The role of teleoperation (with VR/AR) in robotics

Guest Lecture

Ankur Handa

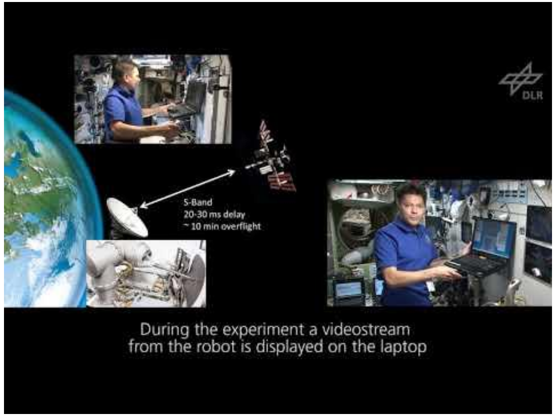
NVIDIA

Motivation: Using robots as tools

Most challenging problems are AI complete: robots serve as tools rather than doing things autonomously

DRC

EVA with NASA JSC



DLR's Justin Robot

Telesurgery (Da Vinci Robot)

Stanford's PR1 robot

All these systems show that we are not lacking in hardware rather the missing piece is clever software

Motivation: teaching robots skills

One way to think about the value of teleoperation system is to help provide demonstrations for teaching skills

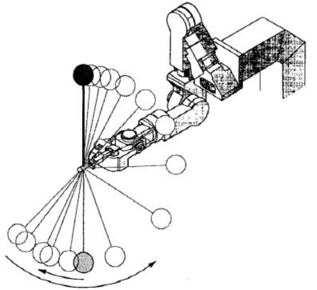
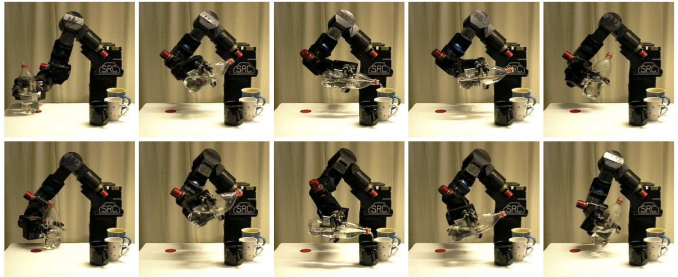


Figure 1: The SARCOS robot arm with a pendulum gripped in the hand. The pendulum axis is aligned with the fingers and with the forearm in this arm configuration.

Atkeson et al. 1997

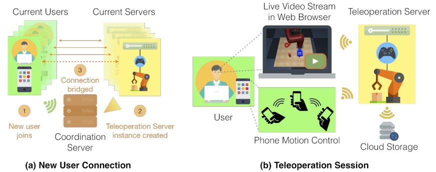
Figure 3: Humanoid robot learning a forehand swing from a human demonstration.

Ijspeert et al. 2003

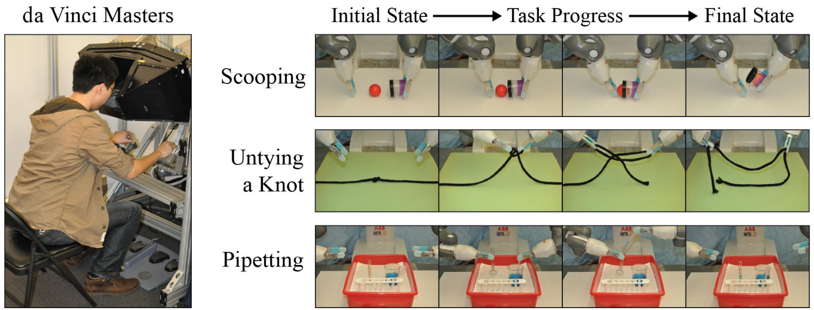


Pastor et al. 2009

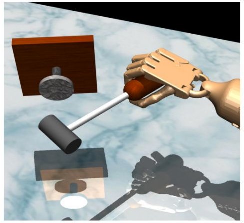
Dyrstad et al. 2018



RoboTurk: Mandlekar et al. 2018

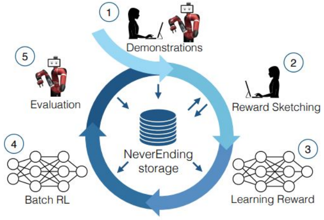


Liang et al. 2017



Rajeswaran et al. 2018

Zhang et al. 2018



Cabi et al. 2019

- We will now cover contemporary work on using teleoperation to teach skills to robots.
- An expert operator teleoperates a robot to collect demonstrations that show how to solve a particular task.
- The data is logged in the form or state (or observations), action pairs and used to learn the mapping from state (or observation) to the corresponding action.
- Various learning mechanisms: supervised learning (behaviour cloning), RL, batch RL etc.
- Learning is faster with demonstrations and the policies exhibit desired human-like behaviours.

Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations

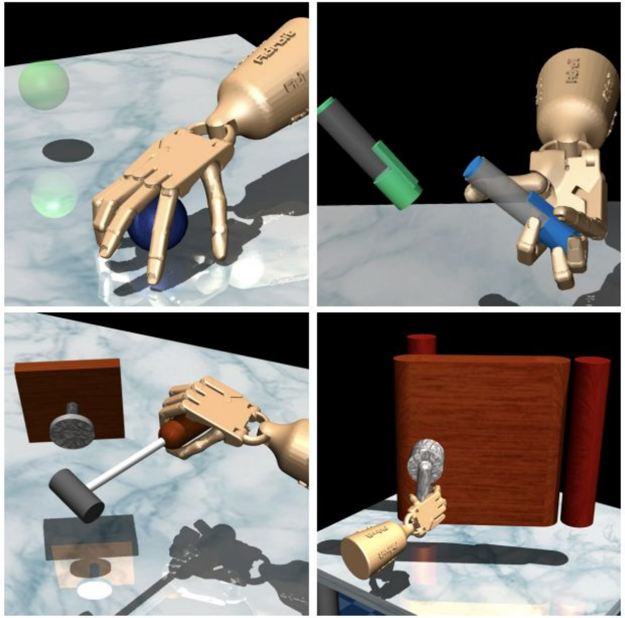


Figure 1: We demonstrate a wide range of dexterous manipulation skills such as object relocation, in-hand manipulation, tool use, and opening doors using DRL methods. By augmenting with human demonstrations, policies can be trained in the equivalent of a few real-world robot hours.

- Collects human demonstrations of robot hand performing these tasks in VR in MuJoCo.

- Demonstrates, in simulation, dexterous manipulation with high-dimensional human-like five-finger hands using model-free DRL.

- Shows that with a small number of human demonstrations, the sample complexity can be reduced dramatically and brought to levels which can be executed on physical systems.

- Policies trained with demonstrations are more human-like as well as robust to variations in the environment. Attributed this to human priors in the demonstrations which bias the learning towards more robust strategies.

Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations

What happens if we don’t use VR demonstrations? We can still learn but do the learned policies look natural?

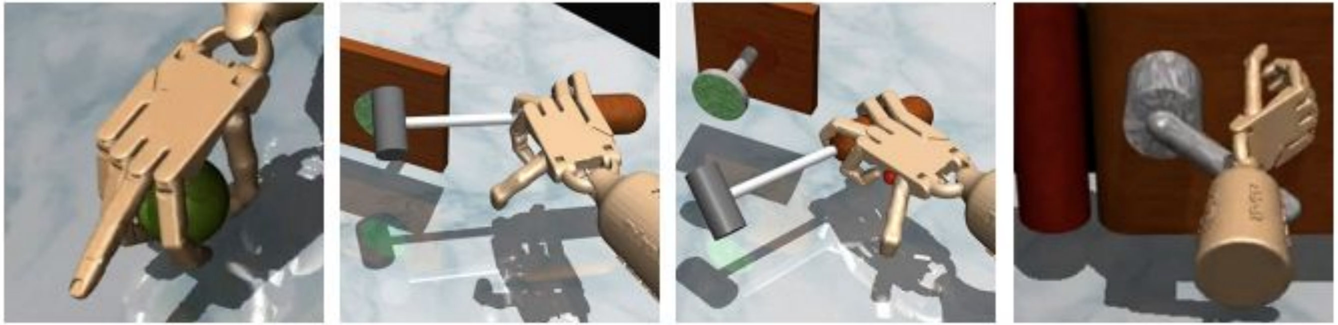


Figure 8: Unnatural movements observed in the execution trace of behavior trained with pure reinforcement leaning. From left to right: (a) unnatural, socially unacceptable, finger position during pick up. (b/c) unnatural grasp for hammer (d) unnatural use of wrist for unlatching the door.

This is precisely why we need human demonstrations to provide the bias in learning the policies.

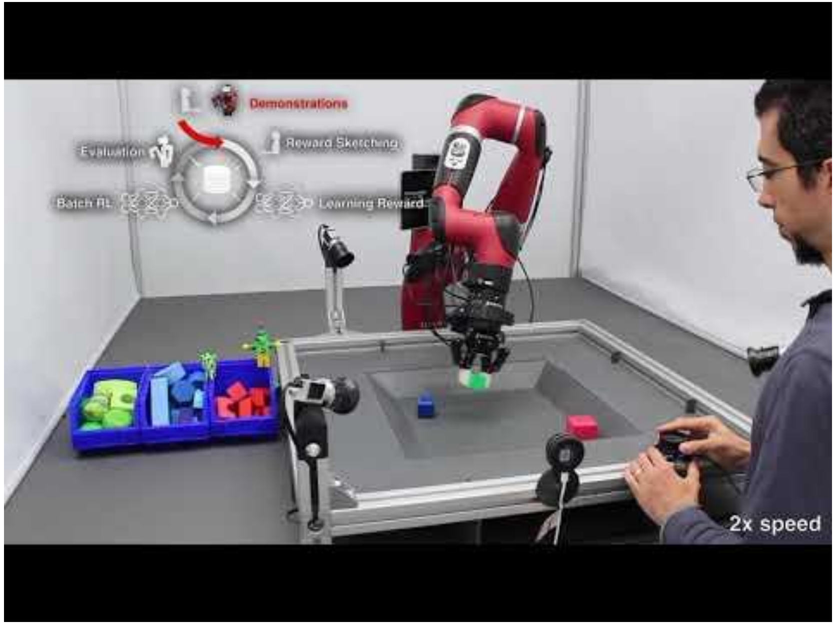
Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations, *rajeswaran et al. 2018*

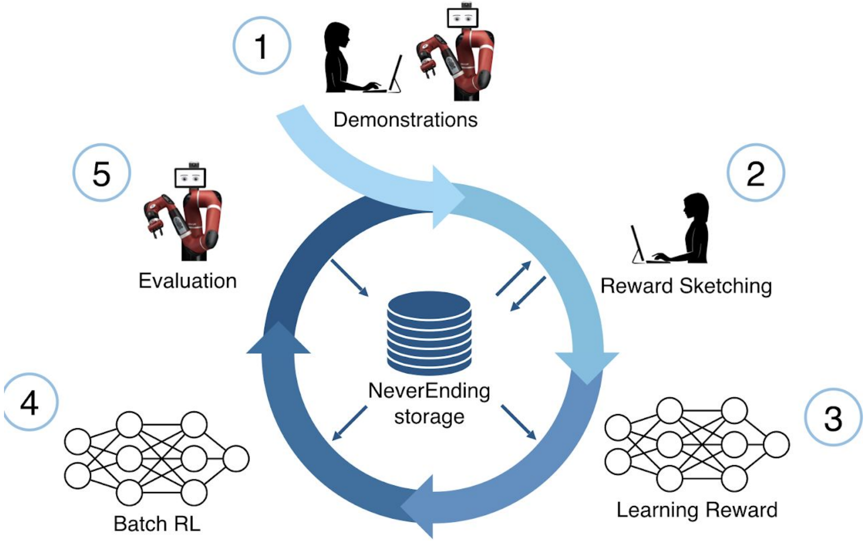
What happens if we don't use VR demonstrations? We can still learn but do the learned policies look natural?

This is precisely why we need human demonstrations to provide the bias in learning the policies.

Scaling data-driven robotics with reward sketching and batch reinforcement learning, *cabi et al. 2019*

Collecting teleoperation demonstrations to guide the learning and annotating the reward functions





https://sites.google.com/view/data-driven-robotics/

Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation, *mccarthy et al. 2018*

Fig. 1: Virtual Reality teleoperation in action

- Inexpensive teleoperation system that allows intuitive robotic manipulation and collection of high-quality demonstrations suitable for learning.

- With high-quality demonstrations, can imitation learning succeed in solving a wide range of challenging manipulation tasks using a practical amount of data?

- VR teleoperation system on a real PR2 robot using consumer-grade VR devices.

- For each task, < 30 minutes of demonstration data is sufficient to learn a successful policy, with the same hyperparameters and neural network architecture used across all tasks.

Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation, *mccarthy et al. 2018*

Neural network used to map the input RGB-D frames to end-effector pose in 6D (pos + rotation)

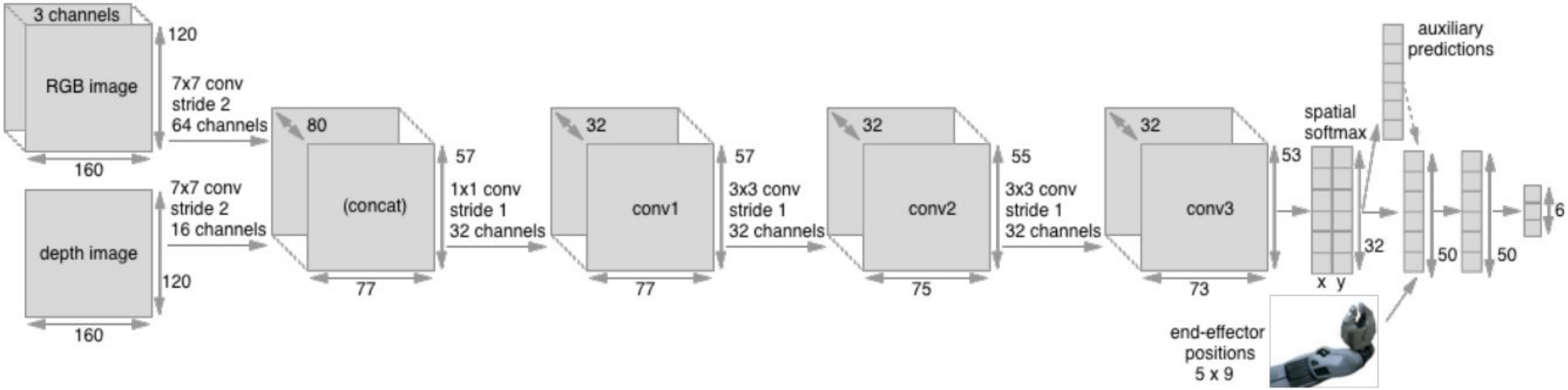


Fig. 3: Architecture of our neural network policies

They use basic supervised learning (also called behaviour cloning) to learn the policies. This work does not use RL.

Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation, *mccarthy et al. 2018*

Tasks used in this work

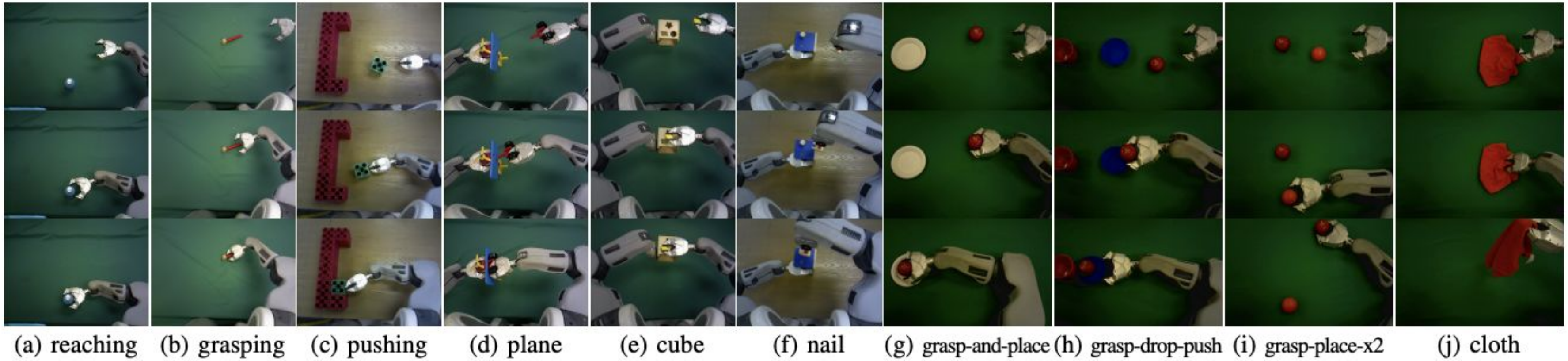
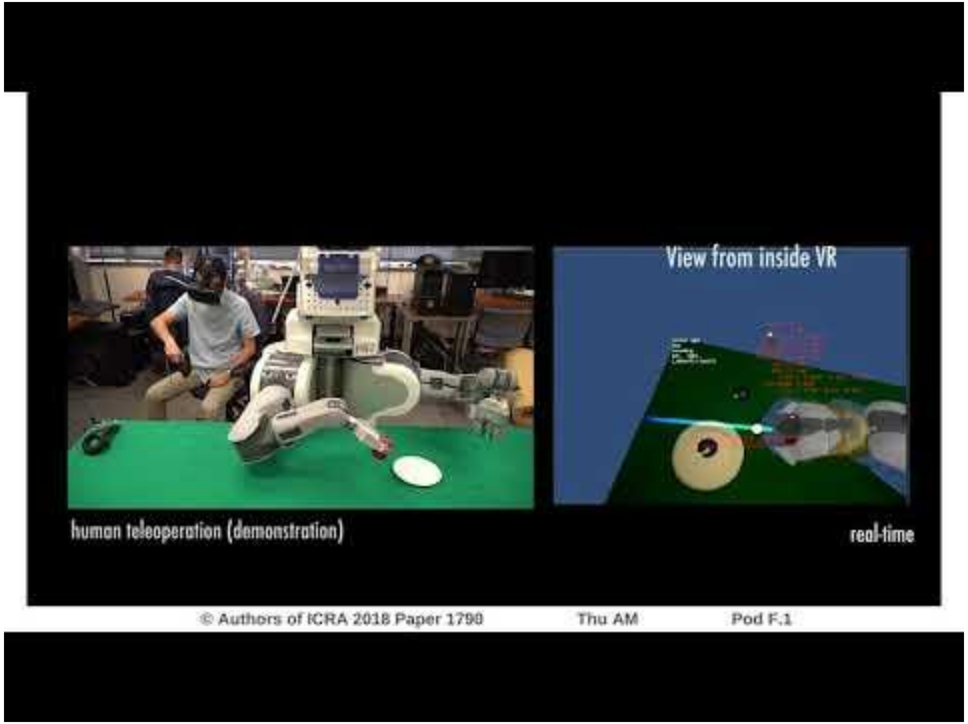


Fig. 4: Examples of successful trials performed by the learned policies during evaluation. Each column shows the image inputs I_t at $t = 0, \frac{T}{2}, T$ for the corresponding task.

Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation, *mccarthy et al. 2018*



ROBOTURK: A Crowdsourcing Platform for Robotic Skill Learning Through Imitation, *mandlekar et al. 2018*

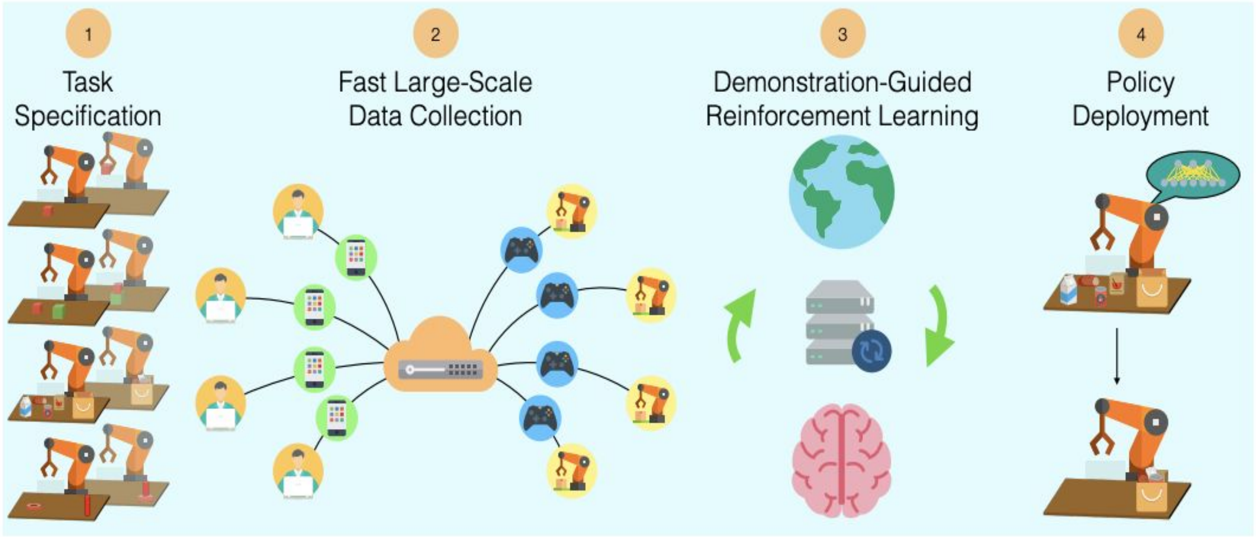
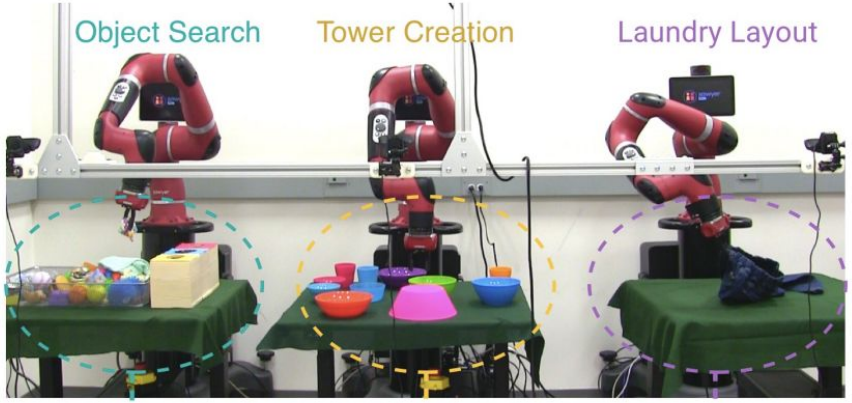


Figure 1: **System overview of ROBOTURK.** ROBOTURK enables quick imitation guided skill learning. Our system consists of the following major steps: 1) specifying a task, 2) collecting a large set of task demonstrations using ROBOTURK, 3) using demonstration-augmented reinforcement learning to learn a policy, and 4) deploying the learned skill in the domain of interest.

- Dataset consisting of over 2200 task demonstrations, amounting to 137 hours of data collected in 20 hours of system usage with contracted workers.

- Data collected through ROBOTURK enables policy learning on challenging manipulation tasks with sparse rewards and that using larger quantities of demonstrations during policy learning provides benefits in terms of both learning consistency and final performance.

ROBOTURK: A Crowdsourcing Platform for Robotic Skill Learning Through Imitation, *mandlekar et al. 2018*



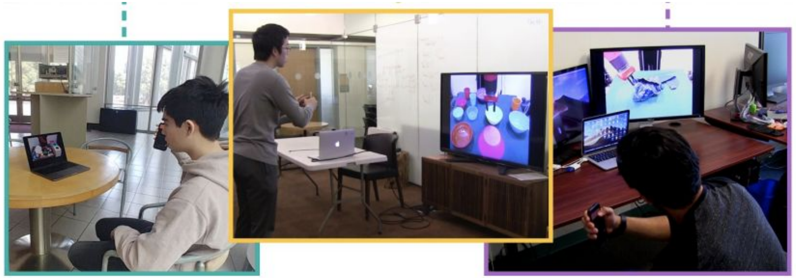
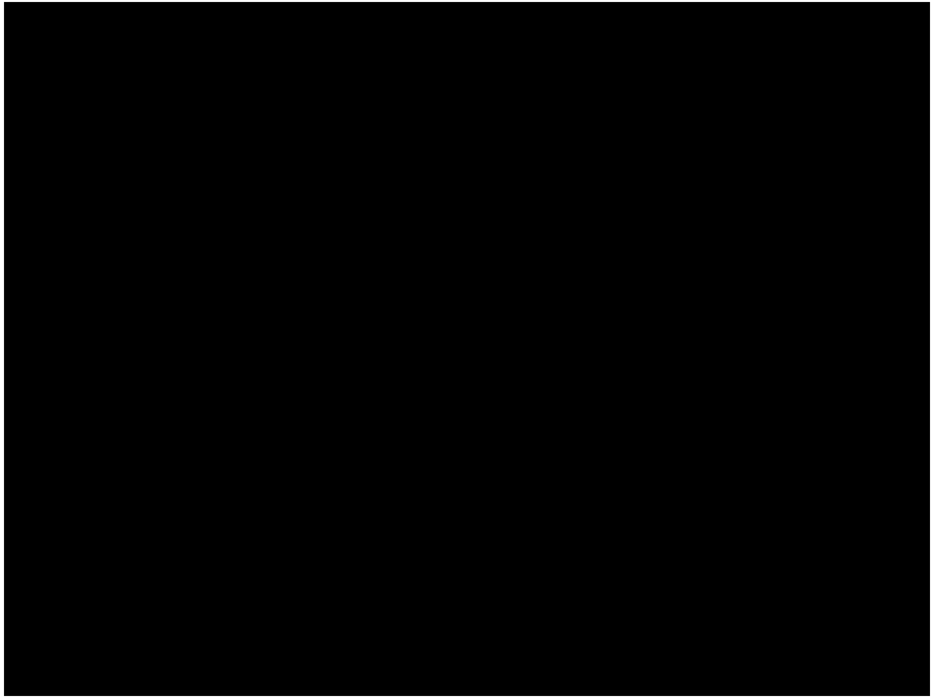


Fig. 1: **Collecting data on physical robot arms with the RoboTurk platform.** To collect task demonstrations, users connect to our platform from remote locations using a web browser and use their smartphone as a motion controller to move the physical robot arm in free space. Users are provided a video stream of the robot workspace in their web browser.

TABLE I: **Dataset Comparison.** We compare our dataset to similar robot datasets collected via human supervision in prior work. Items marked with $*$ are estimates that were extrapolated using other reported information, and interfaces marked with † are not real-time.

Name	Interface	Task	Avg. Task Length (sec)	Number of Demos	Total Time (hours)
JIGSAWS[10]	daVinci	surgery	60*	103	1.66
Deep Imitation [43]	VR	pick, grasp, align	5*	1664	2.35
DAML[42]	Human demos	pick, place, push	5*	2941	4.08
MIME[34]	Kinesthetic	pick, place, push, pour	6*	8260	13.7*
PbD[8]	GUI†	pick, place	207*	465	25.8*
Roboturk-Real (Our)	iPhone AR	long horizon object manip	186	2144	**111.25**

ROBOTURK: A Crowdsourcing Platform for Robotic Skill Learning Through Imitation, *mandlekar et al. 2018*

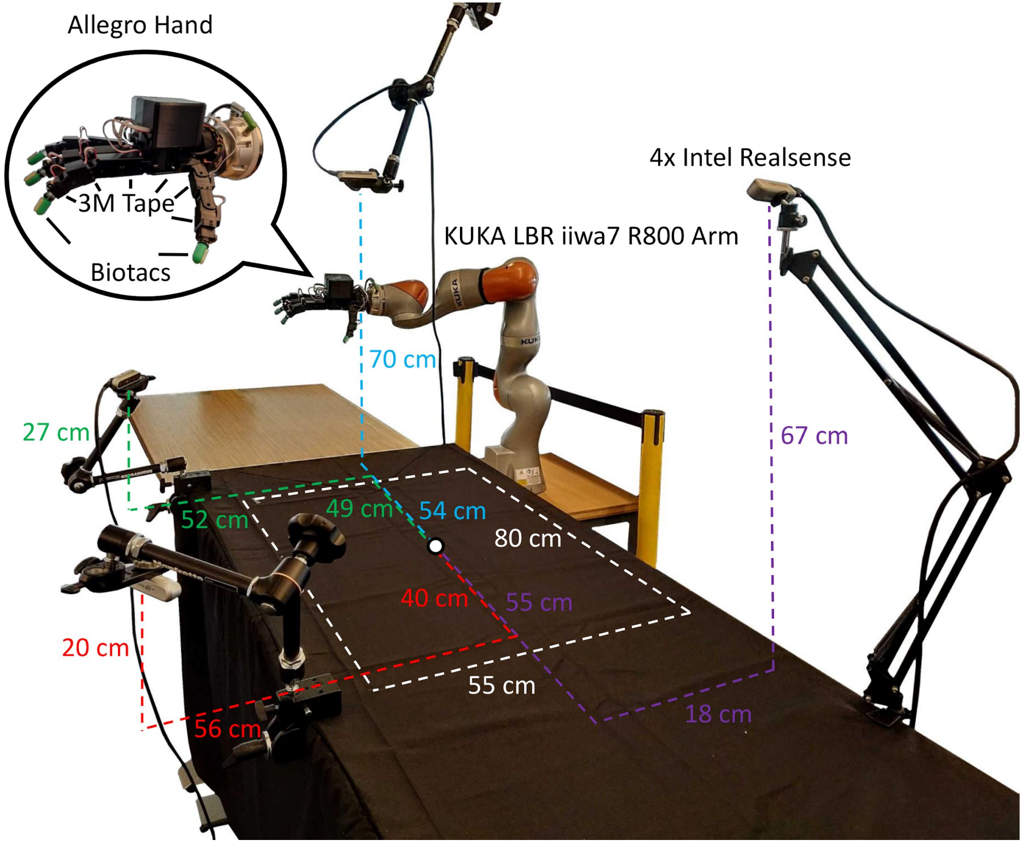


DexPilot: Vision based teleoperation of dexterous robotic hand-arm system, *handa et al. 2019*

- Markerless, glove-free and entirely vision-based teleoperation system that dexterously articulates a highly actuated robotic hand-arm system with direct imitation.

- Demonstration of teleoperation system on a wide variety of tasks particularly involving fine manipulations and dexterity, e.g., pulling out paper currency from wallet and grasping two cubes with four fingers

System Overview



The system is housed in a studio next to the robot

- We use Intel D415 cameras and extrinsically calibrate them apriori using standard calibration toolbox.

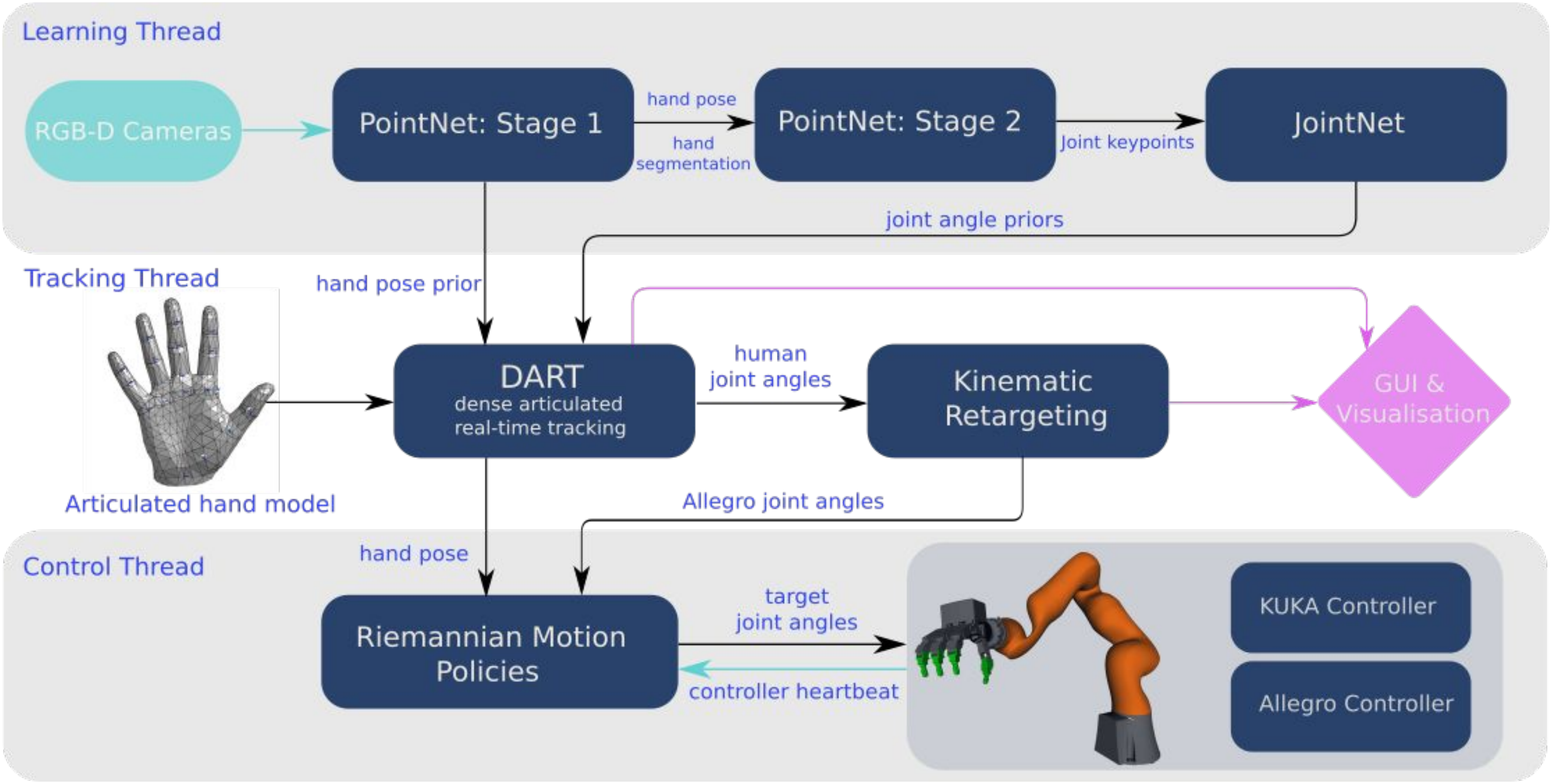
- The hand moves over the table covered in black cloth.

- Given the calibration we can project the point clouds from all the cameras into one global reference frame.

- We assume line of sight driven teleoperation. The user doesn't wear any VR headset.

- Because it's line of sight driven, the user has to be standing in close proximity to the robot for better depth perception.

Architecture



Each thread runs on a different computer: we need 3 computers to run this, although 2 might suffice as well.

DexPilot: Some remarks on teleoperation

- Large majority of **these tasks are complicated** and even simulators struggle to simulate them. If we can teleop something, we have a lot more hope of doing this task via imitation. **Defining rewards is almost impossible** for many of the tasks we do.

- The fact the the **system works despite any tactile feedback** says a lot about how our brains are able adapt to any tasks with vision only input.
    - **HaptX,** a company that builds robot hand-arm teleoperation systems with tactile feedback, **told us that in their system the user only needed tactile feedback at the start for a few trials** but they naturally adapted to the tasks pretty quickly from then on without needing any tactile feedback. It is quite an out-of-body experience.

- When the user does reattempts *e.g.* to pick up a dropped object, the data still provides valuable information about **recovery from failures** which the robots can learn from.

BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning, *zhang & irpan et al. 2021*

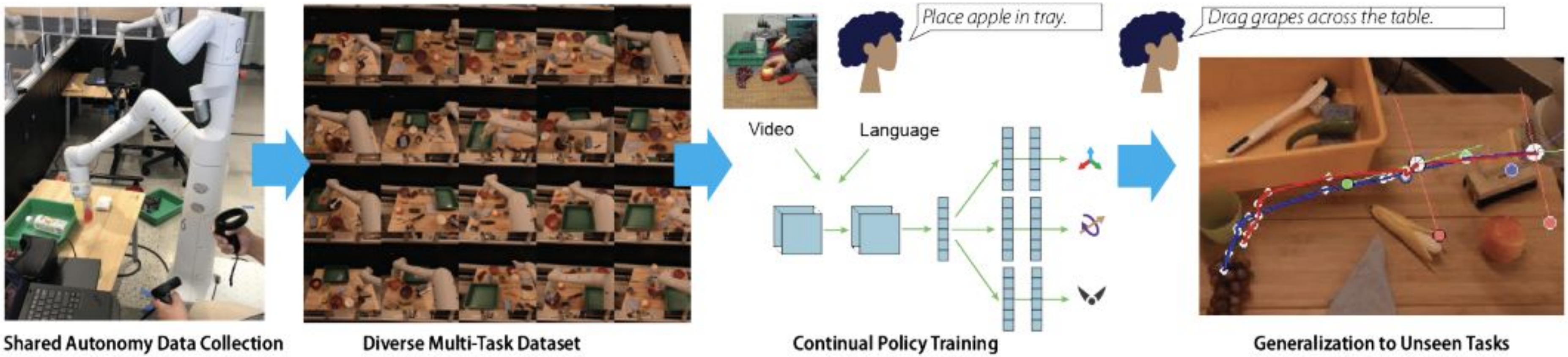


Figure 1: Overview of BC-Z. We collect a large-scale dataset (25,877 episodes) of 100 diverse manipulation tasks, and train a 7-DoF multi-task policy that conditions on task language strings or human video. We show this system produces a policy that is capable of generalizing zero-shot to new unseen tasks.

BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning, *zhang & irpan et al. 2021*

- Interactive and flexible imitation learning system that can learn from both demonstrations and interventions and can be conditioned on different forms of information that convey the task, including pretrained embeddings of natural language or videos of humans performing the task.

- When scaling data collection on a real robot to more than 100 distinct tasks, we find that this system can perform 24 unseen manipulation tasks with an average success rate of 44%, without any robot demonstrations for those tasks.

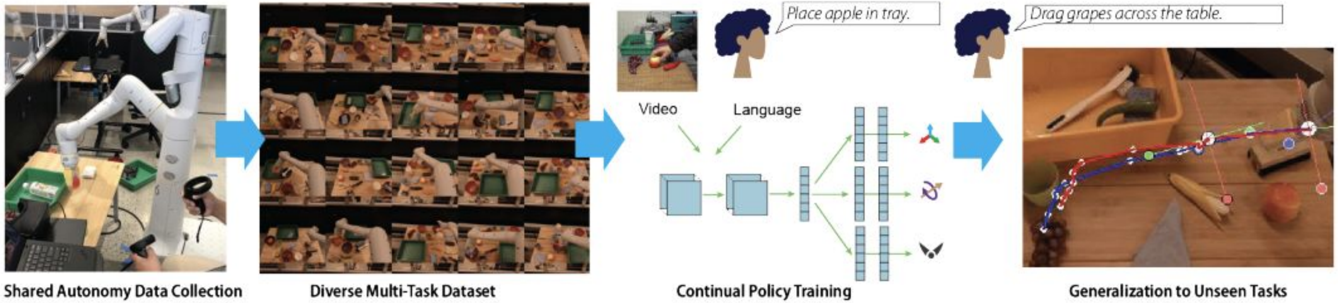


Figure 1: Overview of BC-Z. We collect a large-scale dataset (25,877 episodes) of 100 diverse manipulation tasks, and train a 7-DoF multi-task policy that conditions on task language strings or human video. We show this system produces a policy that is capable of generalizing zero-shot to new unseen tasks.

BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning, *zhang & irpan et al. 2021*

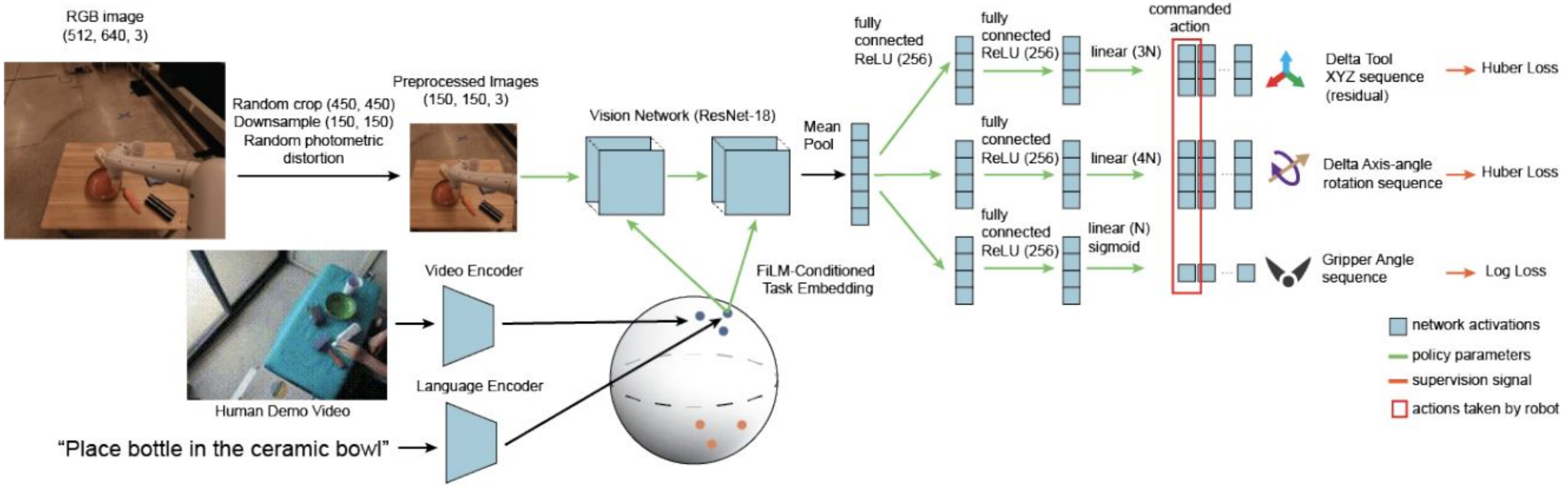
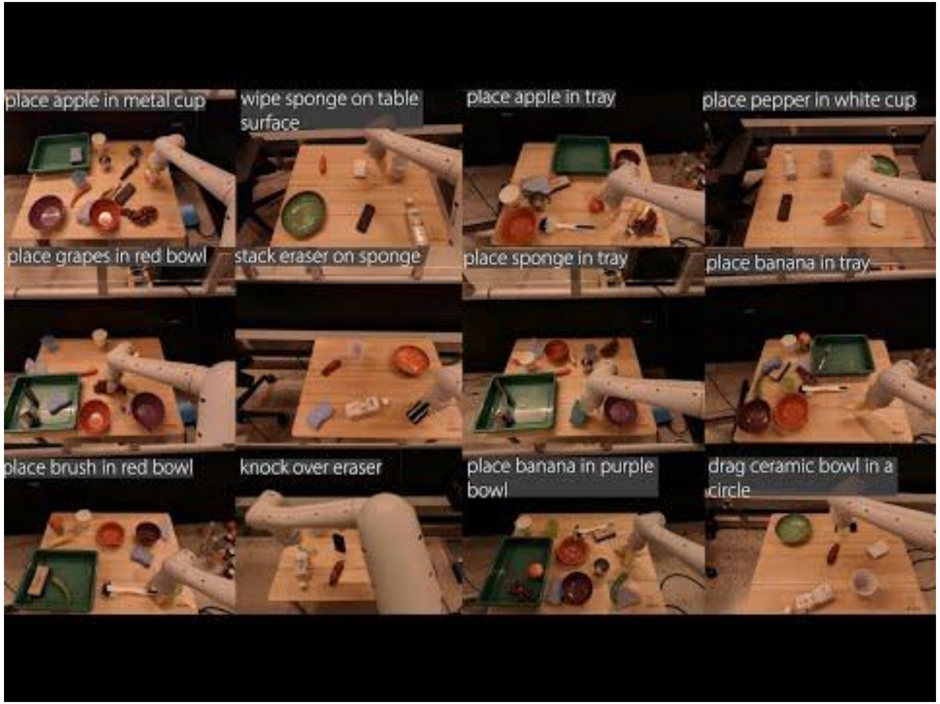


Figure 3: BC-Z network architecture. A monocular RGB image from the head-mounted camera is passed through a ResNet18 encoder, then through a two-layer MLP to predict each action modality (delta XYZ, delta axis-angle, and gripper angle). FiLM layers [47] condition the architecture on a task embedding z computed from language w_ℓ or video w_h.

BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning, *zhang & irpan et al. 2021*

Table 2: Success rates for zero-shot (language) and few-shot (video) generalization to tasks not in the training dataset. The first 4 tasks only use objects from the 79-task family. The remaining tasks mix objects between the 21-task and 79-task families, requiring further generalization. Numbers in parentheses are 1 unit standard deviation. The language conditioning generalizes to several holdout tasks, whereas the video conditioning shows promise on tasks that do not mix objects between task families. Overall performance improves slightly with fewer distractor objects.

Skill	Held-out tasks (no demos during training)	Lang-conditioned (1 distractor)	Lang-conditioned (4-5 distractors)	Video-conditioned (4-5 distractors)
pick-place	'place sponge in tray'	83% (6.8)	82% (9.2)	22% (2.2)
	'place grapes in red bowl'	87% (6.2)	75% (10.8)	12% (7.8)
	'place apple in paper cup'	30% (8.4)	33% (12.2)	14% (7.8)
pick-wipe	'wipe tray with sponge'	40% (8.9)	0% (0)	28% (10.6)
pick-place	'place banana in ceramic bowl'	50% (15.8)	75% (9.7)	7.5% (4.2)
	'place bottle in red bowl'	50% (15.8)	75% (9.7)	0% (0)
	'place grapes in ceramic bowl'	70% (14.5)	70% (10.3)	0% (0)
	'place bottle in table surface'	0	50% (11.2)	5% (3.5)
	'place white sponge in purple bowl'	70% (14.9)	45% (11.2)	0% (0)
	'place white sponge in tray'	50% (15.8)	40% (11.0)	0% (0)
	'place apple in ceramic bowl'	30% (14.5)	20% (8.9)	0% (0)
	'place bottle in purple bowl'	30% (14.5)	20% (8.9)	0% (0)
	'place banana in ceramic cup'	10% (9.5)	0% (0)	0% (0)
	'place banana on white sponge'	40% (15.5)	0% (0)	0% (0)
	'place metal cup in red bowl'	0% (0)	0% (0)	0% (0)
grasp	'pick up grapes'	70% (14.5)	65% (10.7)	0% (0)
	'pick up apple'	20% (12.7)	55% (11.2)	5% (3.5)
	'pick up towel'	50% (15.8)	42.8% (18.7)	0% (0)
	'pick up pepper'	50% (15.8)	35% (10.7)	12.5% (5.2)
	'pick up bottle'	40% (15.5)	30% (10.3)	17.5% (6.0)
	'pick up the red bowl'	30% (14.5)	0% (0)	0% (0)
pick-drag	'drag grapes across the table'	0% (0)	14% (13.2)	0% (0)
pick-wipe	'wipe table surface with banana'	0% (0)	10% (6.7)	0% (0)
	'wipe tray with white sponge'	20% (12.7)	0% (0)	0% (0)
	'wipe ceramic bowl with brush'	10% (9.49)	0% (0)	0% (0)
push	'push purple bowl across the table'	50% (15.8)	30% (10.3)	0% (0)
	'push tray across the table'	30% (14.5)	25% (9.7)	0% (0)
	'push red bowl across the table'	60% (15.5)	0% (0)	0% (0)
	Holdout Task Overall	38%	32%	4%



CLIPort: What and Where Pathways for Robotic Manipulation, *Shridhar et al, 2021*

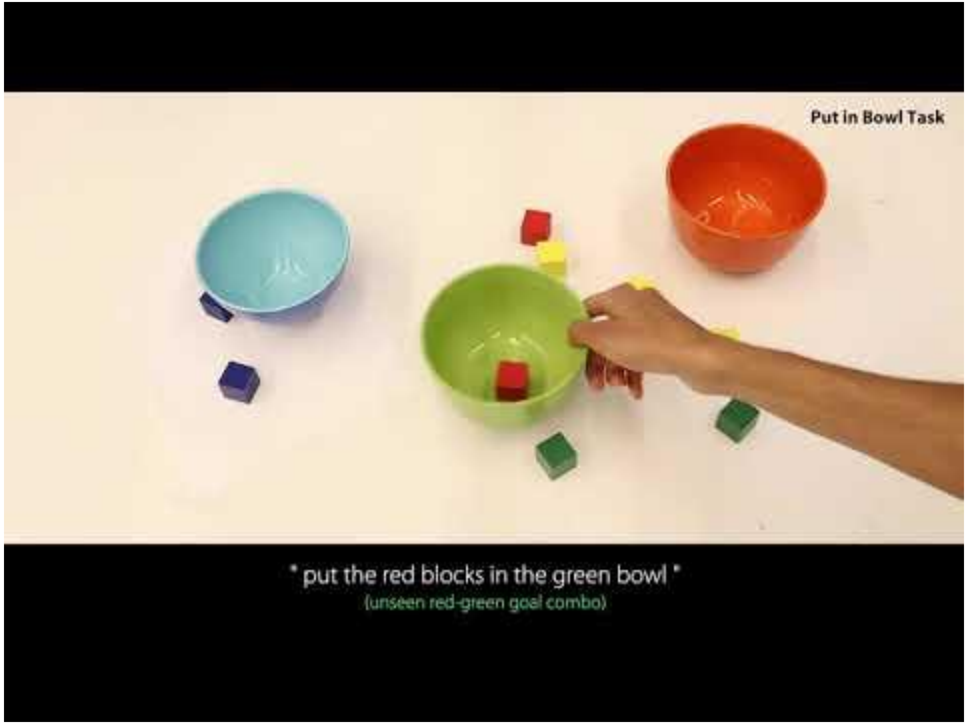


Figure 8. Real-Robot Experimental Setup.

Closing Remarks

- Teleoperation is a great way to direct the behaviour of policies to enable human-like operations.

- Although time consuming and not scalable, teleoperation still offers a way to speed up the learning process.

- Better teleoperation tools and systems will enable faster and scalable data collection.

- Policies should copy high-level behaviours than low-level trajectories.

Thank you