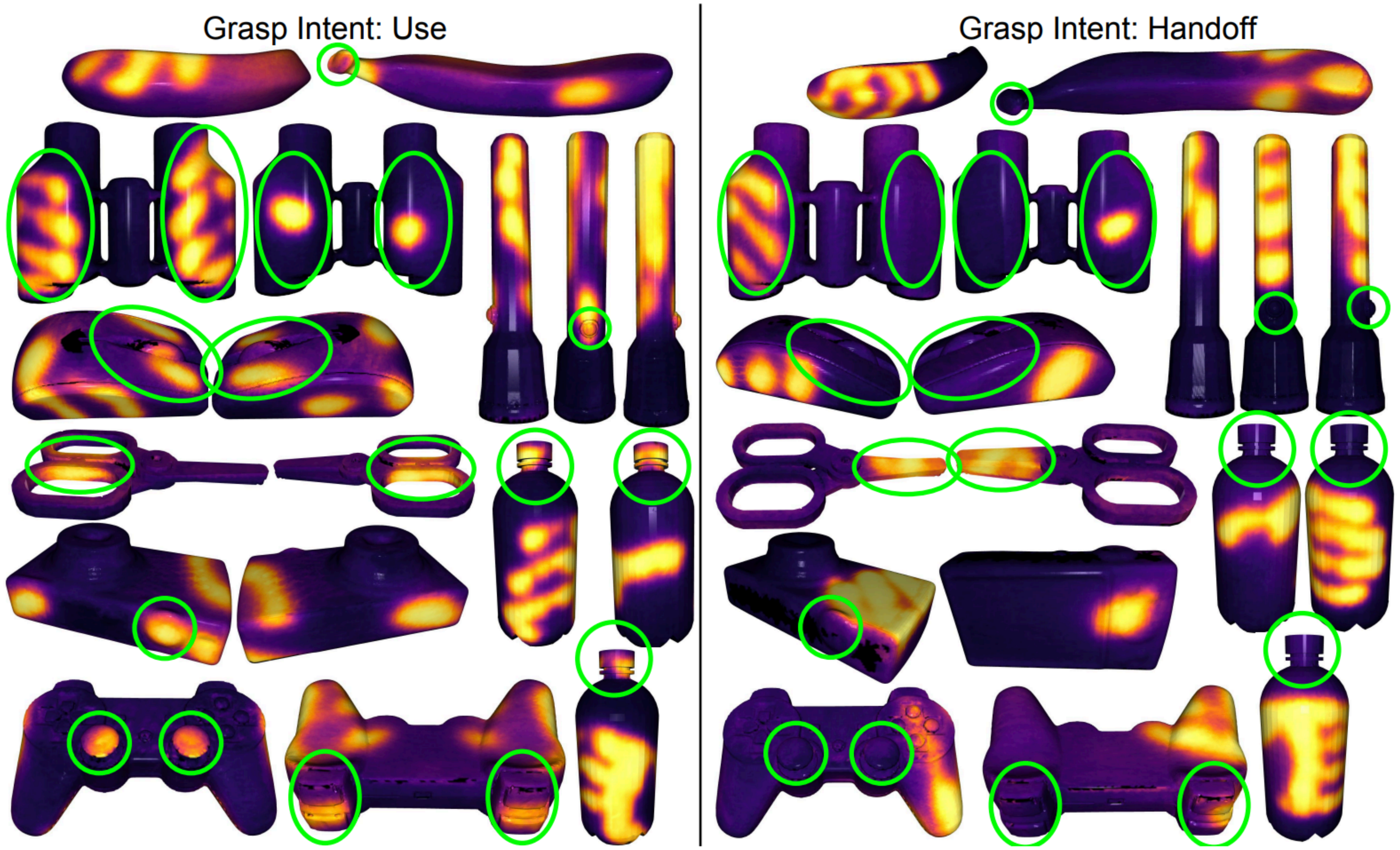

HandManipulateBlock-v0
Orient a block using a robot hand.

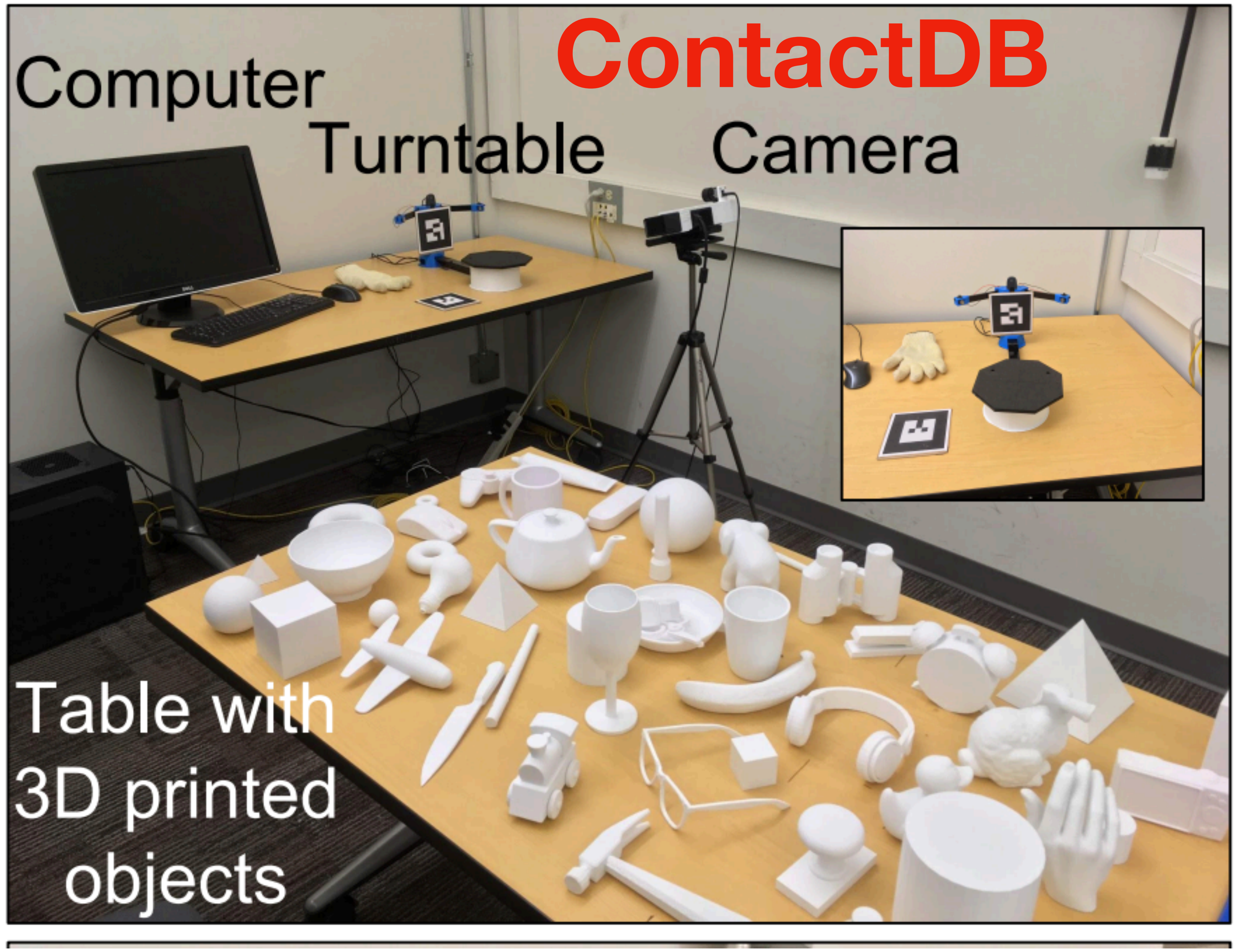

HandManipulateEgg-v0
Orient an egg using a robot hand.


HandManipulatePen-v0


HandReach-v0


YCB
CHEEZ-IT
Pringles
Soft Scrub
Comet
Color Cubes


Grasp Intent: Use
Grasp Intent: Handoff


ContactDB
Computer
Turntable
Camera
Table with
3D printed
objects

# More Implementation Details

- State:
  - joint angles
  - joint velocities
  - object position
  - object orientation
  - object desired orientation
  - object velocity
  - object angular velocity
  - positions and surface normals of 128 randomly sampled points (geometry-aware policy only .. points are resampled at each tilmestep)

- Action:
  - joint angle positions for 20 actuated joints of the Shadow Hand (the other 4 joints are coupled DoF)

- Reward:
  - 1 if the orientation is within 0.1 radians of the desired orientation
  - 0 otherwise

- Initialization and goal selection:
  - The initial and goal orientation are sampled independently and randomly about the z-axis for each episode.

# Research Questions

- Can vanilla multi-task policy attain competitive performance on a large number of objects?
- Leveraging object representation, can a single geometry-aware policy interpolate its experience and outperform single-task oracles?
- What are the generalization properties of a geometry- aware policy?

- Can vanilla multi-task policy attain competitive performance on a large number of objects?

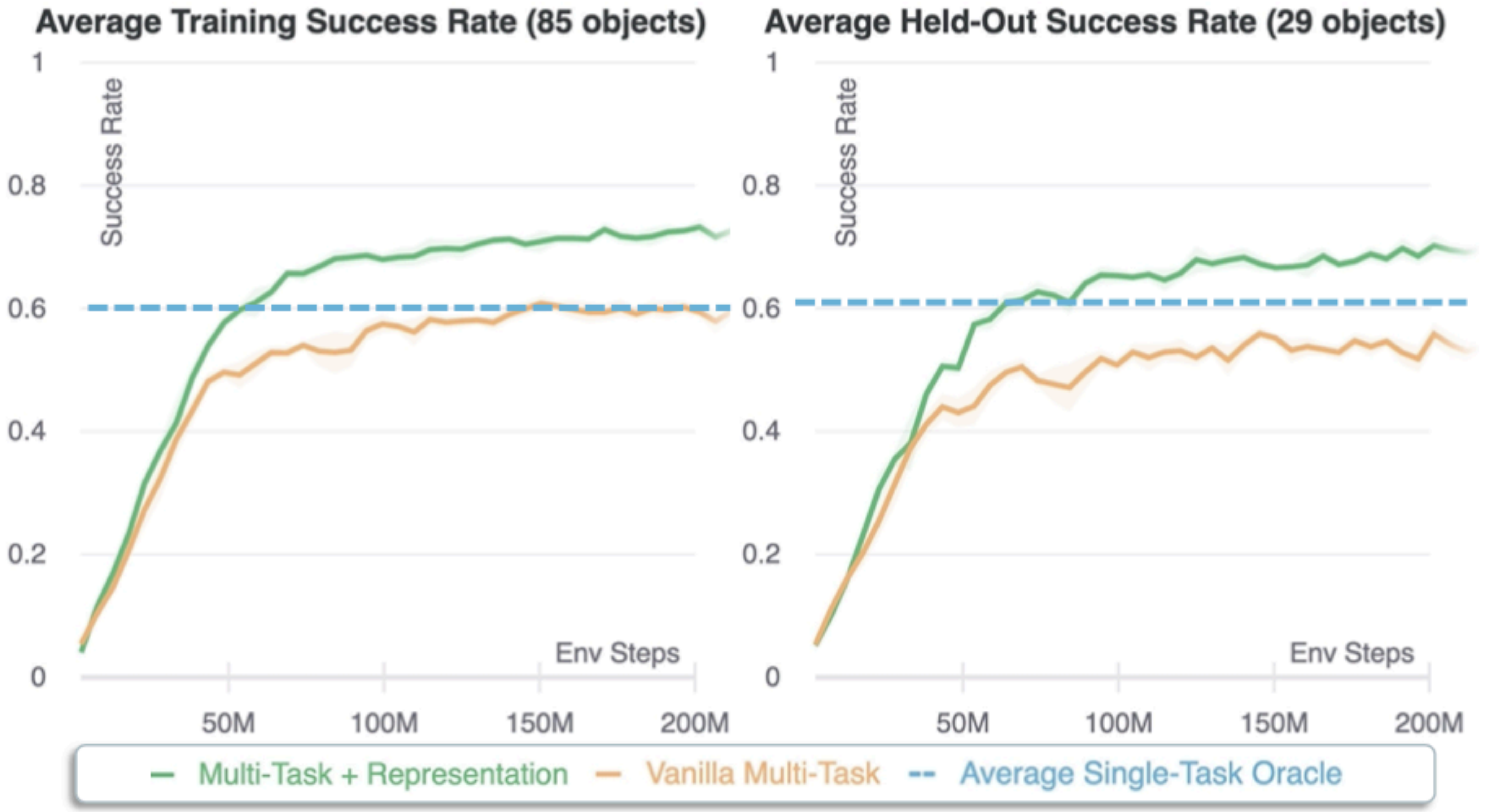
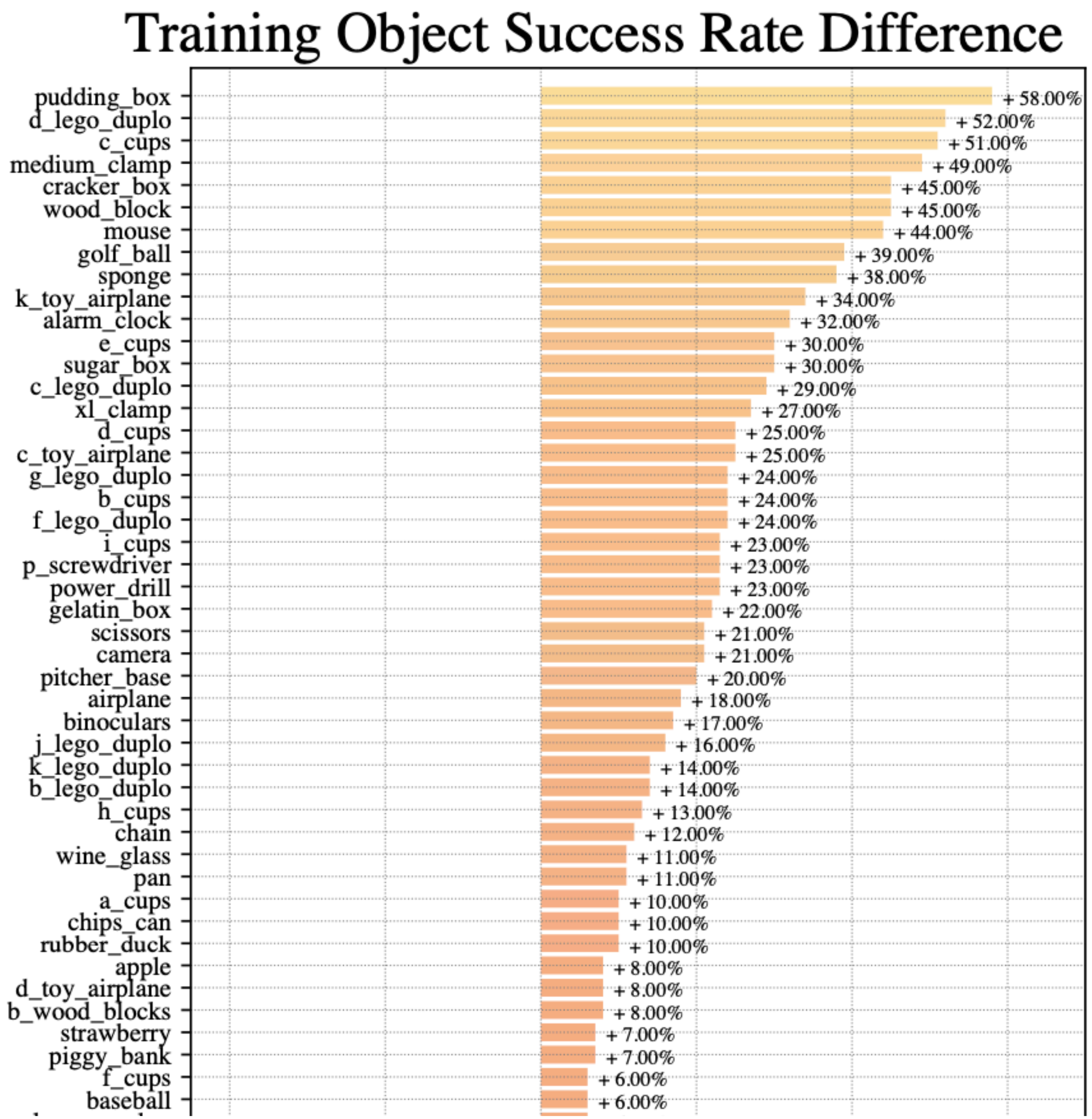



Fig. 4: Average success rate across 85 training objects and across 29 held-out objects. The plot shows that multi-task joint training can lead to a surprisingly robust policy on both training and testing, with similar performance compared to the average of individual single-task oracle trained for each object. Furthermore, when combined with object representation, a joint policy can even outperform the oracles on held-out objects, in a completely zero-shot manner. The success rate reported are averaged across 425 and 145 episodes, respectively for all training objects and all held-out objects.

- Leveraging object representation, can a single geometry-aware policy interpolate its experience and outperform single-task oracles?



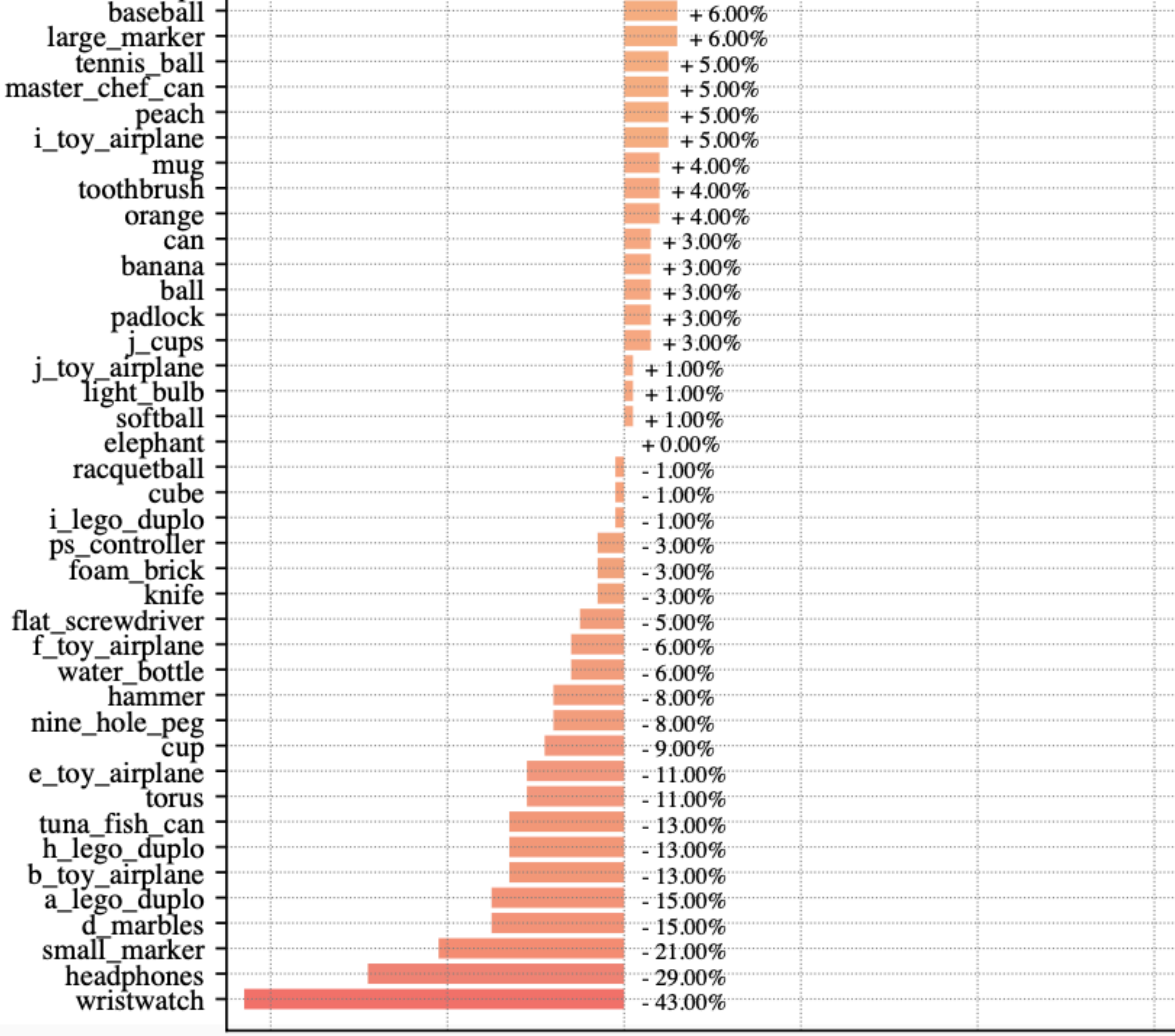




Fig. 3: Success rate difference between geometry-aware multi-task policy and individual oracles on the 85 training objects, calculated as $\Delta S = (S_{\text{ours}} - S_{\text{oracle}})$. It shows that a single geometry-aware multi-task policy can attain even better performance than individual single-task oracles on most training objects. It demonstrates that the policy can leverage skills learned from many tasks, leading to an overall stronger policy. The success rate reported are averaged across 100 episodes.

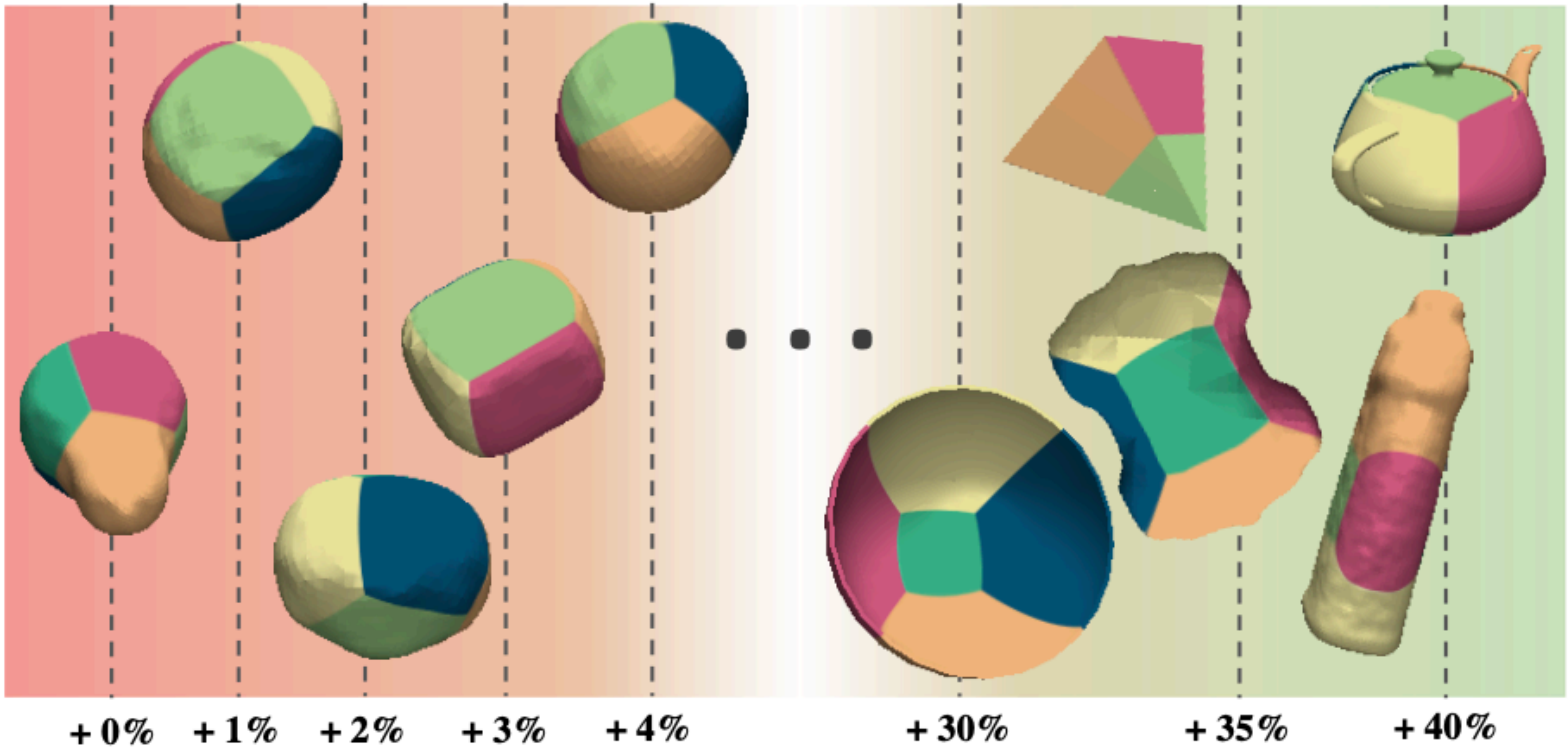



Fig. 5: Visualization of held-out objects ranked by the performance gains of geometry-aware policy, calculated as $\Delta S = (S_{\text{ours}} - S_{\text{vanilla}})$. Notice that the gains are the highest for objects with irregular shapes and the lowest for medium-sized and spherical objects, showing the policy can effectively leverage object representation to adopt specific strategies even for challenging unseen objects.

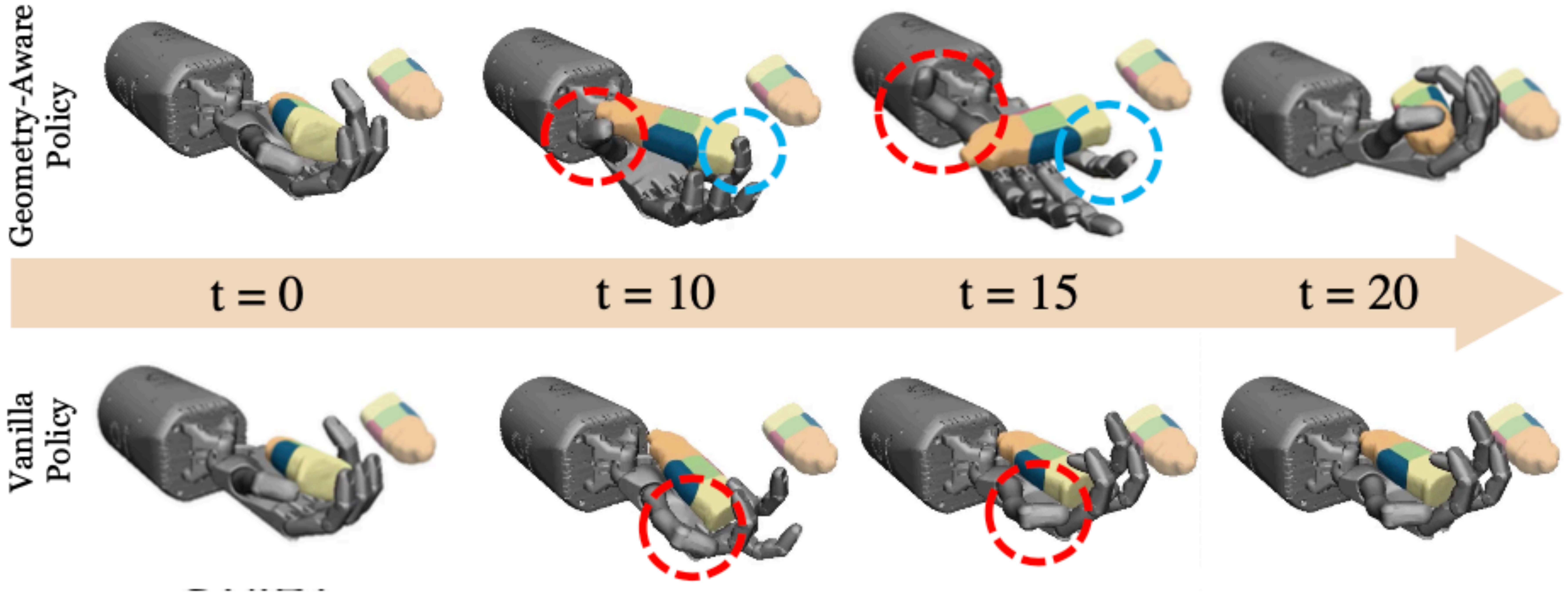



Fig. 6: The top row shows the progression of geometry-aware policy and the bottom row shows the vanilla policy on the ”bleach_cleanser” object. Image on the right of each hand shows goal orientation. Our geometry-aware policy can reason about shape and move thumb and finger accordingly so as to rotate the object while vanilla policy can’t as shown by circles.

- What are the generalization properties of a geometry- aware policy?

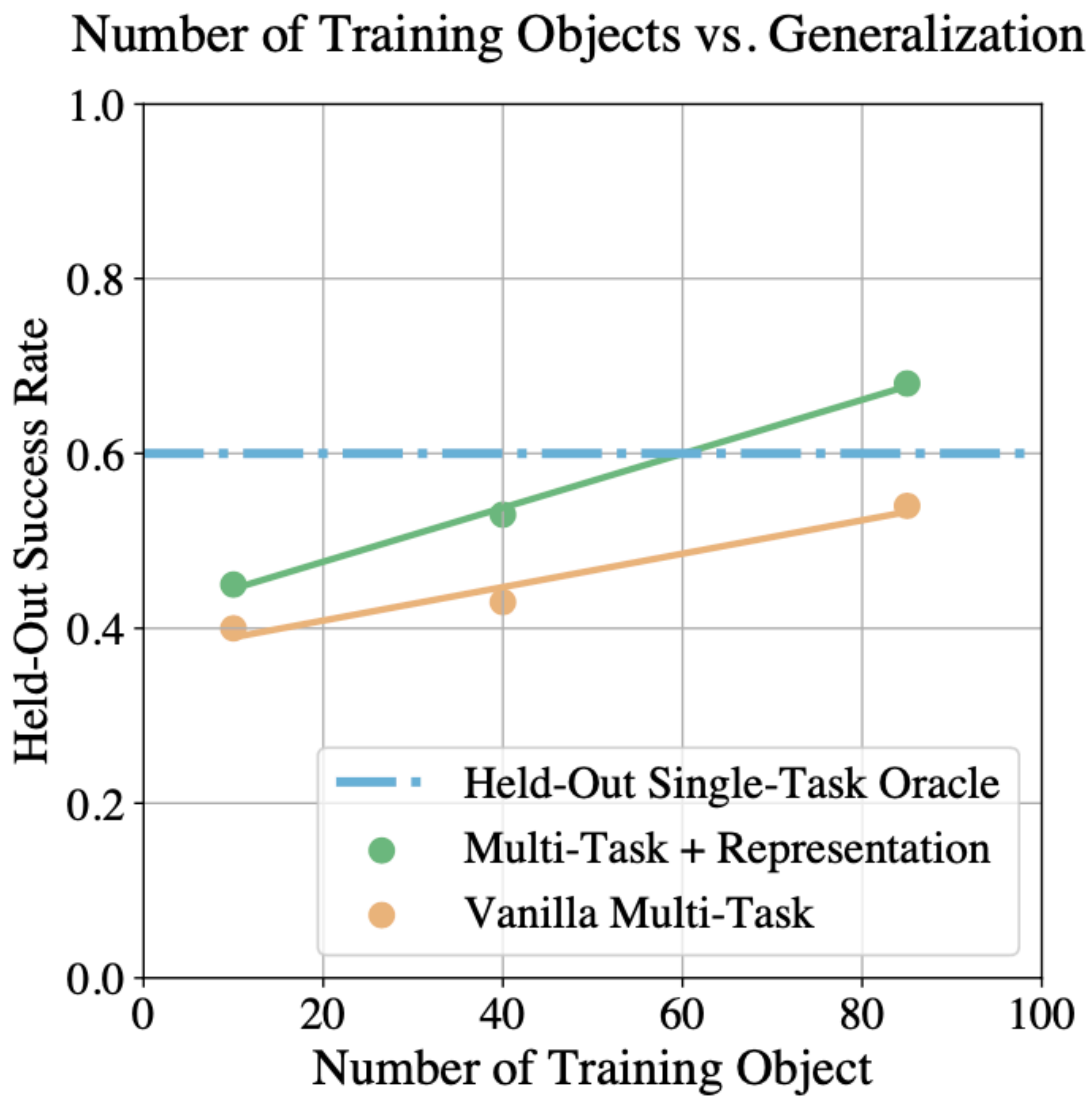



Fig. 7: Comparisons of the effect of the number of training objects on zero-shot generalization. More training objects would lead to a more robust policy that can even surpass single-task oracles on held-out objects.

| | w/ Object Rep. | w/o Object Rep. |
|---|---|---|
| large_clamp | **97.00%** | 46.00% |
| door_knob | **84.00%** | 65.00% |
| large_clamp | **92.00%** | 74.00% |

TABLE I: Effect of object representation used for single-task training. The objects are randomly selected from the held-out set. Even though the frozen encoder has not seen the objects in the pre-training phase, the encoded representation is still shown to be beneficial for single-task RL training, suggesting geometry-awareness is important for dexterous manipulation policies.

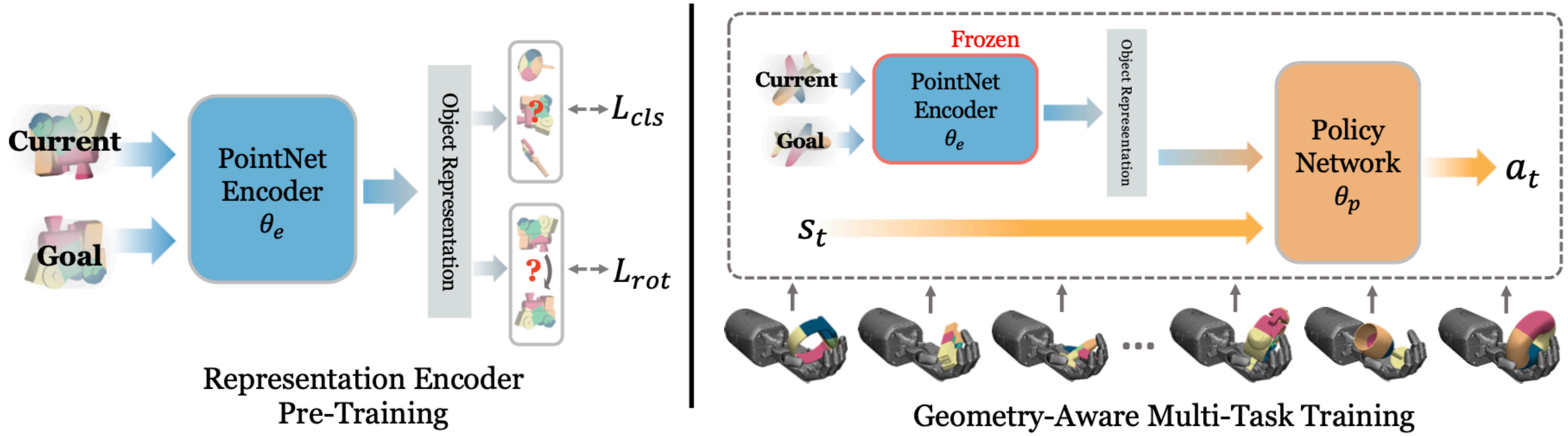



Fig. 2: We show that simple extensions to existing RL algorithms can produce geometry-aware dexterous manipulation policies that are robust to over 100 diverse objects. We first train an object representation encoder using object point clouds (left). Then we perform multi-task RL training on a large number of objects leveraging the encoded object representation (right).

| | Frozen Encoder | Fine-tuned Encoder |
|---|---|---|
| Training Success Rate | **71.88%** | 61.62% |
| Held-Out Success Rate | **68.80%** | 59.84% |

TABLE II: Comparisons of frozen encoder vs. fine-tuned encoder. Frozen encoder has much better performance than fine-tuned variant, whose performance is similar to that of vanilla multi-task policy without an encoder. The success rate reported are averaged across 425 and 145 episodes, respectively for 85 training objects and 29 held-out objects.

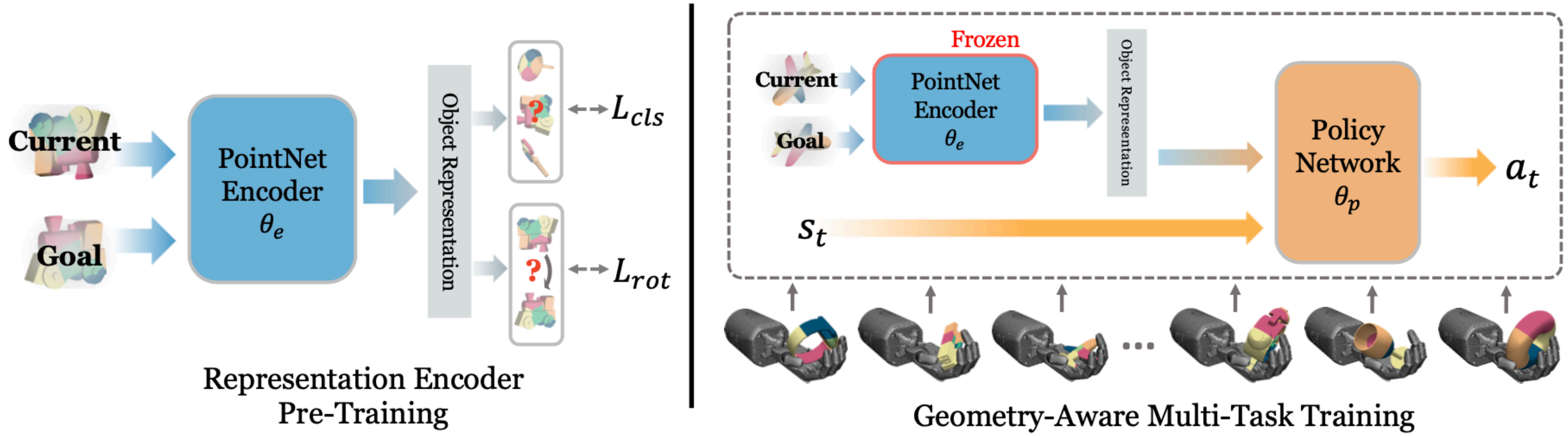



Fig. 2: We show that simple extensions to existing RL algorithms can produce geometry-aware dexterous manipulation policies that are robust to over 100 diverse objects. We first train an object representation encoder using object point clouds (left). Then we perform multi-task RL training on a large number of objects leveraging the encoded object representation (right).

# Results Videos

https://wenlong.page/geometry-dex/

# Comparison to Monday’s paper

*“A parallel work also studies dexterous manipulation on a variety of objects [40]. However, their approach does not condition the multi-task learning on the object geometric representation. Hence, it forces the policy to discover a common ” generally” good strategy that works across many distinct objects of simpler shapes but may suffer to generalize to objects of more challenging geometries.”*

# Differences

- Start and goal: SO(3) vs. rotations around z-axis
- Objects: YCB+EGAD vs. YCB+ContactDB
- Environments: Isaac Gym vs. OpenAI Gym
- Learning algorithms: PPO vs. DDPG

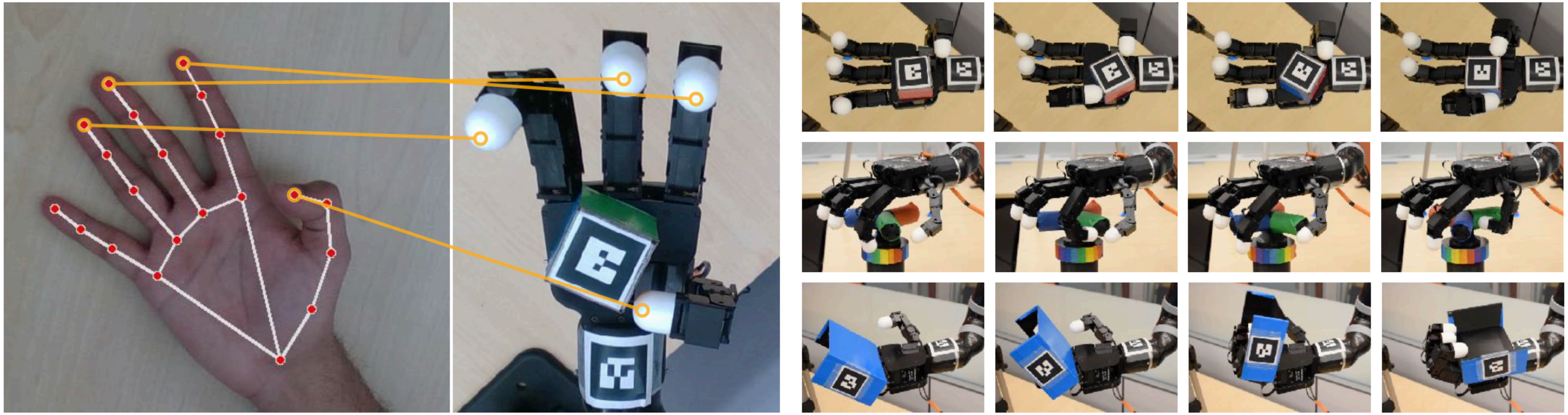
Table 1: Success rates (%) of policies tested on different dynamics distribution. $\bar{\theta} = 0.1$rad. DR: domain randomization and observation/action noise. X→Y: distill policy X into policy Y. The full table is in Table D.5.

| | | | | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| | | | | Train without DR | | Train with DR |
| Exp. ID | Dataset | State | Policy | Test without DR | Test with DR | Test with DR |
| B | EGAD | Full state | RNN | $95.95 \pm 0.8$ | $84.27 \pm 1.0$ | $88.04 \pm 0.6$ |
| E | | Reduced state | RNN→RNN | $91.96 \pm 1.5$ | $78.30 \pm 1.2$ | $80.29 \pm 0.9$ |
| G | YCB | Full state | RNN | $80.40 \pm 1.6$ | $65.16 \pm 1.0$ | $72.34 \pm 0.9$ |
| J | | Reduced state | RNN→RNN | $81.04 \pm 0.5$ | $64.93 \pm 0.2$ | $65.86 \pm 0.7$ |

**Which do you like better?**

# Dexterous Imitation Made Easy: A Learning-Based Framework for Efficient Dexterous Manipulation

Sridhar Pandian Arunachalam[†], Sneha Silwal[†], Ben Evans, and Lerrel Pinto
New York University



(a) Teleoperation through a single RGB camera.

(b) Learned policies for dexterous manipulation.

Fig. 1: The framework for dexterous manipulation consists of two phases. (a) Demonstrations are collected using a real-time hand tracker on a single visual stream of a human operator's hand. The estimated fingertip 2D pixel coordinates are retargeted to 3D coordinates in the robot frame. (b) Given these demonstrations, dexterous manipulation policies are learned on both a real Allegro Hand, using nearest neighbor-based imitation, and on a simulated Allegro Hand using RL augmented with demonstrations.


**Sridhar Pandian Arunachalam** · 3rd

Graduate Research Assistant at Courant Institute of Mathematical Sciences

**Sneha Silwal** ·

ssilwal.com

Sneha

studying reinforcement learning and robotics at NYU

worked as a software engineer at Pilot AI & IBM

studied Computer Science & Economics at Columbia

Ben Evans

2nd Year PhD student

GRAIL at New York University

# Biography

I am a second-year PhD Student in the CILVR group advised by Lerrel Pinto. My interests include Robot and Reinforcement Learning to enable robots to sense, think, and act in unstructured environments.

I completed my undergraduate degree in Computer Science at the University of Washington, working at first in the Robotics and State Estiation Lab with Arunkumar Byravan on video prediction, then in the Movement Control Lab with Kendall Lowrey and Aravind Rajeswaran on off-policy reinforcement learning and nonlinear model predictive control.

## Interests

- Robot Learning
- Reinforcement Learning
- Directed Exploration

## Education

- PhD in Computer Science, 2020-present
  New York University
- BS/MS in Computer Science, 2015-2020
  University of Washington

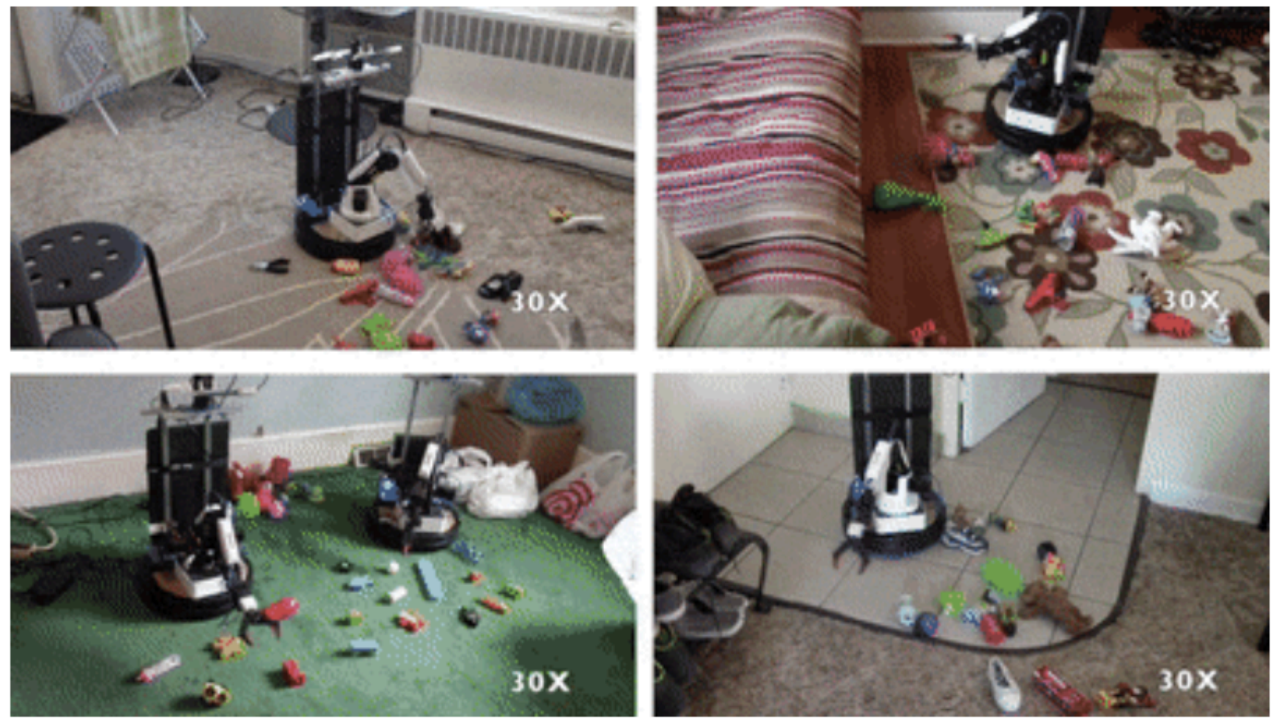
## Welcome!

I am an Assistant Professor of Computer Science at NYU Courant working on problems in Robotics and Machine Learning. I am also affiliated with the Center for Data Science. Together with several wonderful colleagues I am part of the CILVR (Computational Intelligence, Learning, Vision and Robotics) group.

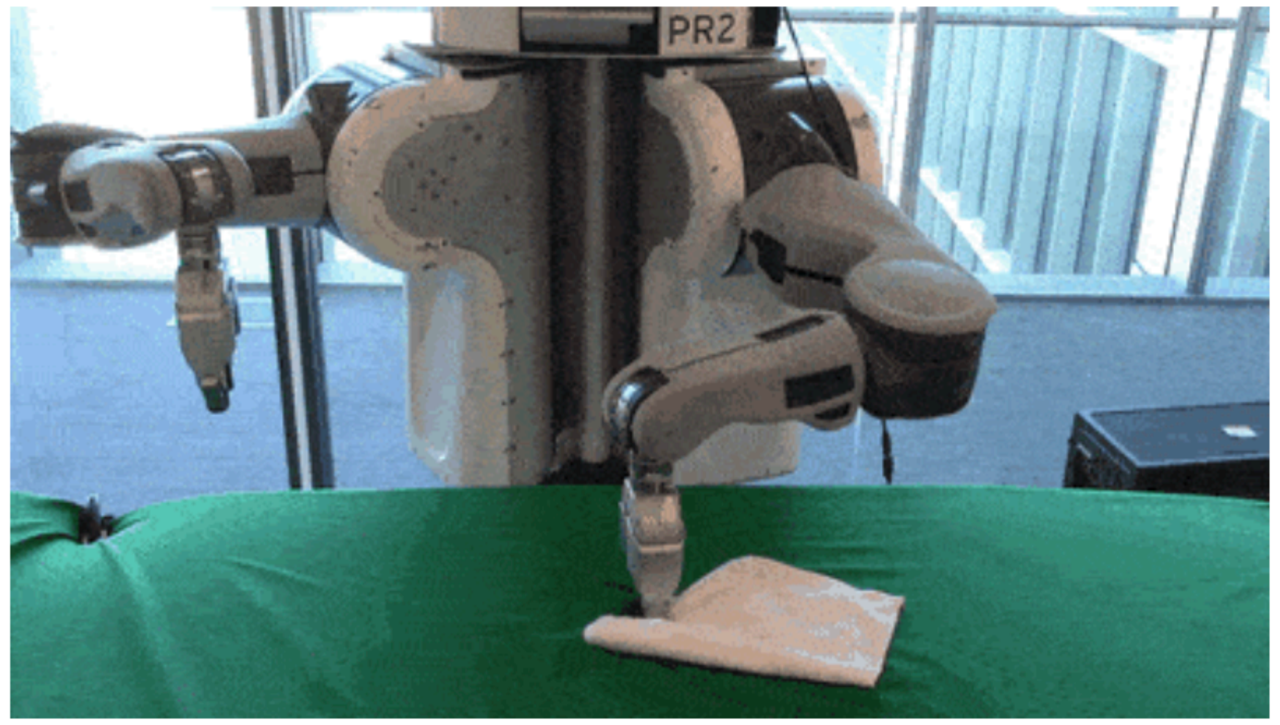
My lab's goal is to get robots to generalize and adapt in the diverse world we live in. To this end, my research touches the areas of Robot Learning, Representation Learning, Reinforcement Learning, and Affordable Robotics.


Learning to Grasp

Learning to Fly

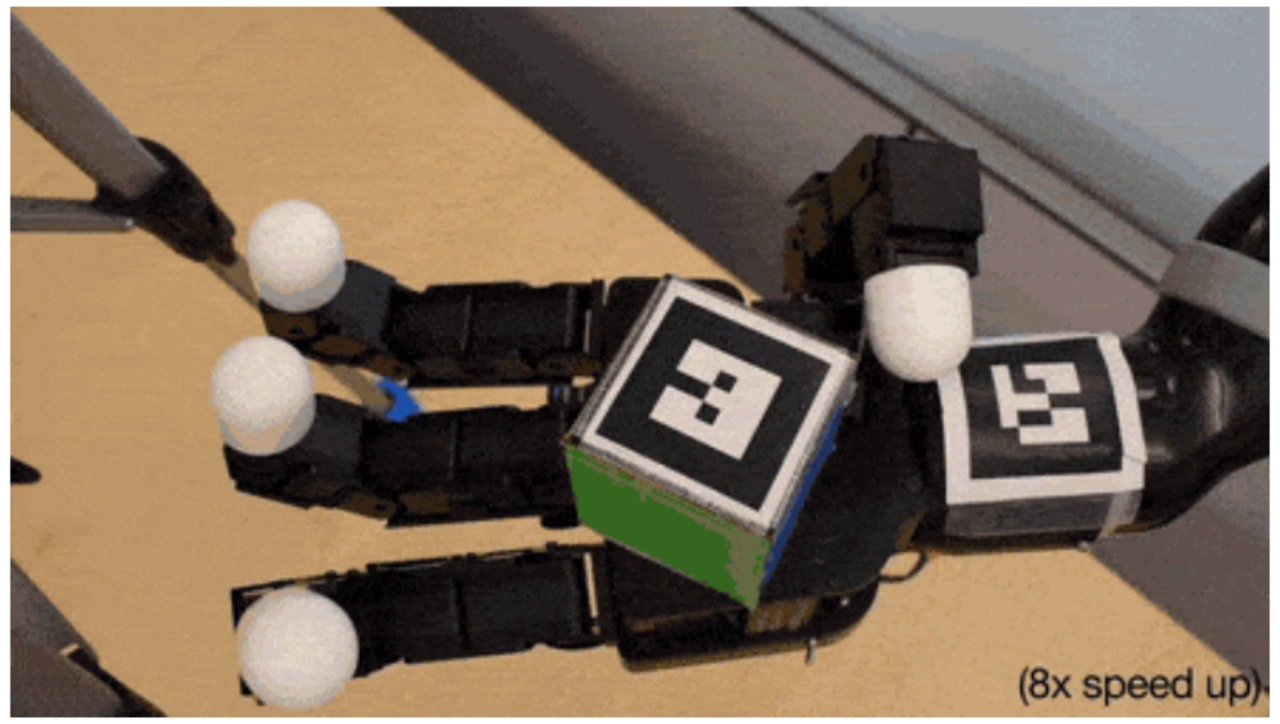



Learning in Homes

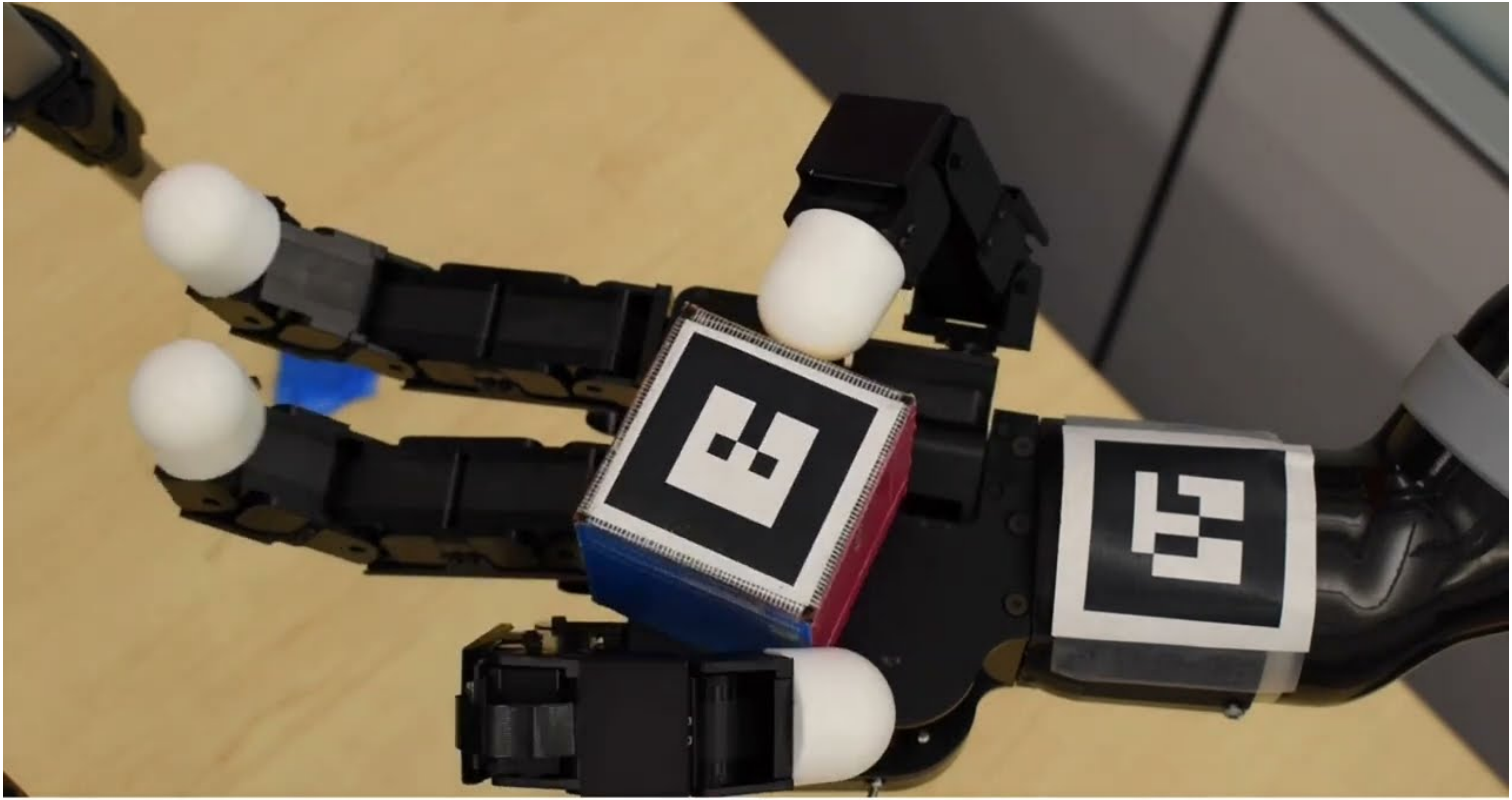



Manipulating Deformable Objects


Visual Imitation Learning




Dexterous Manipulation



Cube rotation policy using **image-based** NN. (8x speed up)

# Setup

- 3 tasks:  Flipping, Spinning, Rotating

- Collect **30** human demonstrations of each task

  - flipping — 30s on average to teleoperate

  - spinning — 120 seconds on average..

  - rotating — 150 seconds on average..

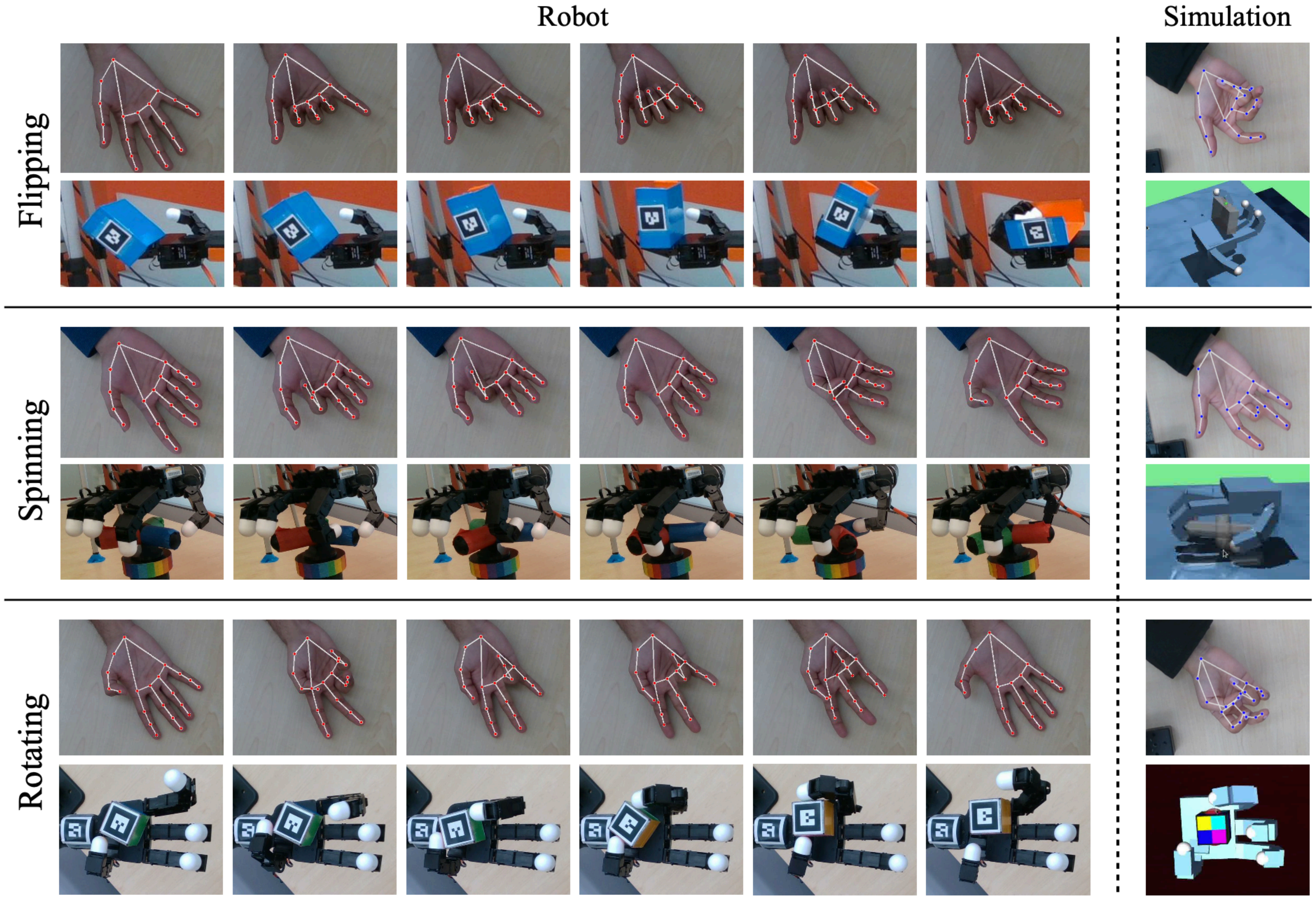


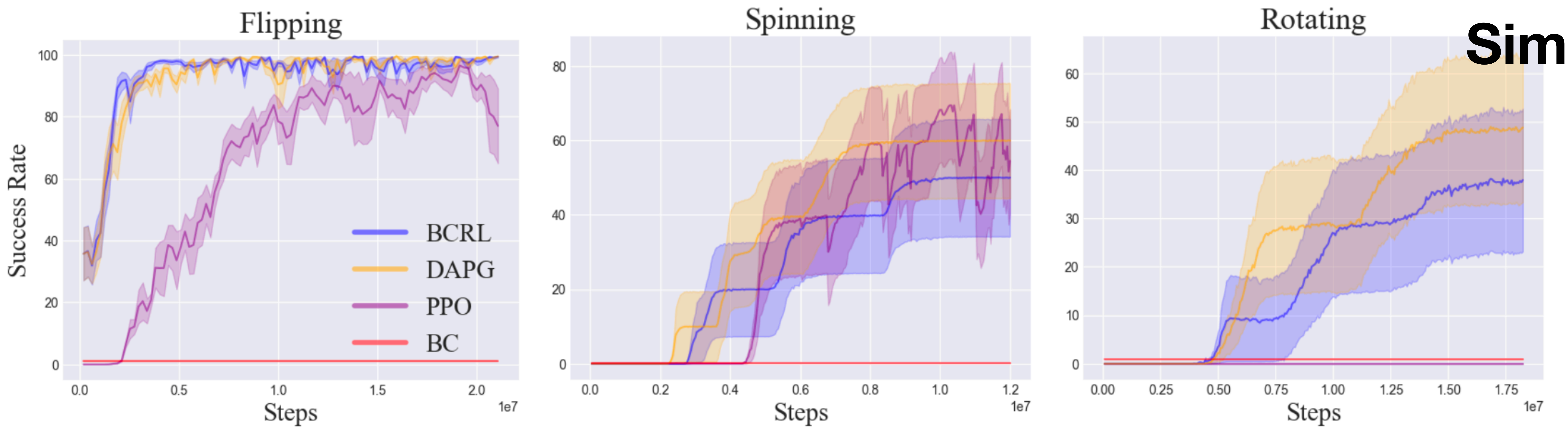

Fig. 3: The demonstration collection process for the three tasks. For each task, the operator's hand in the upper row is depicted followed by the corresponding state of the robot's hand in the lower row. The rightmost column visualizes the operator's actions during demonstration collection along with the simulated MuJoCo environments for each task.

# Setup

- Simulation experiments
    - Behavior cloning alone (BC)
    - PPO alone (no use of demonstrations)
    - BCRL (behavior cloning fine tuned with RL)
    - DAPG (essentially PPO initialized with BC)

- Real robot experiments
    - Behavior cloning alone (BC)
    - Nearest neighbor
        - state based (INN)
        - vision-based (VINN)

# Results - Simulation




*The policies trained with BCRL and DAPG produced similar results to one another, whereas PPO methods had more erratic movements.*

*The policies that used demonstrations were also qualitatively better than the teleoperated demonstrations, which would often take longer to record and have more abrupt starts and stops.*

*Upon visualizing the policies, we noticed that the PPO policies were successful because extreme, random movements of the middle and last finger were enough to spin the handle. Whereas the policies learned with demonstrations were more 'human-like' and smoother.*

*Pure behavior cloning (BC) fails on all tasks. Since the number of demonstrations used is relatively small, BC policies are unable to remain in the support of demonstration data and fail.*

# Results - Real Robot

TABLE I: Success rates on our real Allegro hand using DIME.

| Method Used | Flipping | Turning | Rotation | |
|---|---|---|---|---|
| | | | 90° | 180° |
| INN (State Based) | 80% | 60% | 100% | 80% |
| Behavior Cloning (State Based) | 0% | 0% | 0% | 0% |
| VINN (Image Based) | 90% | 0% | 70% | 50% |
| Behavior Cloning (Image Based) | 0% | 0% | 0% | 0% |

*We notice that non-parametric nearest neighbors (INN) outperforms parametric behavior cloning approaches across all tasks*

# Results - Real Robot

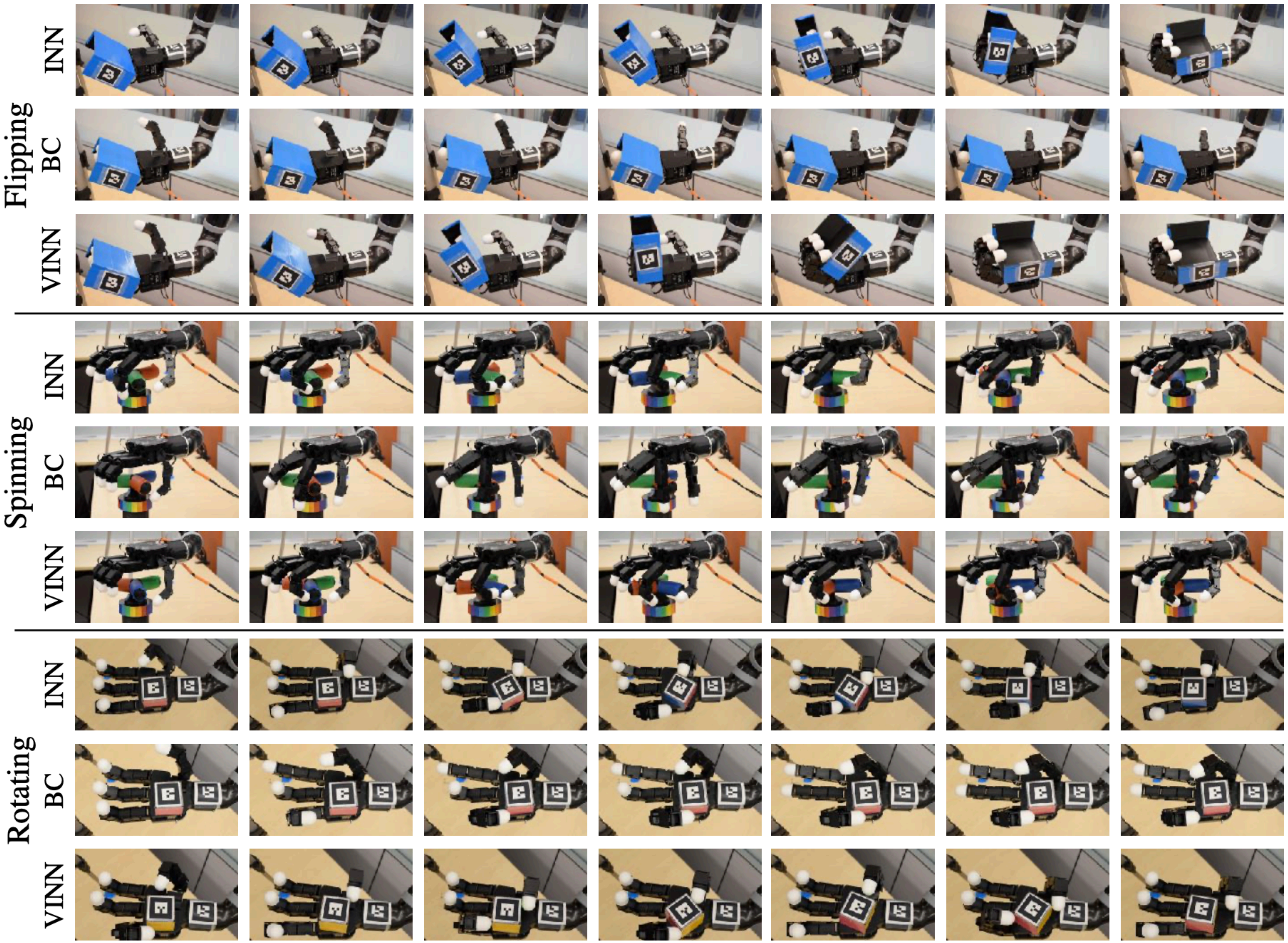
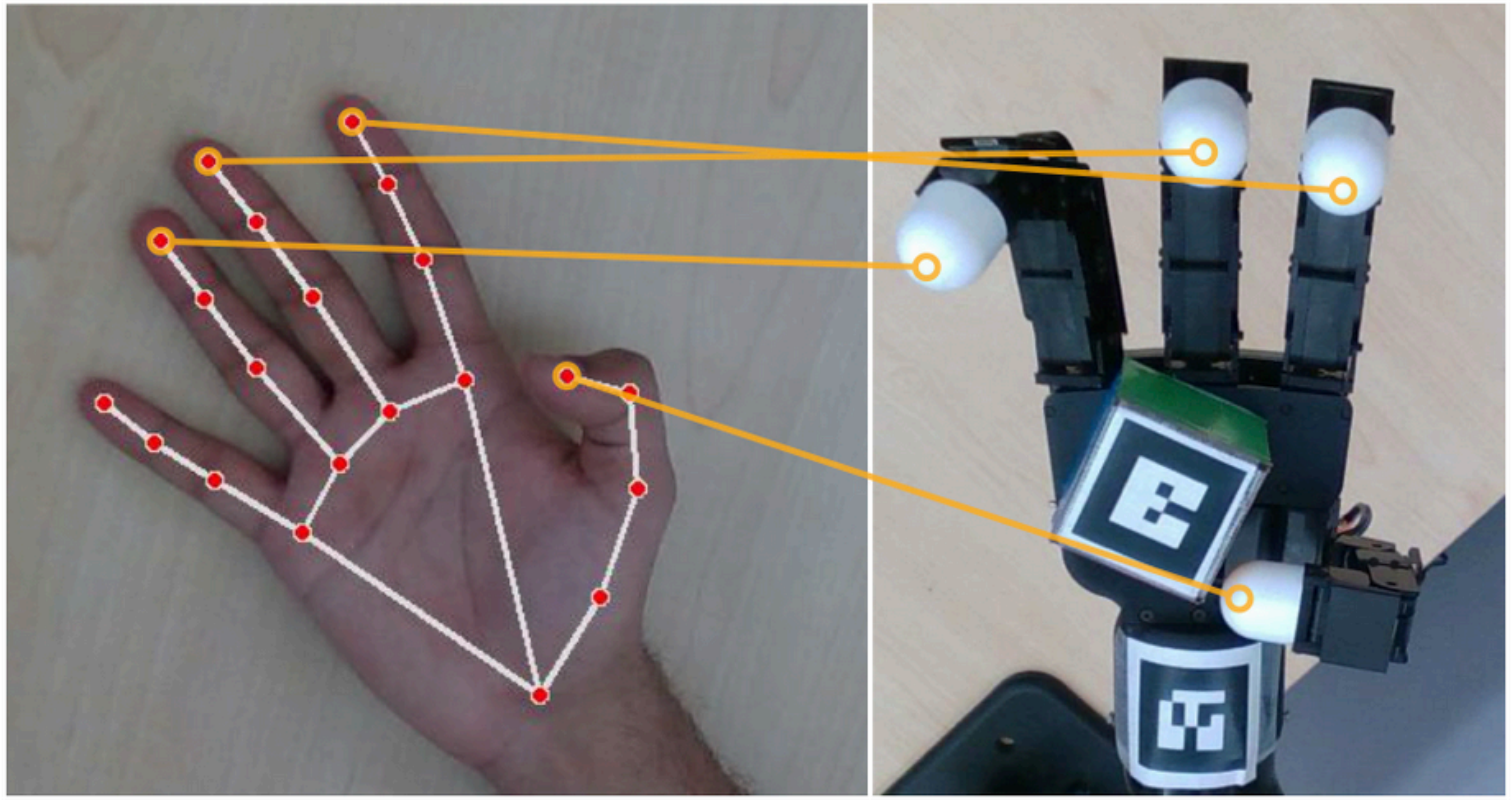


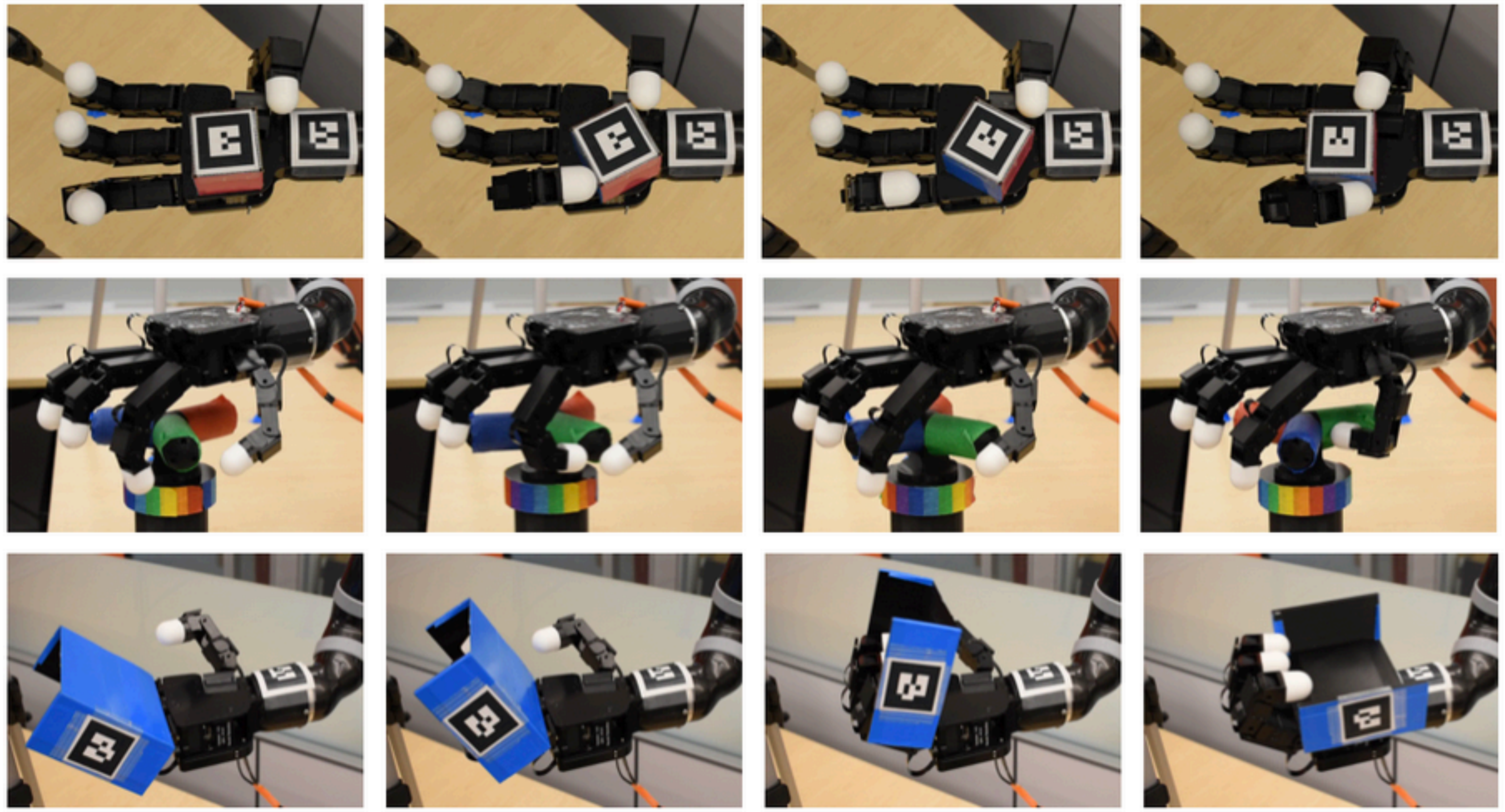

Fig. 5: Robot runs for both state-based (INN) and image-based (VINN) non-parametric nearest neighbors along with state-based parametric behavior cloning (BC) are visualized across the three tasks. INN performs the best across all three tasks, while VINN is able to solve the *flipping* and *rotating* tasks. BC is unable to solve any task and suffers from distributional mismatch [40].

# Conclusion

*first step towards training dexterous robots from inexpensive demonstrations*



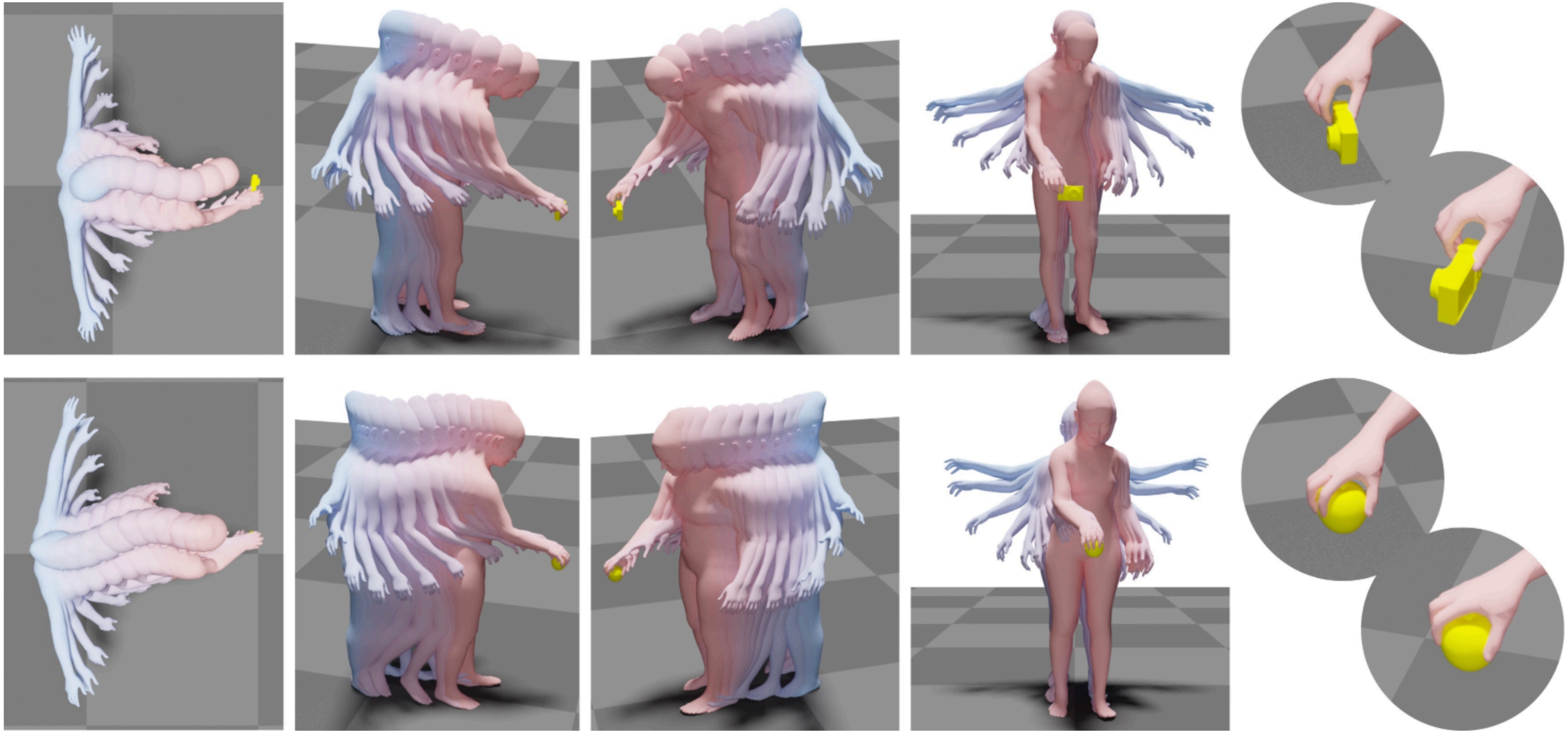
(a) Teleoperation through a single RGB camera.



(b) Imitation learning for dexterous manipulation.

# GOAL: Generating 4D Whole-Body Motion for Hand-Object Grasping

**Omid Taheri**, **Vasileios Choutas**, **Michael J. Black**, and **Dimitrios Tzionas**



# Omid Taheri

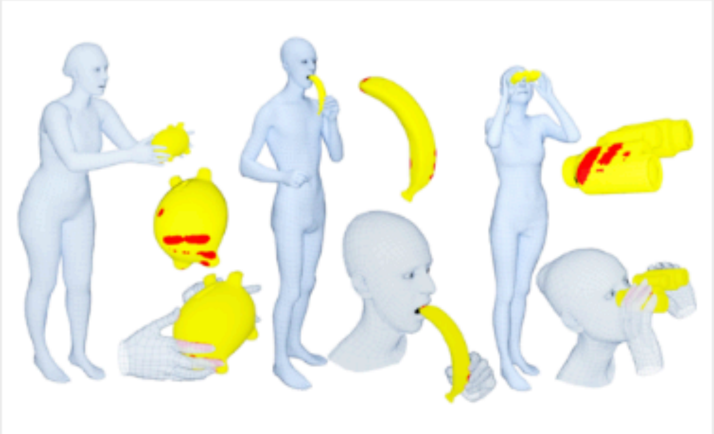
Ph.D. Student

Perceiving Systems

Overview Projects **Publications**

1 result (View BibTeX file of all listed publications)

2020



GRAB: A Dataset of Whole-Body Human Grasping of Objects

Taheri, O., Ghorbani, N., Black, M. J., Tzionas, D.

In *Computer Vision – ECCV 2020*, LNCS 12355, pages: 581-600, Springer International Publishing, Cham, August 2020 (inproceedings)

# Vassilis Choutas

Ph.D. Student

Perceiving Systems

Office: N3.004
Max-Planck-Ring 4

Overview CV Projects **Publications**

5 results (View BibTeX file of all listed publications)

Collaborative Regression of Expressive Bodies using Moderation

Feng, Y., Choutas, V., Bolkart, T., Tzionas, D., Black, M.

*International Conference on 3D Vision (3DV)*, pages: 792-804, December 2021 (conference)

Abstract

Biography | Research | Students | PostDocs and Research Scientists | Affiliations | Contact | Talks | Publications


Michael Black

Director

Perceiving Systems

Office: N3.019
Max-Planck-Ring 4
72076 Tübingen
Germany


## Research Interests

I am interested in motion.  What does motion tell us about the structure of the world and how can we compute this from video?  How do humans and animals move?  What goals drive behavior? My work combines computer vision, graphics and machine learning to develop new models and algorithms to capture, analyze, and synthesize the motion of humans, animals and the world.

My **Computer Vision** research addresses:

- articulated human motion pose estimation and tracking;
- the estimation of human body shape from images and video;
- the estimation of scene structure and physical properties from video;
- the estimation of optical flow;
- vision as inverse graphics.

My **Graphics** research addresses:

- virtual humans;
- next-generation motion capture;
- articulated and non-rigid shape representation;
- human and animal shape and motion capture;
- human animation, AR/VR, and Metaverse applications;
- capture and animation of clothing.

My **Machine Learning** reserarch addresses

- learning representations of 3D shape
- implicit functions
- neural rendering
- regressing 3D models from images
- temporal models of human motion
- learning 3D models from 2D data

I also work on industrial applications in **Fashion Science**:

- Body scanning and measurement;
- clothing sizing;
- cloth capture and modeling;
- virtual try-on.

My previous work on **Computational Neuroscience research** addressed:

- modeling the neural control of reaching and grasping;
- novel neural decoding algorithms;
- neural prostheses and cortical brain-machine interfaces;
- markless animal motion capture.

What ties this all together is my ultimate goal of understanding humans and their behavior by creating vritual humans. If we can can simulate a virtual human that behaves like a real human, then we have a working model of ourselves.

# Dimitrios Tzionas

Research Scientist

Perceiving Systems

Office: N3.009
Max-Planck-Ring 4
72076 Tübingen
Germany

Info | **News** | CV | Data/Models/Code | Students/Collab | Videos | Projects

**Mar 2022**:

- Accepted papers at CVPR 2022:
  - "*GOAL: Generating 4D Whole-Body Motion for Hand-Object Grasping*"
    Omid Taheri, Vasileios Choutas, Michael J. Black, Dimitrios Tzionas
    Project Website
  - "*ICON: Implicit Clothed humans Obtained from Normals*"
    Yuliang Xiu, Jinlong Yang, Dimitrios Tzionas, Michael J. Black
    Project Website
  - "*Accurate 3D Body Shape Regression using Metric and Semantic Attributes*"
    Project Website
  - "*Human-Aware Object Placement for Visual Environment Reconstruction*"
    Project Website

Fullscreen Mode

# GOAL - System Overview