Cross-Embodiment Robotic Manipulation: Unifying Grippers Across Robots

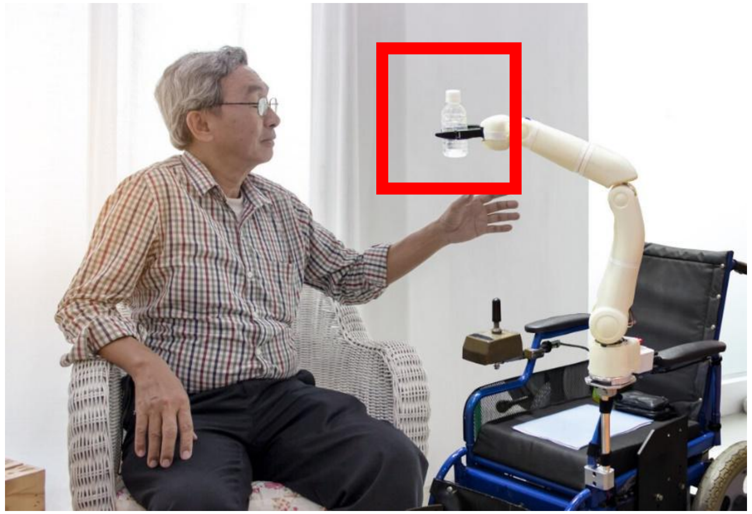
Yu Xiang
Assistant Professor
Intelligent Robotics and Vision Lab
The University of Texas at Dallas

10/30/2025, 10/31/2025

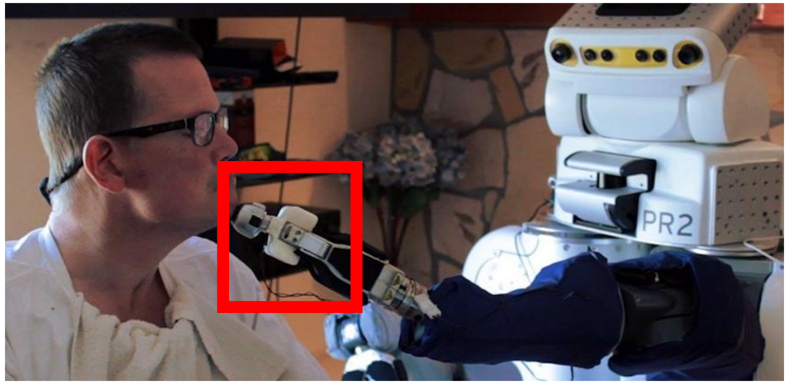
Robotics and AI Institute, Brown University

Future Intelligent Robots in Human Environments

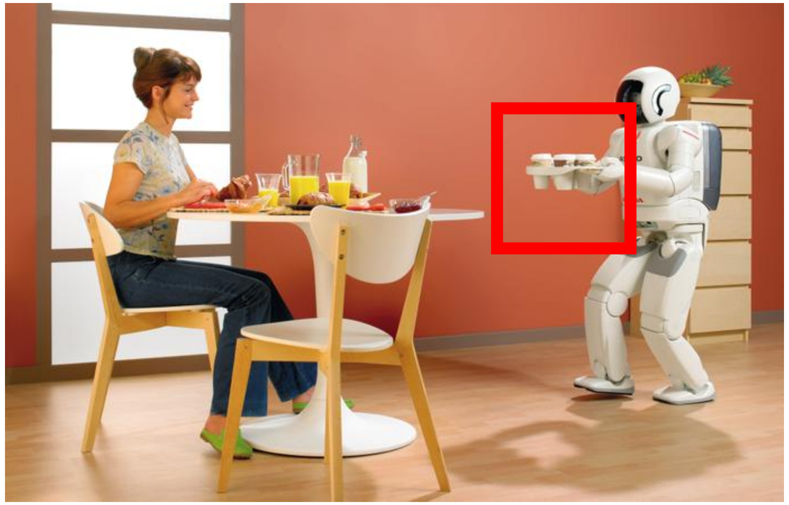
Manipulation



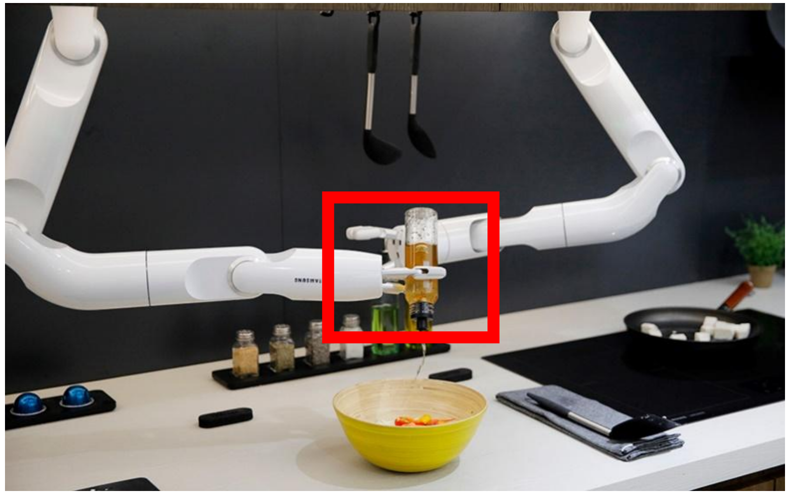
Senior Care



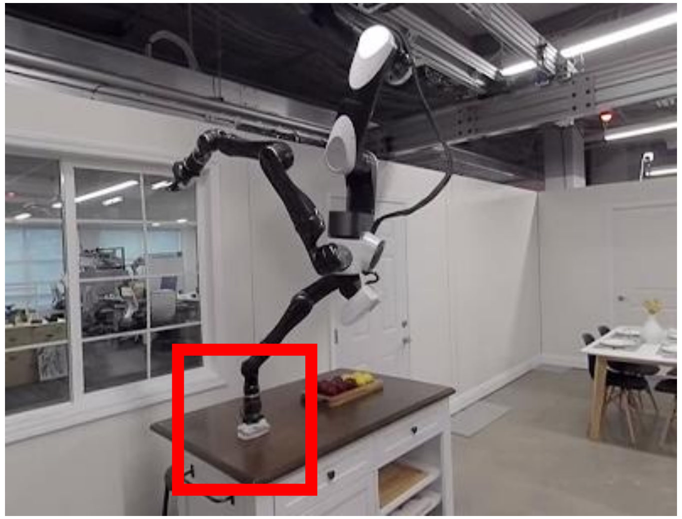
Assisting



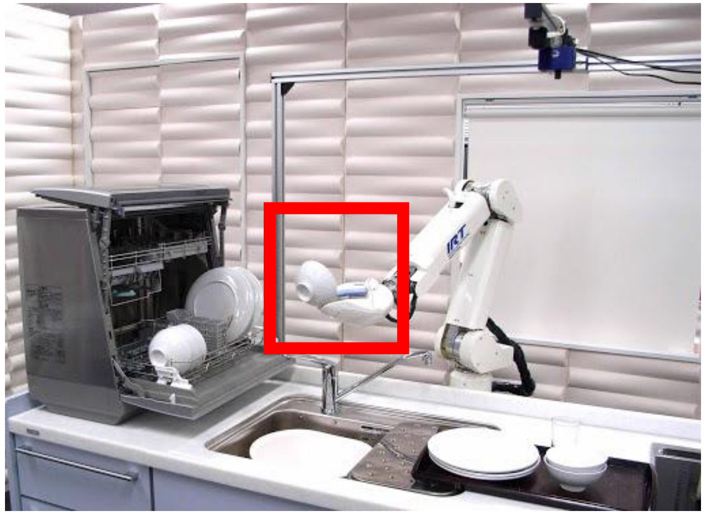
Serving



Cooking



Cleaning



Dish washing

We will have many different robots

Humanoid

Boston Dynamics Atlas

Unitree G1

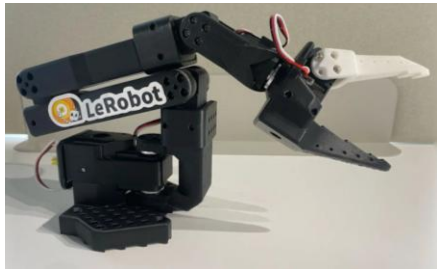
Figure

XPeng Iron

1X

Manipulator

Franka Emika



SO101 arm

Kuka arm

Mobile Manipulator

Fetch

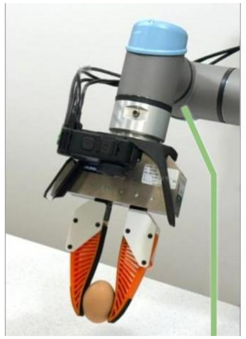
Mobile AI

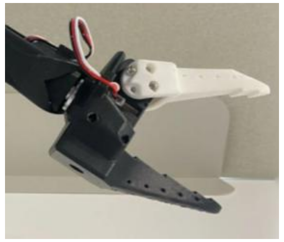
We will have many different grippers/hands


Panda


Fetch


Sawyer


UMI


SO101


Barrett

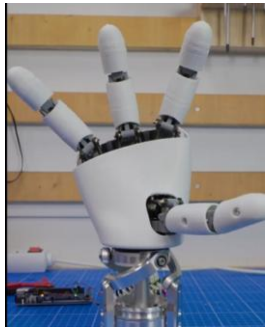

Robotiq


Atlas

Jaco


Allegro


Leap


LeRobot


Shadow


INSIPRE


PAL Hey5 hand


Sharpa

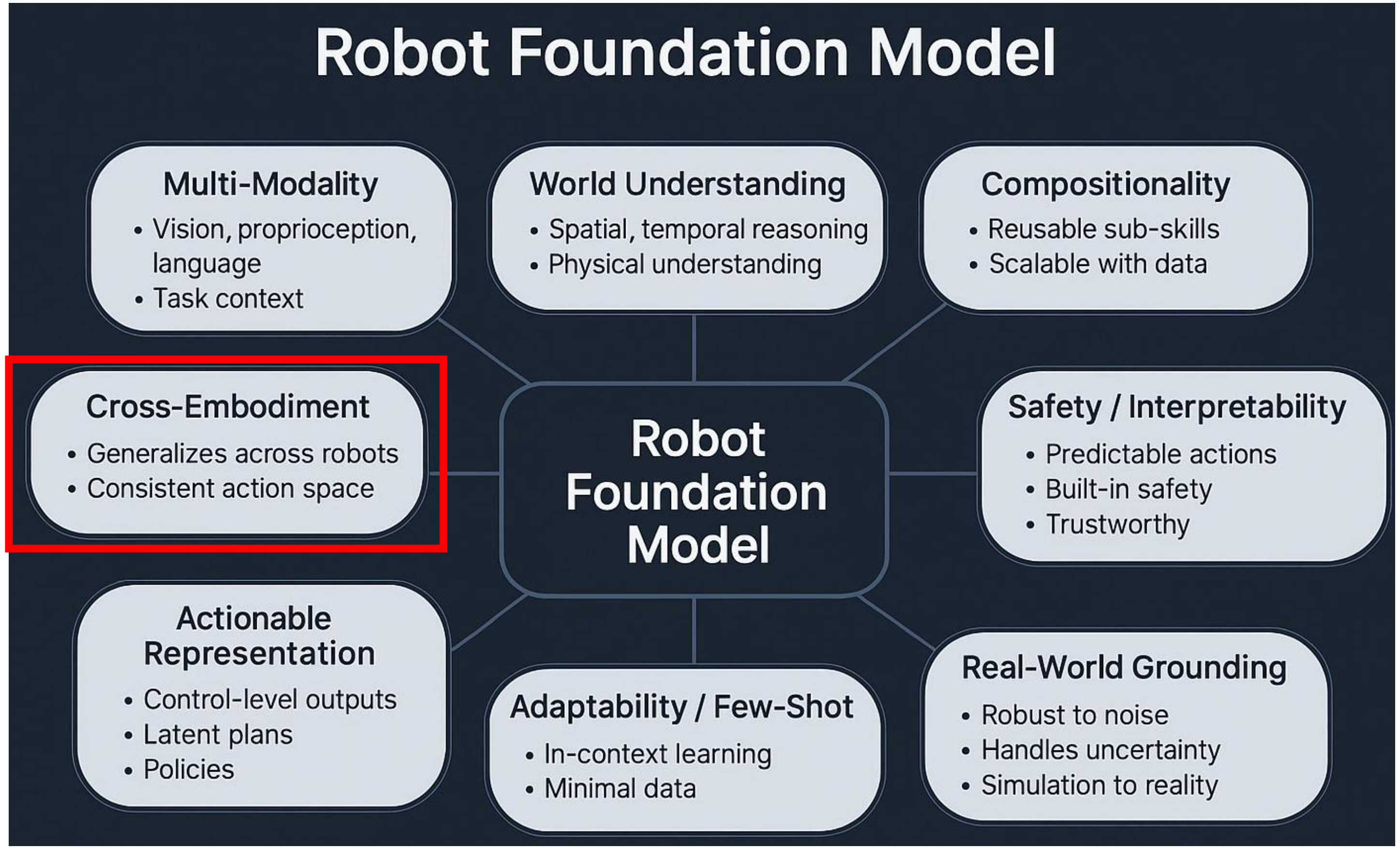

Aero Hand Open

4

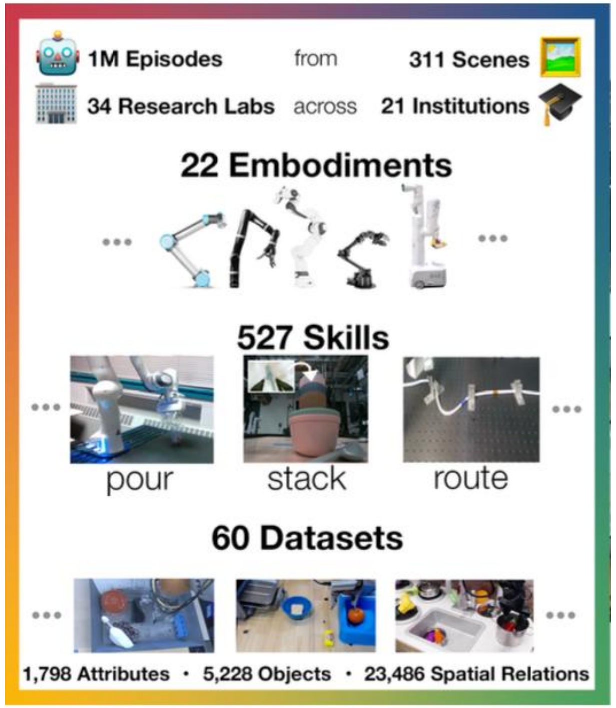
How to make all these robots work?

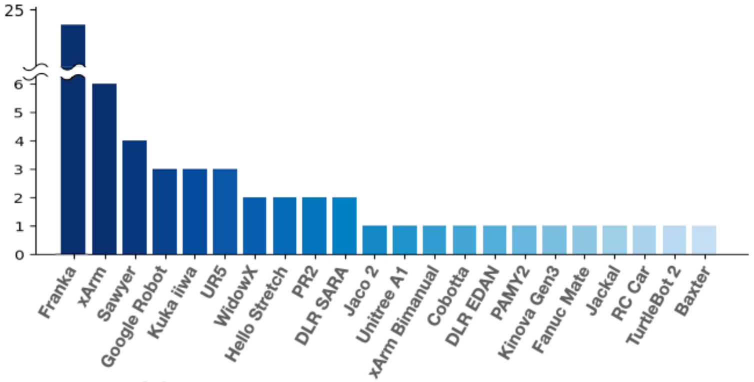
- Robot Foundation Model
  - A **foundation model** in AI refers to a **large, pre-trained model** that serves as a *base* (or "foundation") for building many downstream applications and specialized models.

Our recent focus



Current Robot Data

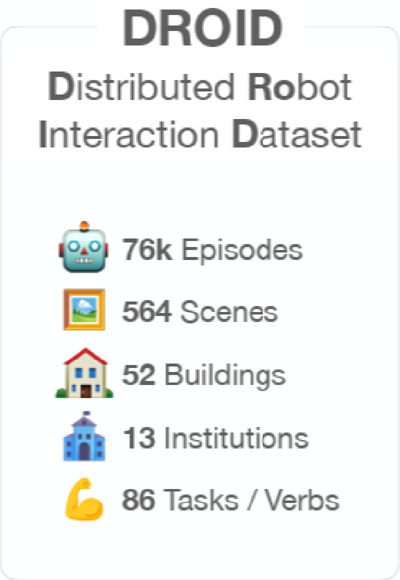


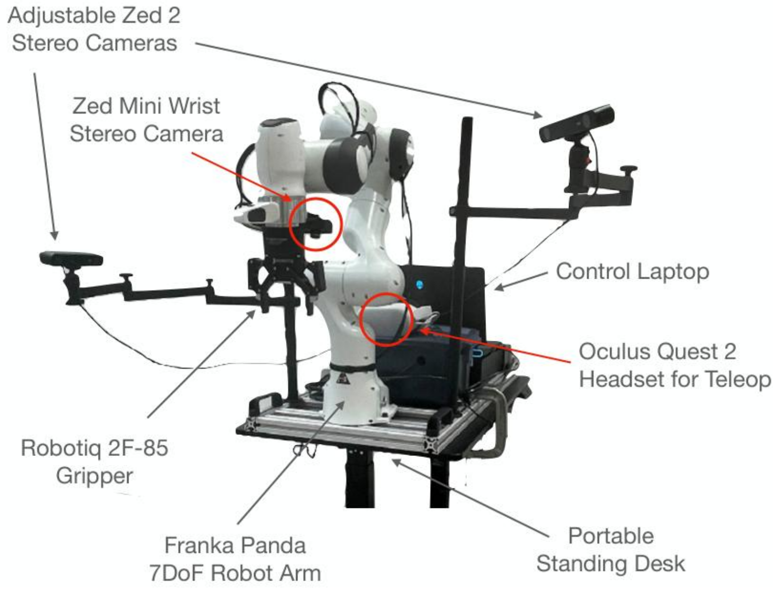


(a) # Datasets per Robot Embodiment

Open X-Embodiment

- Franka
- xArm
- Sawyer
- Google Robot
- Kuka iiwa
- UR5
- WidowX
- Hello Stretch
- PR2
- DLR SARA
- Jaco 2
- Unitree A1
- xArm Bimanual
- Cobotta
- DRL EDAN (5-finger)
- PAMY2
- Kinova Gen3
- Fanuc Mate
- Jackal
- RC Car
- TurtleBot 2
- Baxter



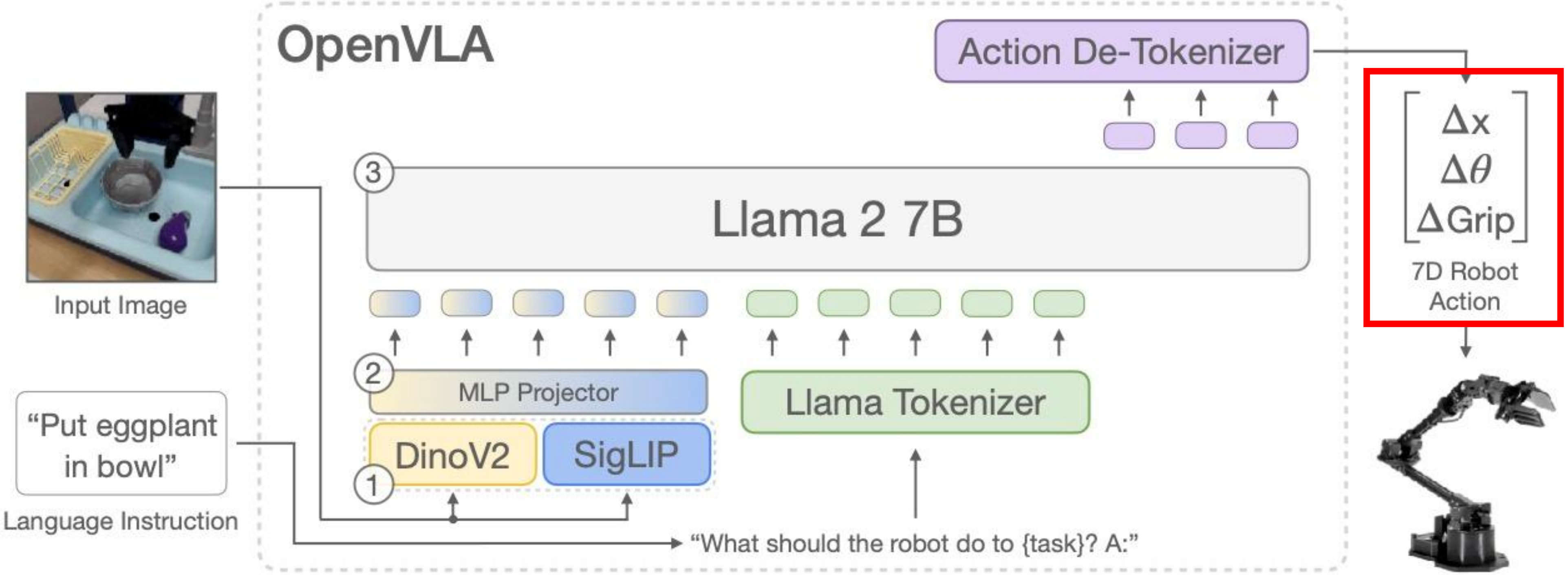


The DROID dataset

Very biased to 2-finger grippers

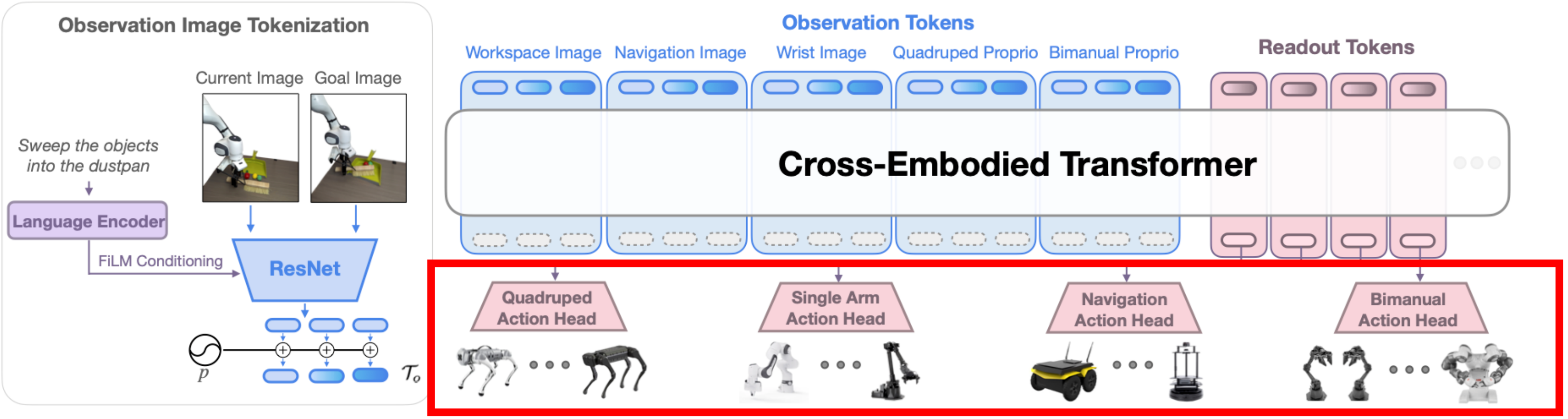
Current Model Architecture

Action
- Gripper pose for two-finger grippers
- Cannot to used for multi-finger hands (no hand joints)

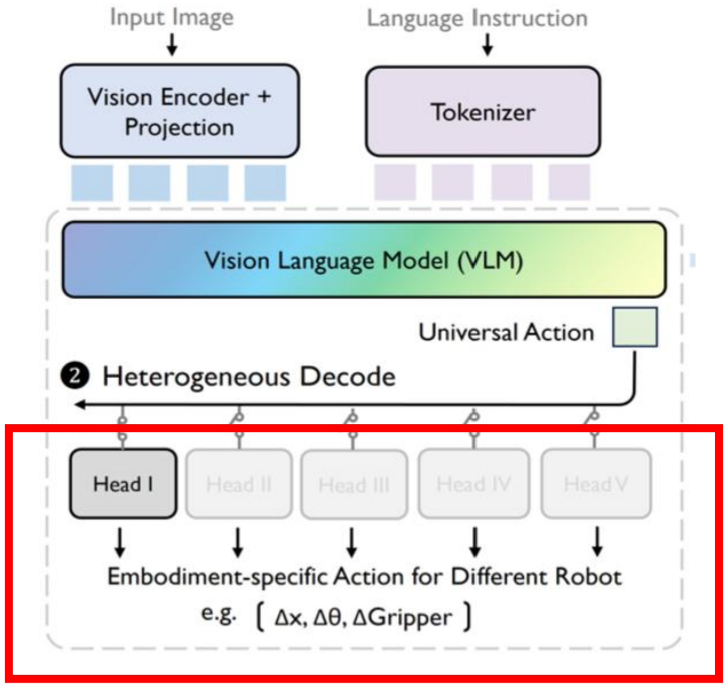


OpenVLA: An Open-Source Vision-Language-Action Model. Kim et al., 2024.

Cross-Embodiment Model Architecture



CrossFormer: Scaling Cross-Embodied Learning for Manipulation, Navigation, Locomotion, and Aviation. Doshi et al., CoRL, 2024.

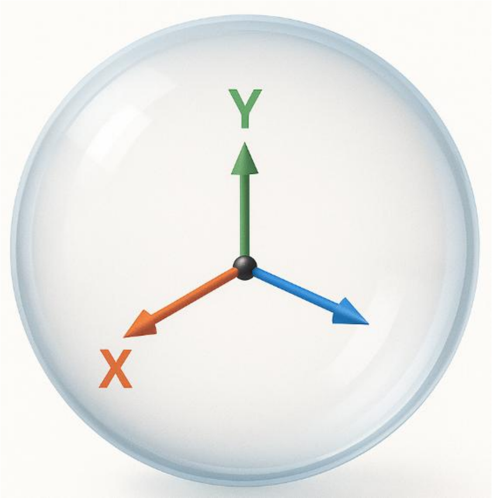


One Action head for each robot type

- Given a new robot, one new head needs to be trained

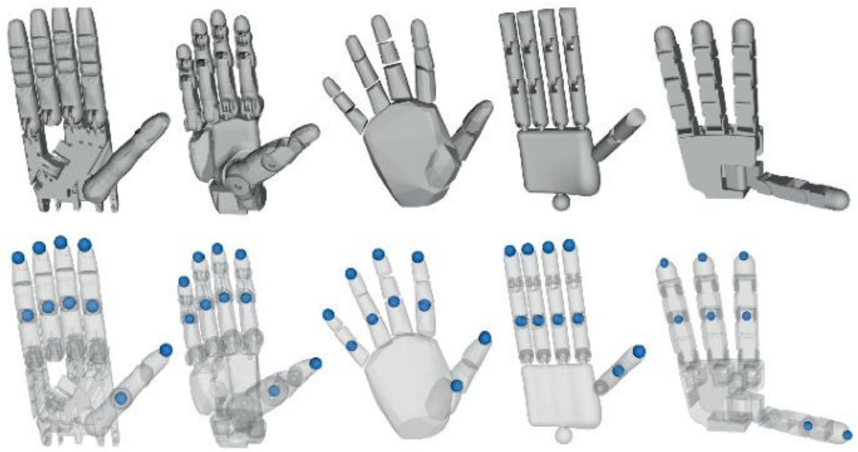
Universal Actions for Enhanced Embodied Foundation Models. Zheng et al., CVPR, 2025.

Can we find a unified action space for different robot grippers?

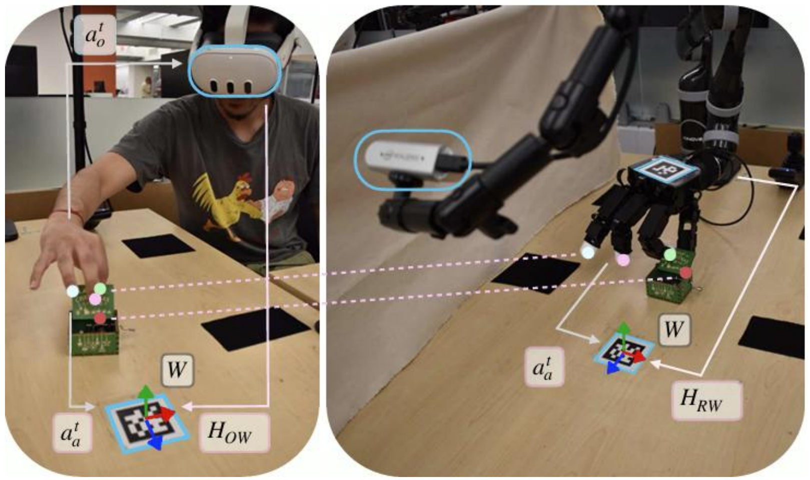


Human Hand
(human data)

Previous work: manual alignment of grippers (retargeting)



Learning Cross-hand Policies for High-DOF Reaching and Grasping (She et al., 2024)



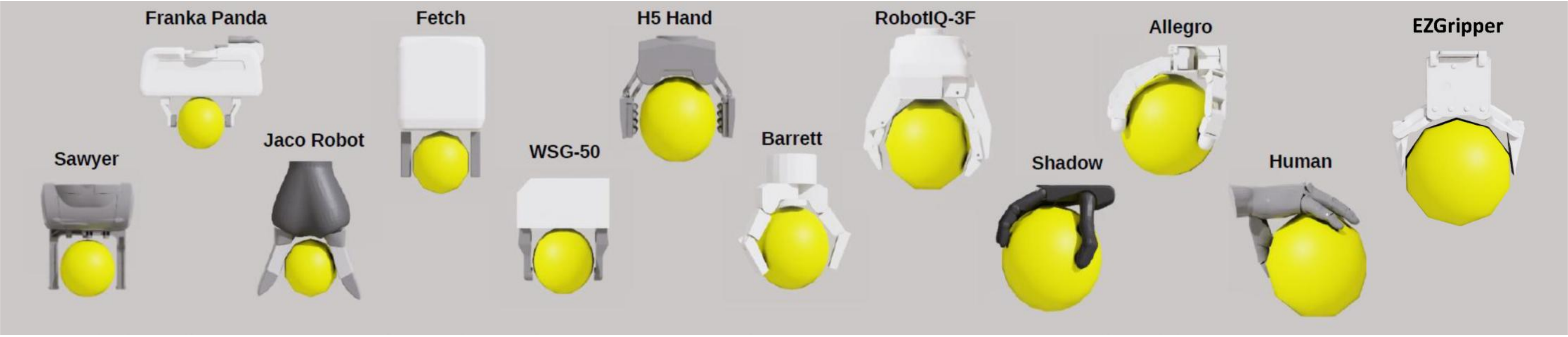
HuDOR, Guzey et al. NYU 2025

https://object-rewards.github.io/

- Manual mapping or hand-designed correspondence
- Hard to deal with different number of fingers
- Cannot handle unseen grippers

Our idea: Let’s use a sphere to align grippers

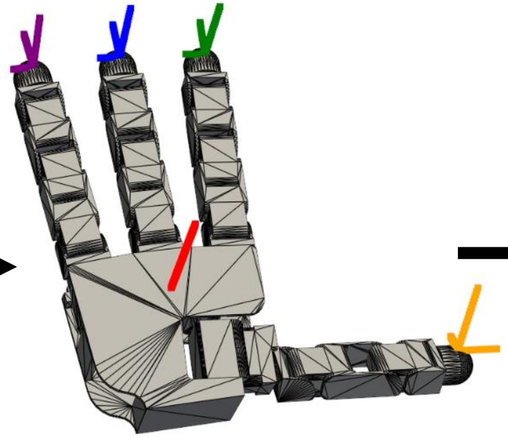
- Because any hand can grasp a sphere! (otherwise, it might not be that useful for manipulation)
- Spheres have some good properties for control (you will see)



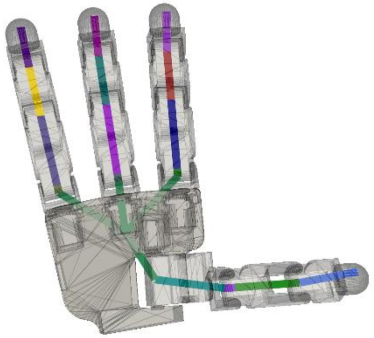
A unified action space for different robot grippers

- Sphere creation

Hand URDF

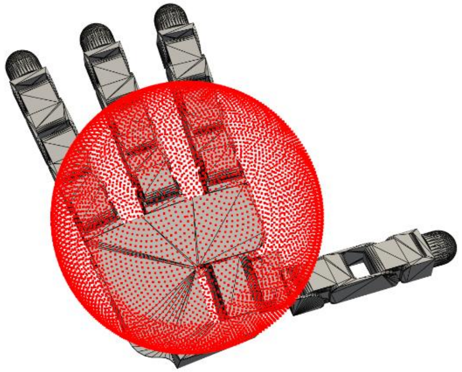


Hand URDF
- Frames for palm center and fingertips

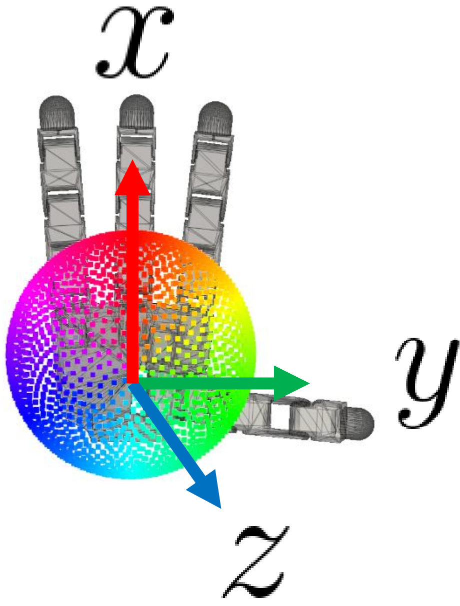


$$l = \frac{l_1 + l_2 + l_3 + l_4}{4}$$

Radius $r = l\frac{2}{\pi}$



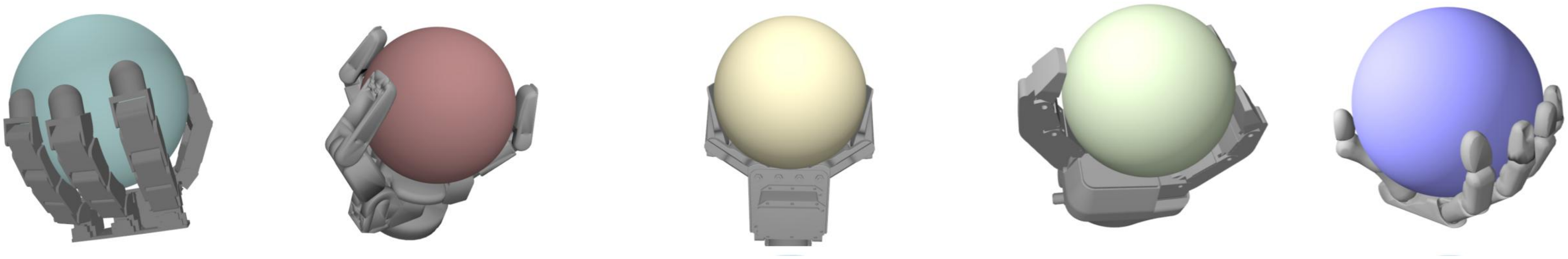
Sphere center above the palm center by r



x-axis towards the middle finger

A unified action space for different robot grippers

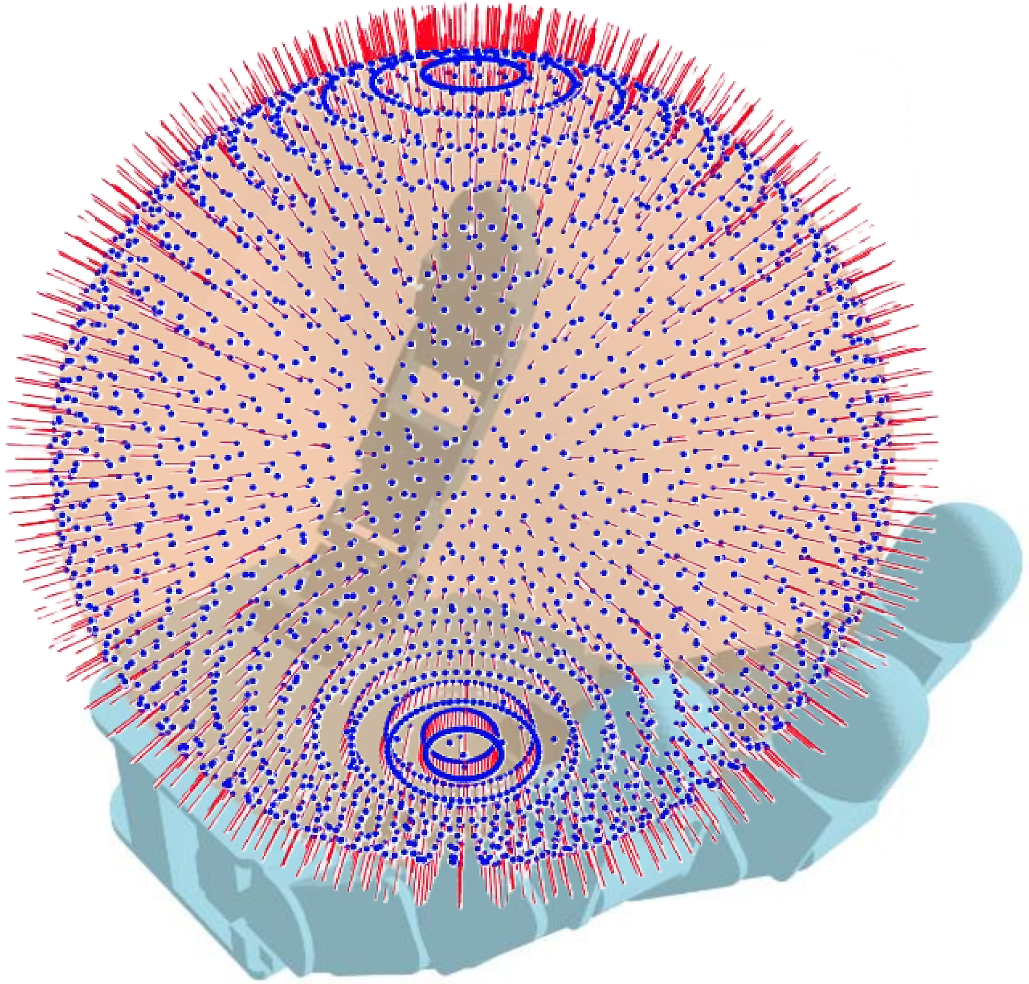
- Sphere creation applies to different grippers

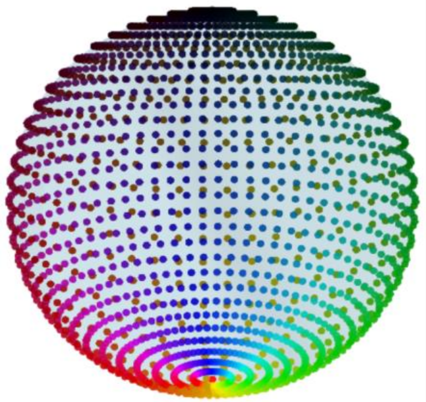


Human Hand

A unified action space for different robot grippers

- Map spherical coordinates to the gripper

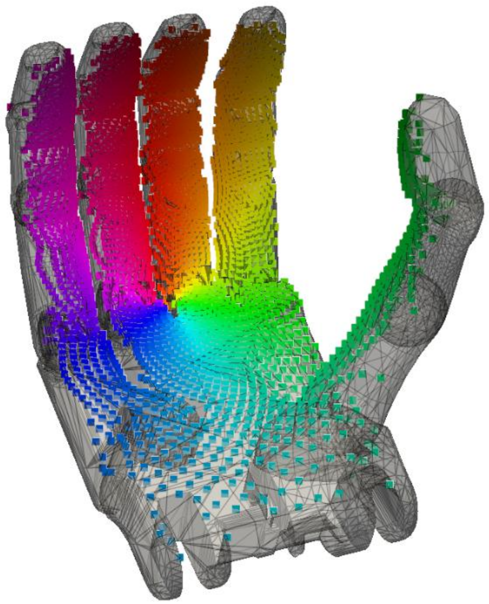


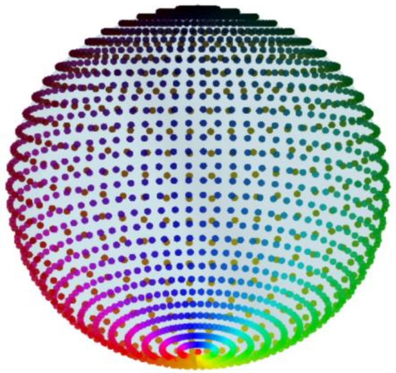


(λ, ϕ)

A unified action space for different robot grippers

- Map spherical coordinates to the gripper (a representation of the gripper)





(λ, ϕ)

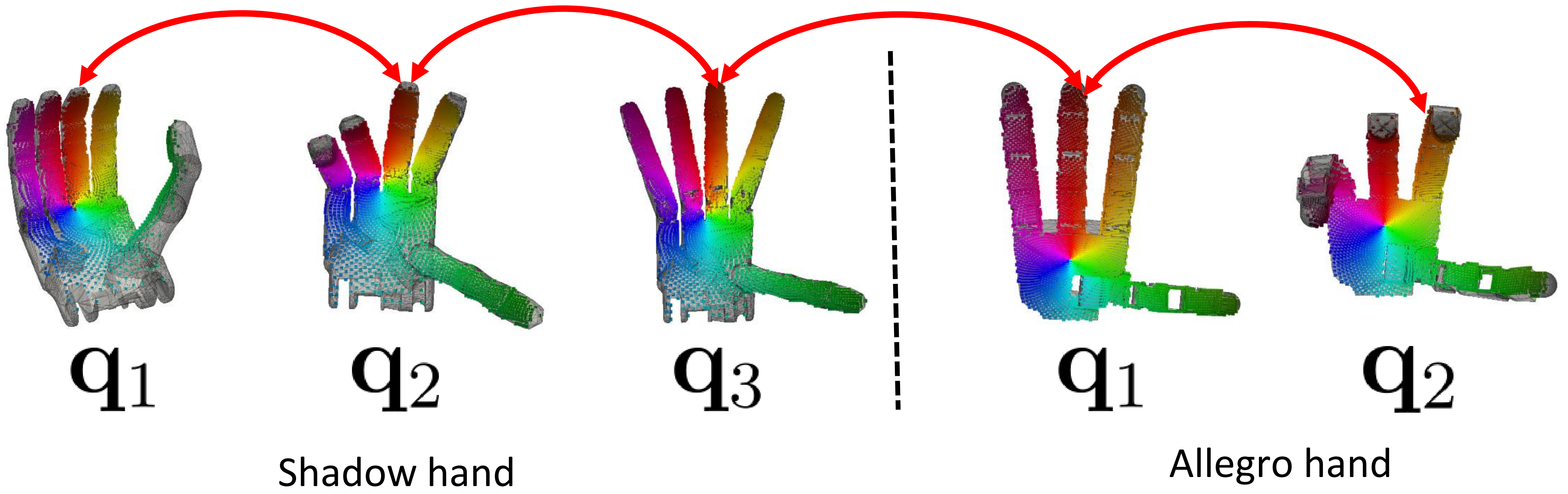
A gripper G is represented by a set of interior points $P_G = \{\mathbf{v}_g \mid \mathbf{v}_g \in \mathbb{R}^3\}$

Each point $\mathbf{v}_g$ is associated with a spherical coordinate (λ, ϕ)

Spherical coordinates $\Phi_G = \{(\lambda_{\mathbf{v}_g}, \phi_{\mathbf{v}_g}) | \mathbf{v}_g \in P_G; \lambda_{\mathbf{v}_g}, \phi_{\mathbf{v}_g} \in [0,1]\}$

Unified Gripper Coordinate Space (UGCS)

- Property 1: the locations of the gripper points change according to grasp configuration $P_G = \{\mathbf{v}_g \mid \mathbf{v}_g \in \mathbb{R}^3\}$ $P_G(\mathbf{q})$

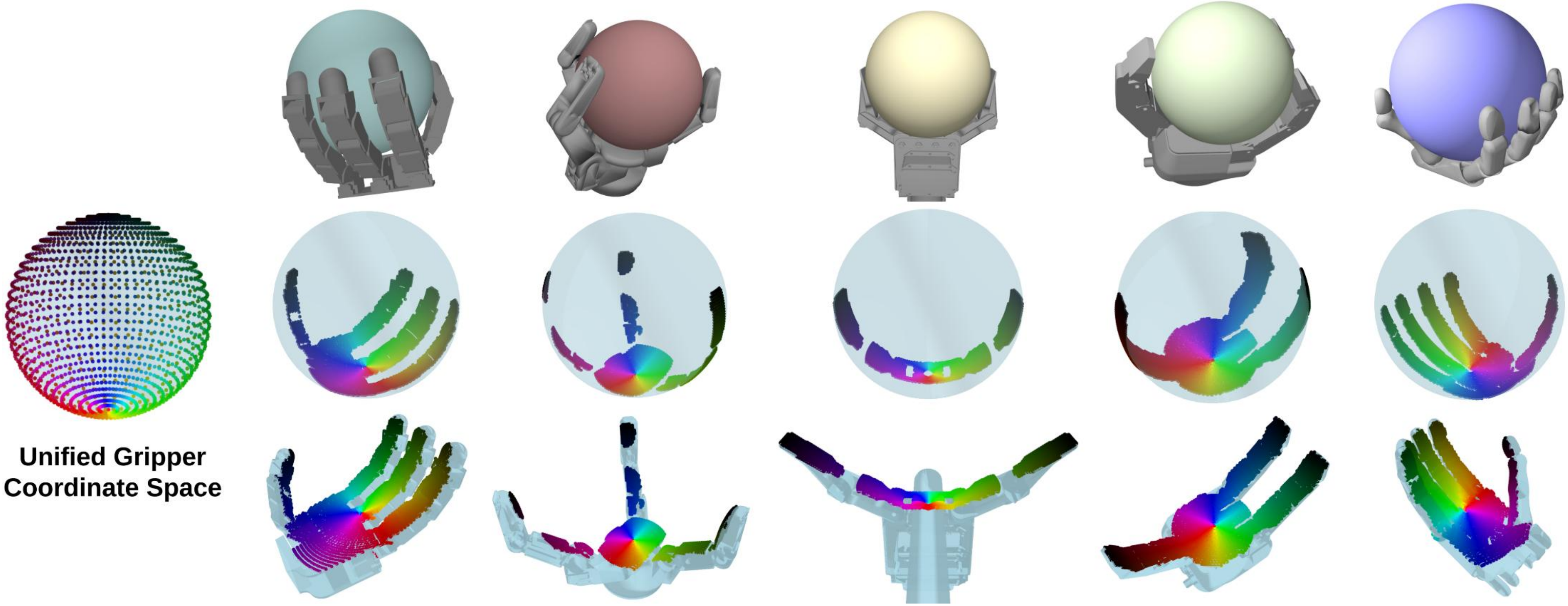


- Property 2: the spherical coordinate for each point **remains the same across configurations and hands** (correspondences!!)

Unified Gripper Coordinate Space

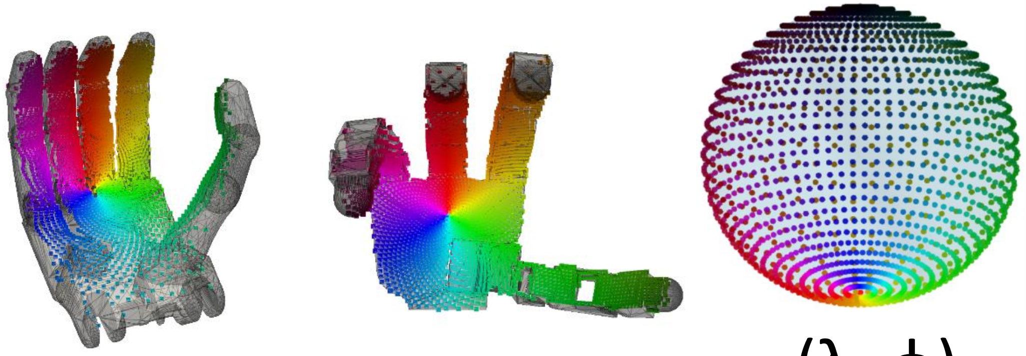
Ninad Khargonkar

Luis Felipe Casas



RobotFingerPrint: Unified Gripper Coordinate Space for Multi-Gripper Grasp Synthesis and Transfer.
Ninad Khargonkar, Luis Felipe Casas, Balakrishnan Prabhakaran, Yu Xiang. In IROS, 2025.

How can we use the UGCS representation for robot manipulation?



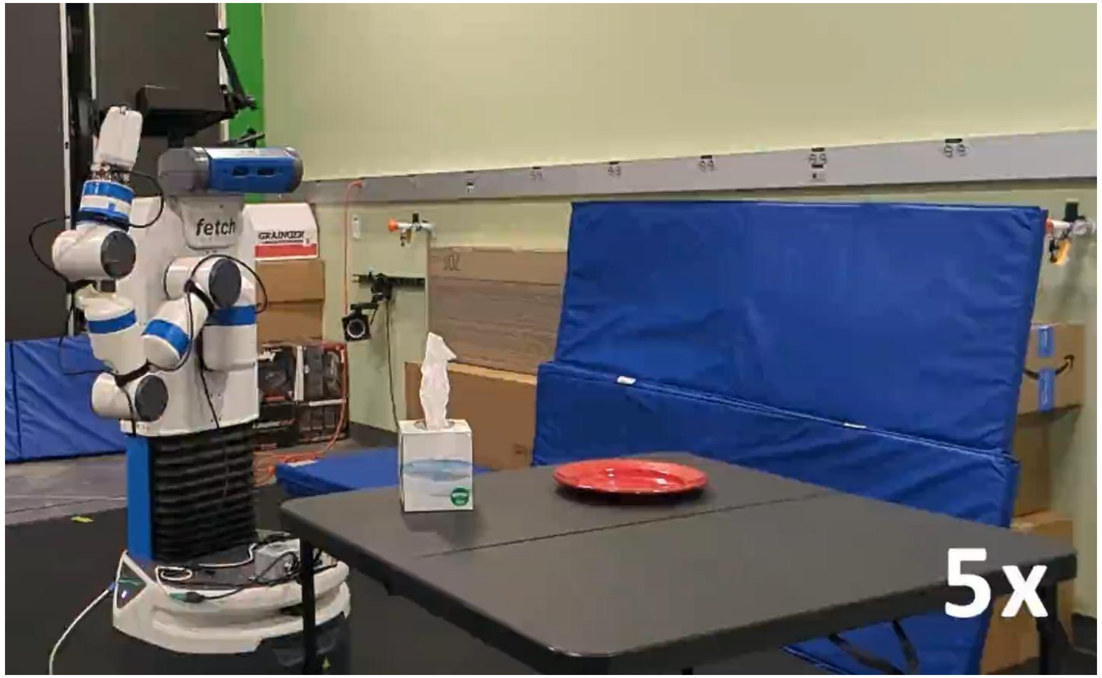
(λ, ϕ)

- Two applications in this talk
- One-shot human-to-robot trajectory transfer
- Cross-embodiment in-hand manipulation

One-Shot Human-to-Robot Trajectory Transfer

One-shot human demonstration

Robot execution in different environment



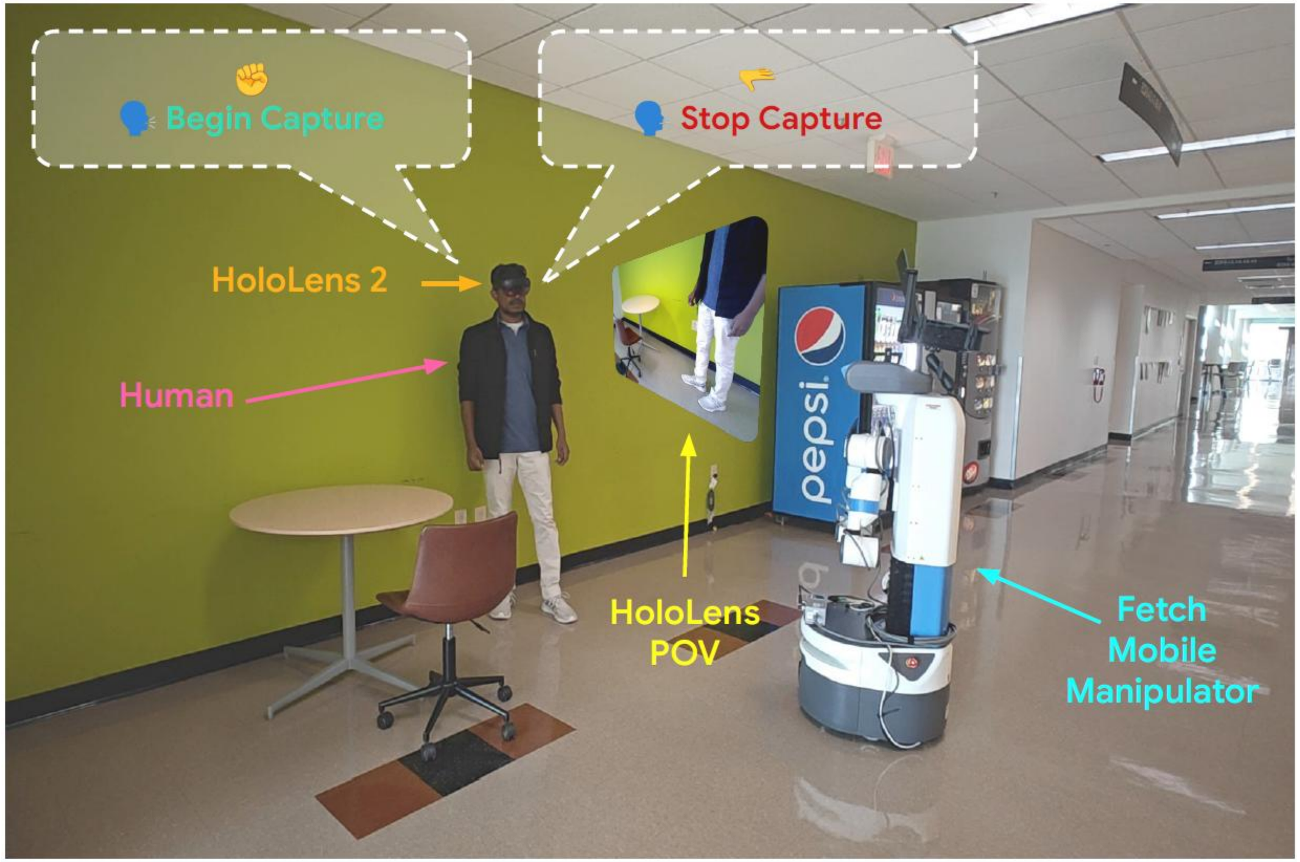
Sai Haneesh Allu

Jishnu Jaykumar P

HRT1: Mobile Manipulation via One-Shot Human-to-Robot Trajectory Transfer. https://irvlutd.github.io/HRT1/
Sai Haneesh Allu*, Jishnu Jaykumar P*, Ninad Khargonkar, Tyler Summers, Jian Yao, Yu Xiang. In arXiv, 2025.

One-Shot Human-to-Robot Trajectory Transfer

Image generated by ChatGPT

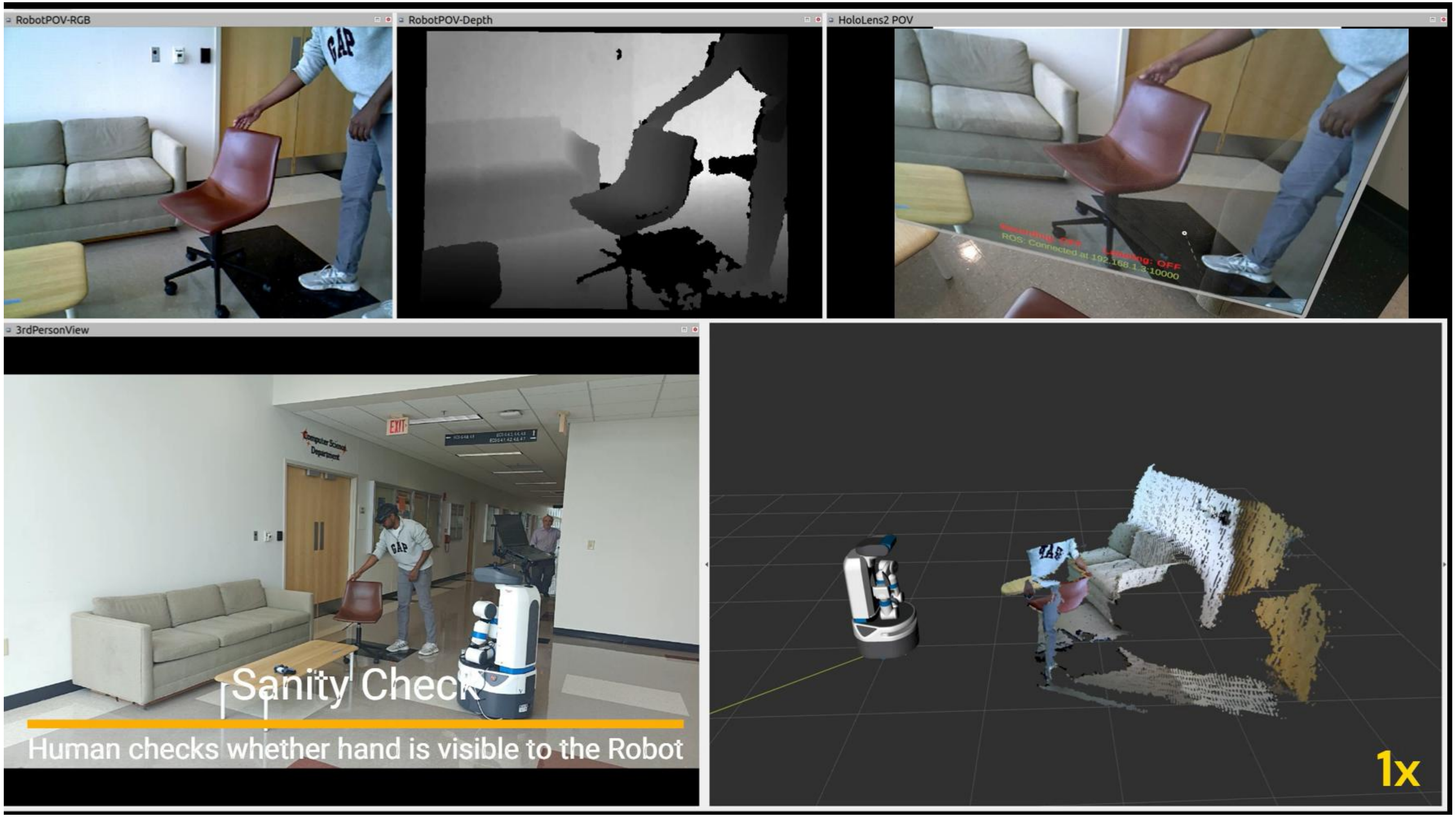


Our setup

iTeach: Interactive Teaching for Robot Perception using Mixed Reality.
Jishnu Jaykumar P, Cole Salvato, Vinaya Bomnale, Jikai Wang, Yu Xiang. In arXiv, 2025.

One-Shot Human-to-Robot Trajectory Transfer

- Human demonstration collection

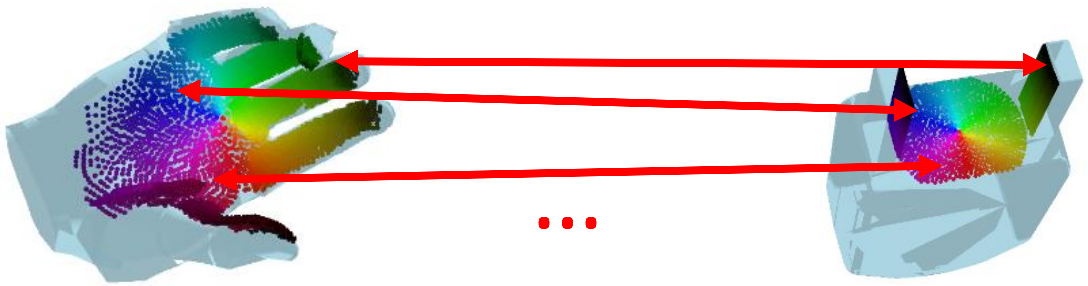


Understanding of the Human Demonstration

Hand Pose Estimation (HaMeR)

Optimization using Depth

Grasp Transfer with UGCS



A Human Grasp

A Fetch gripper

Human hand points

$$P_H(\mathbf{q}_H) = \{\mathbf{v}_H(\mathbf{q}_H) \in \mathbb{R}^3\}$$

Hand configuration (known)

Fetch gripper points

$$P_F(\mathbf{q}_F) = \{\mathbf{v}_F(\mathbf{q}_F) \in \mathbb{R}^3\}$$

Fetch grasp configuration (unknown)

Spherical coordinates (independent of $\mathbf{q}_H$)

$$\Phi_H = \{(\lambda_{\mathbf{v}_g}, \phi_{\mathbf{v}_g}) | \mathbf{v}_g \in P_H; \lambda_{\mathbf{v}_g}, \phi_{\mathbf{v}_g} \in [0,1]\}$$

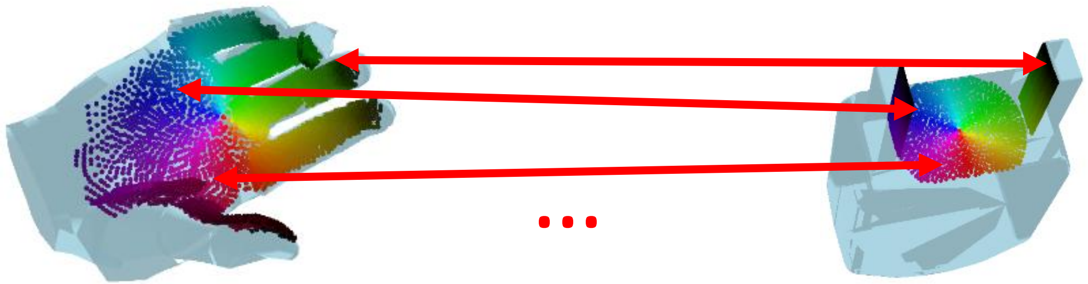
Spherical coordinates (independent of $\mathbf{q}_F$)

$$\Phi_F = \{(\lambda_{\mathbf{v}_g}, \phi_{\mathbf{v}_g}) | \mathbf{v}_g \in P_F; \lambda_{\mathbf{v}_g}, \phi_{\mathbf{v}_g} \in [0,1]\}$$

Matching their UGCS coordinates to establish correspondences (find mutually closest pairs)

$$P_H^c \subset P_H, P_F^c \subset P_F, |P_H^c| = |P_F^c|$$

Grasp Transfer with UGCS



A Human Grasp $\mathbf{q}_H$ A Fetch gripper $\mathbf{q}_F$

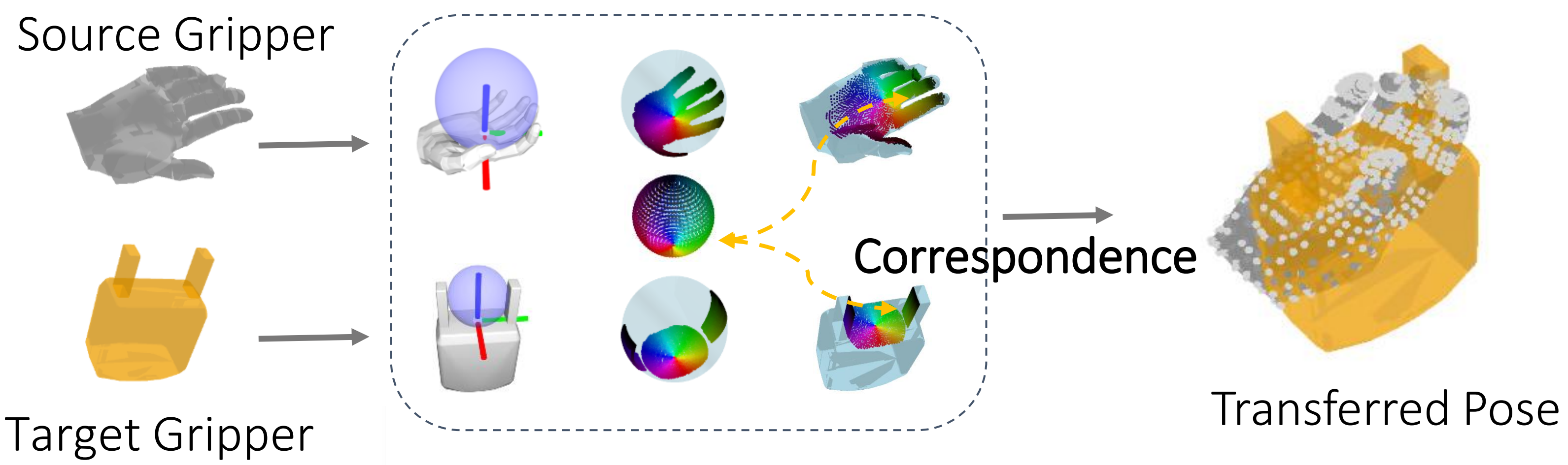
Correspondences from UGCS $P_H^c \subset P_H, P_F^c \subset P_F, |P_H^c| = |P_F^c|$

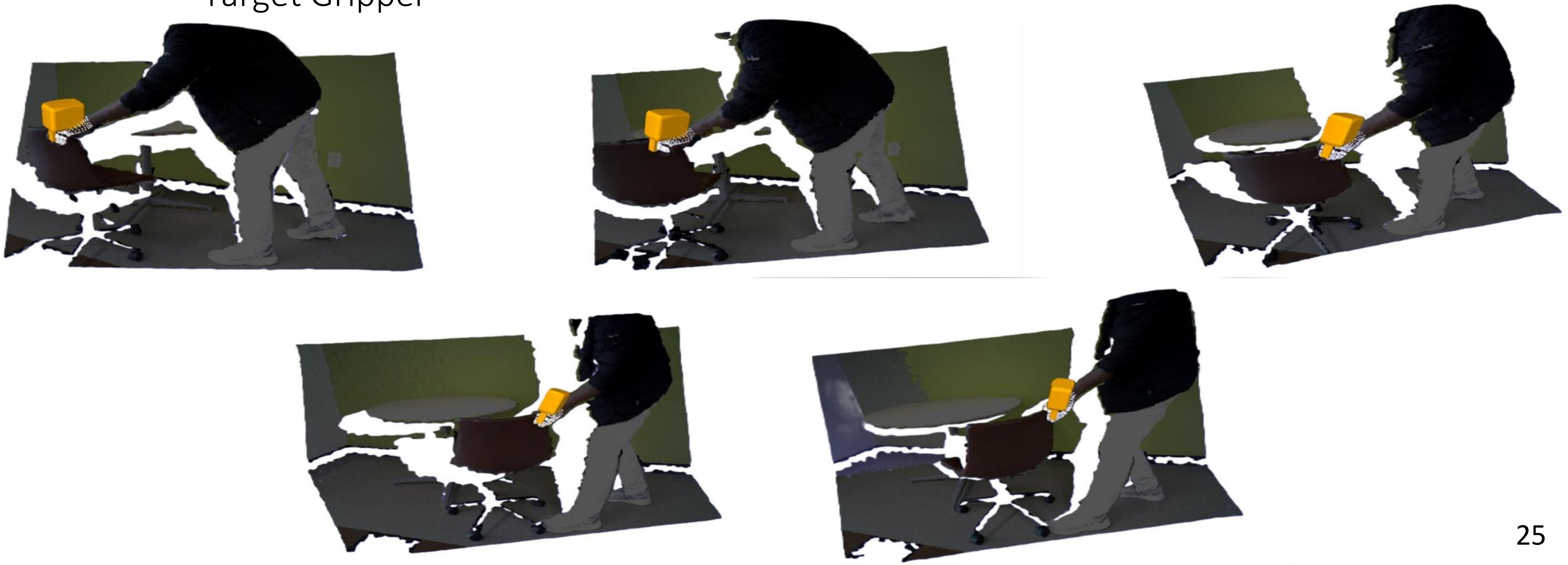
Optimize the target grasp using the

$$\mathbf{q}_F^* = \arg\min_{\mathbf{q}_F} \; E_{\text{dist}}(P_H^c(\mathbf{q}_H), P_F^c(\mathbf{q}_F)) + E_{\text{n}}(\mathbf{q}_F)$$

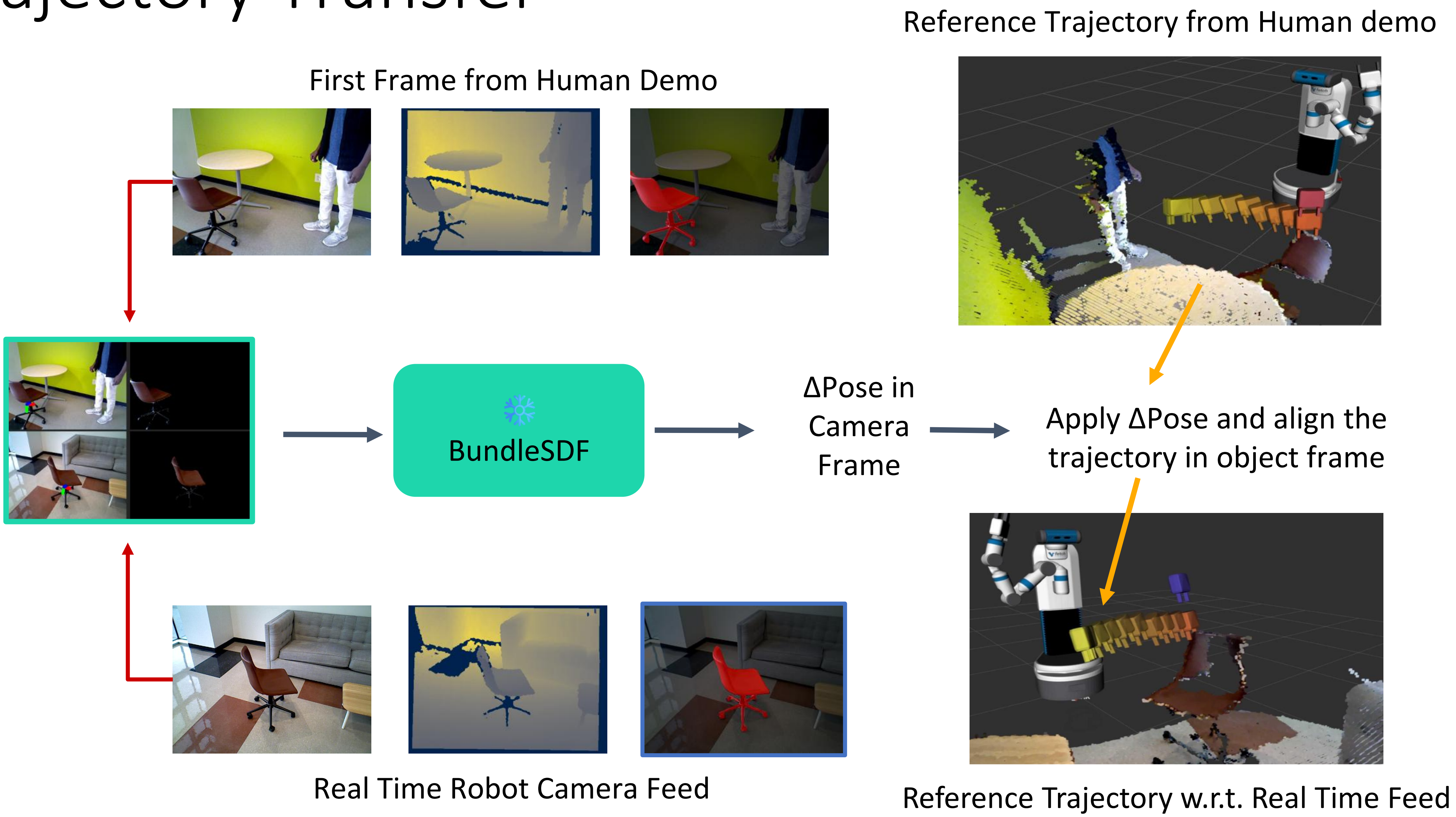
Reference grasp Joint limits

Understanding of the Human Demonstrations





Trajectory Transfer



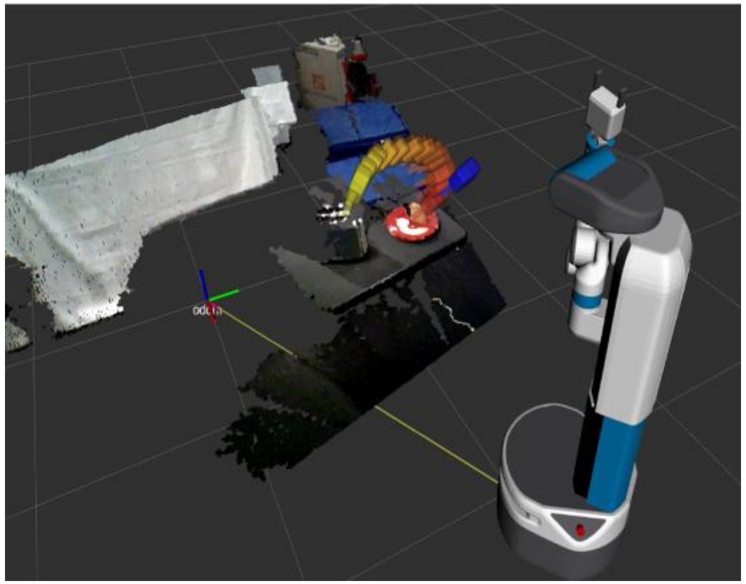
Trajectory Transfer

- Dual-object tasks

Demo

ΔPose for object 1

ΔPose for object 2



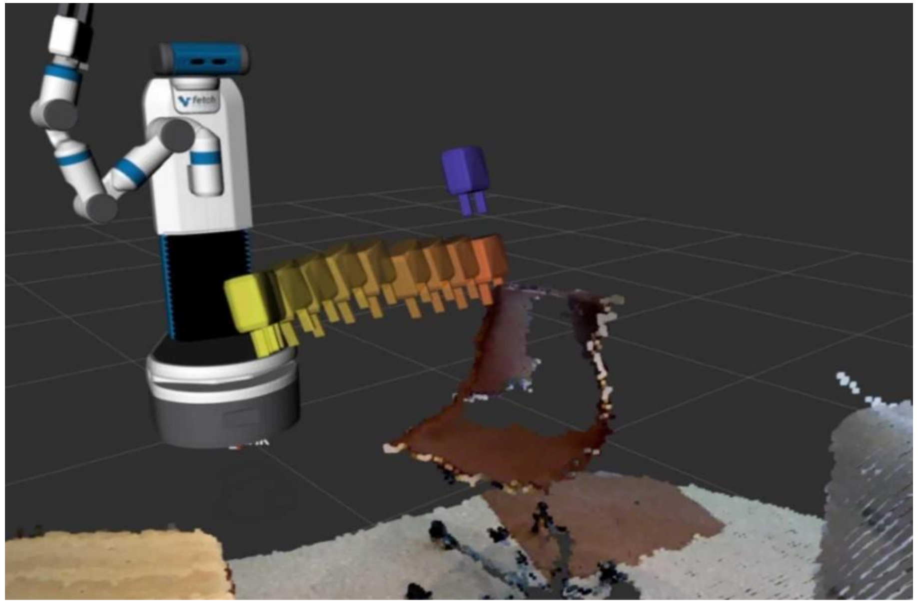
Transferred trajecotry

Trajectory Transfer

- How to follow the transferred gripper trajectory?

Task Space

Robot View

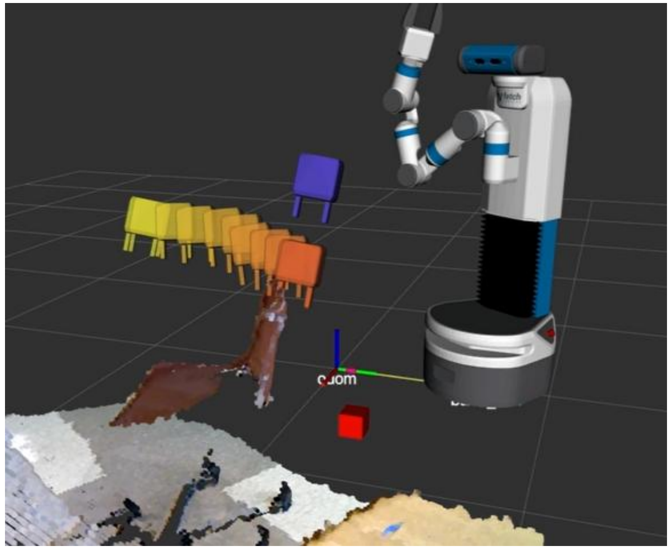


Reference Trajectory w.r.t. Real Time Feed

Optimizing the Robot Base Location

- Find the base position that can reach N gripper poses from the trajectory

New base relative to current base $\mathbf{x} = \begin{bmatrix} x \\ y \\ \theta \end{bmatrix}$ $\mathbf{T}(\mathbf{x}) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & x \\ \sin\theta & \cos\theta & 0 & y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ Unknown



Gripper pose in current base $\mathcal{T} = \{\mathbf{T}_1, \mathbf{T}_2 \ldots, \mathbf{T}_N\}$ Known

Arm configuration $\mathcal{Q} = \{\mathbf{q}_1, \mathbf{q}_2 \ldots, \mathbf{q}_N\}$ Unknown

$$\arg\min_{\mathbf{x},\mathcal{Q}} \left(\lambda_{\text{effort}} \|\mathbf{x}\|^2 + \lambda_{\text{goal}} \sum_{i=1}^{N} c_{\text{goal}}(\mathbf{T}(\mathbf{q}_i), \mathbf{T}(\mathbf{x})^{-1} \cdot \mathbf{T}_i) \right)$$

$$\text{s.t.,} \quad -\mathbf{x}_{\min} \leq \mathbf{x} \leq \mathbf{x}_{\max}$$

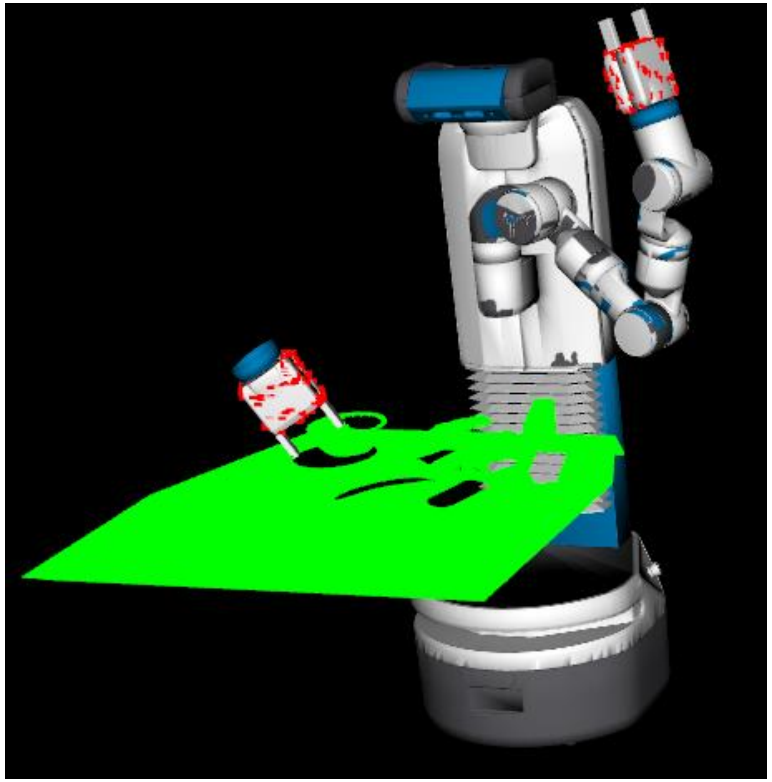
$$\mathbf{q}_l \leq \mathbf{q}_i \leq \mathbf{q}_u, i = 1, \ldots, N,$$

Forward kinematics

Gripper goal in new base

Goal-reaching Cost Function

- Point Cloud-based Cost Function for Goal Reaching



Gripper pose

$$\mathbf{T} = (\mathbf{R}, \mathbf{t})$$

Goal pose

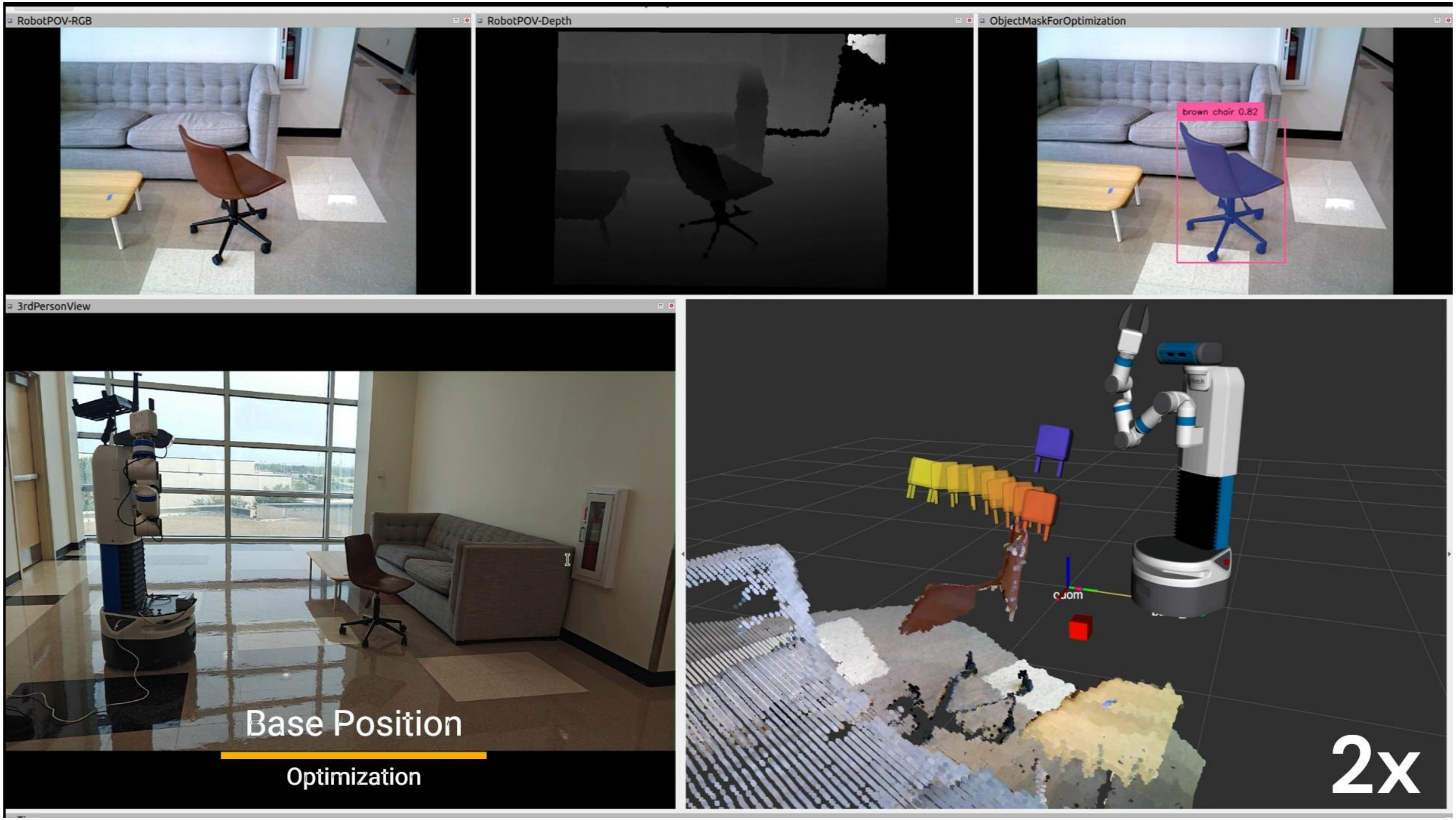
$$\mathbf{T}_g = (\mathbf{R}_g, \mathbf{t}_g)$$

$$c_{\text{goal}}(\mathbf{T}, \mathbf{T}_g) = \sum_{i=1}^{m} \|(\mathbf{R}\mathbf{x}_i + \mathbf{t}) - (\mathbf{R}_g\mathbf{x}_i + \mathbf{t}_g)\|^2$$

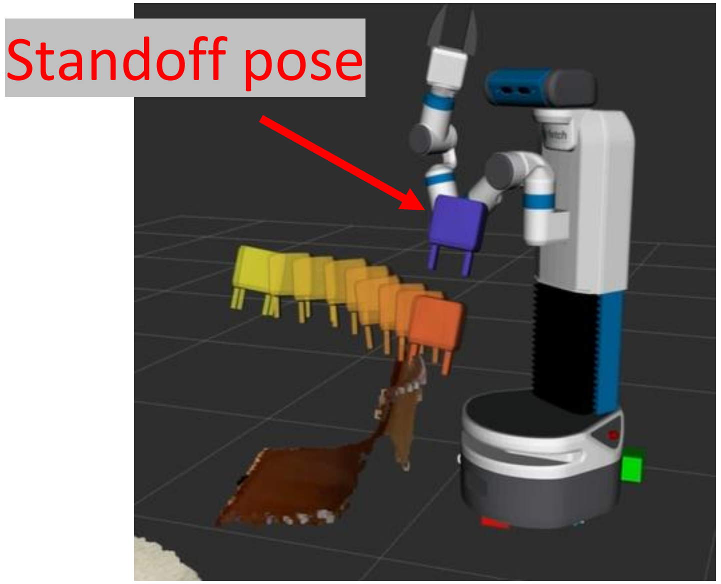
Points on the gripper CAD model

Grasping Trajectory Optimization with Point Clouds. Yu Xiang, Sai Haneesh Allu, Rohith Peddi, Tyler Summers, Vibhav Gogate. In IROS, 2024.

Optimizing the Robot Base Location



Optimizing the Robot Trajectory



- Find the trajectory to follow the gripper poses well

Gripper trajectory in new robot base

Known $\mathcal{T} = (\mathbf{T}_0, \mathbf{T}_1, \ldots, \mathbf{T}_T)$

Standoff pose

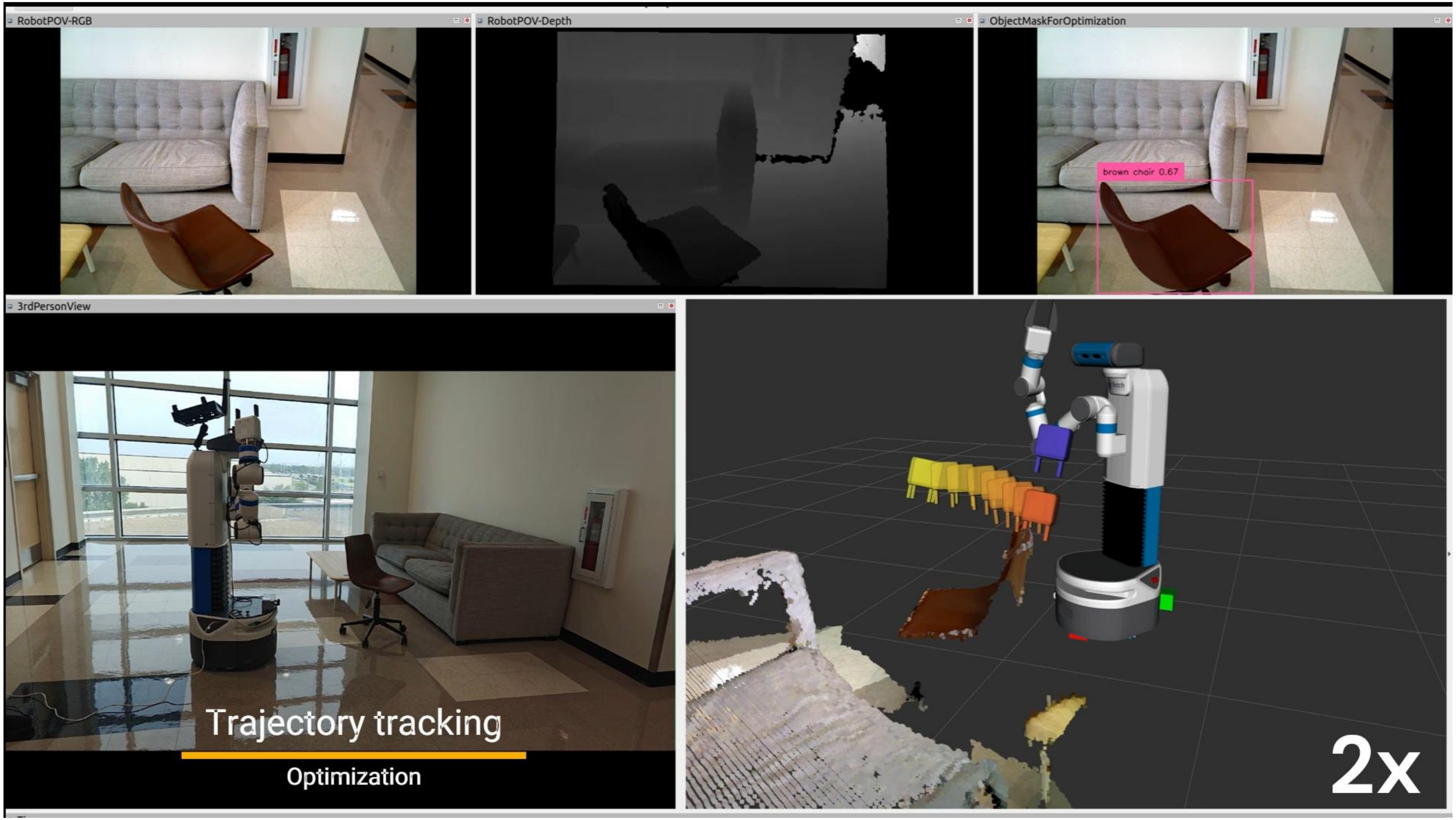
Unknown $\mathcal{Q} = (\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_T) \qquad \dot{\mathcal{Q}} = (\dot{\mathbf{q}}_0, \dot{\mathbf{q}}_1, \ldots, \dot{\mathbf{q}}_T)$

$$\arg\min_{\mathcal{Q}, \dot{\mathcal{Q}}} \sum_{i=0}^{T} \left(\lambda c_{\text{goal}}(\mathbf{T}(\mathbf{q}_i), \mathbf{T}_i) + \lambda_1 c_{\text{collision}}(\mathbf{q}_i) + \lambda_2 ||\dot{\mathbf{q}}_i||^2 \right)$$

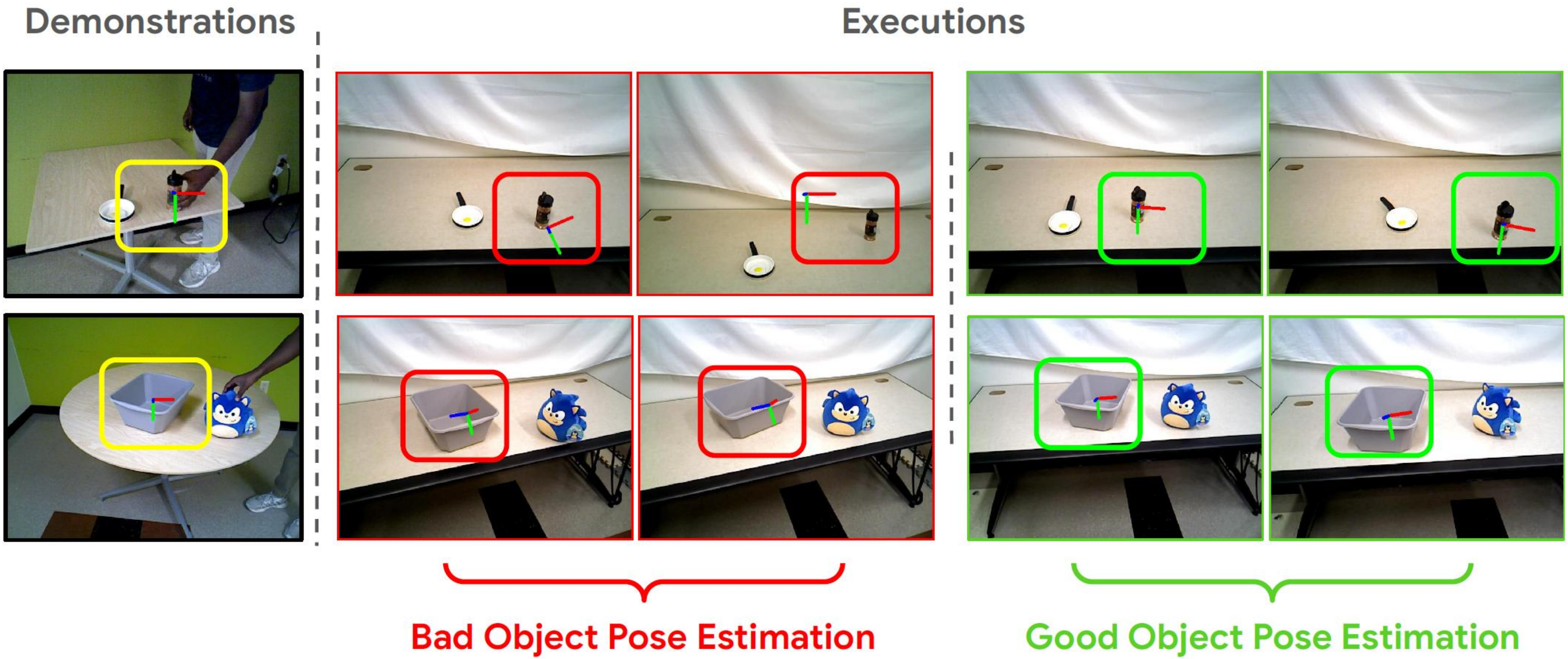
$$\begin{aligned} \text{s.t.,} \quad & \dot{\mathbf{q}}_0 = \mathbf{0}, \dot{\mathbf{q}}_T = \mathbf{0} \\ & \mathbf{q}_{i+1} = \mathbf{q}_i + \dot{\mathbf{q}}_i dt, i = 0, \ldots, T-1 \\ & \mathbf{q}_l \leq \mathbf{q}_i \leq \mathbf{q}_u, i = 0, \ldots, T \\ & \dot{\mathbf{q}}_l \leq \dot{\mathbf{q}}_i \leq \dot{\mathbf{q}}_u, i = 0, \ldots, T \end{aligned}$$

We utilize the Interior Point OPTimizer (Ipopt) with the CasADi framework to solve it.

Optimizing the Robot Trajectory

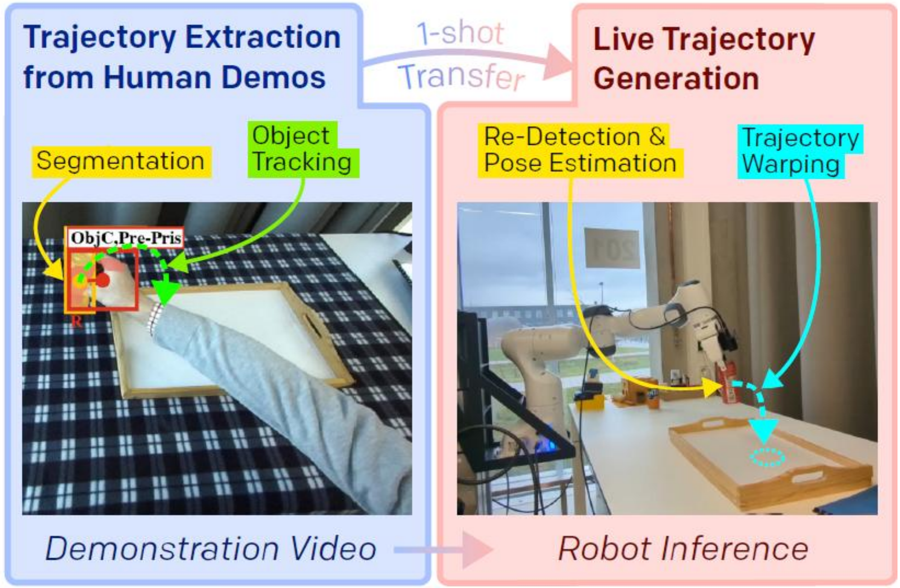


Object Pose Verification



Quantitative Evaluation

- 16 tasks
- Baseline: DITTO (transfer object trajectory)



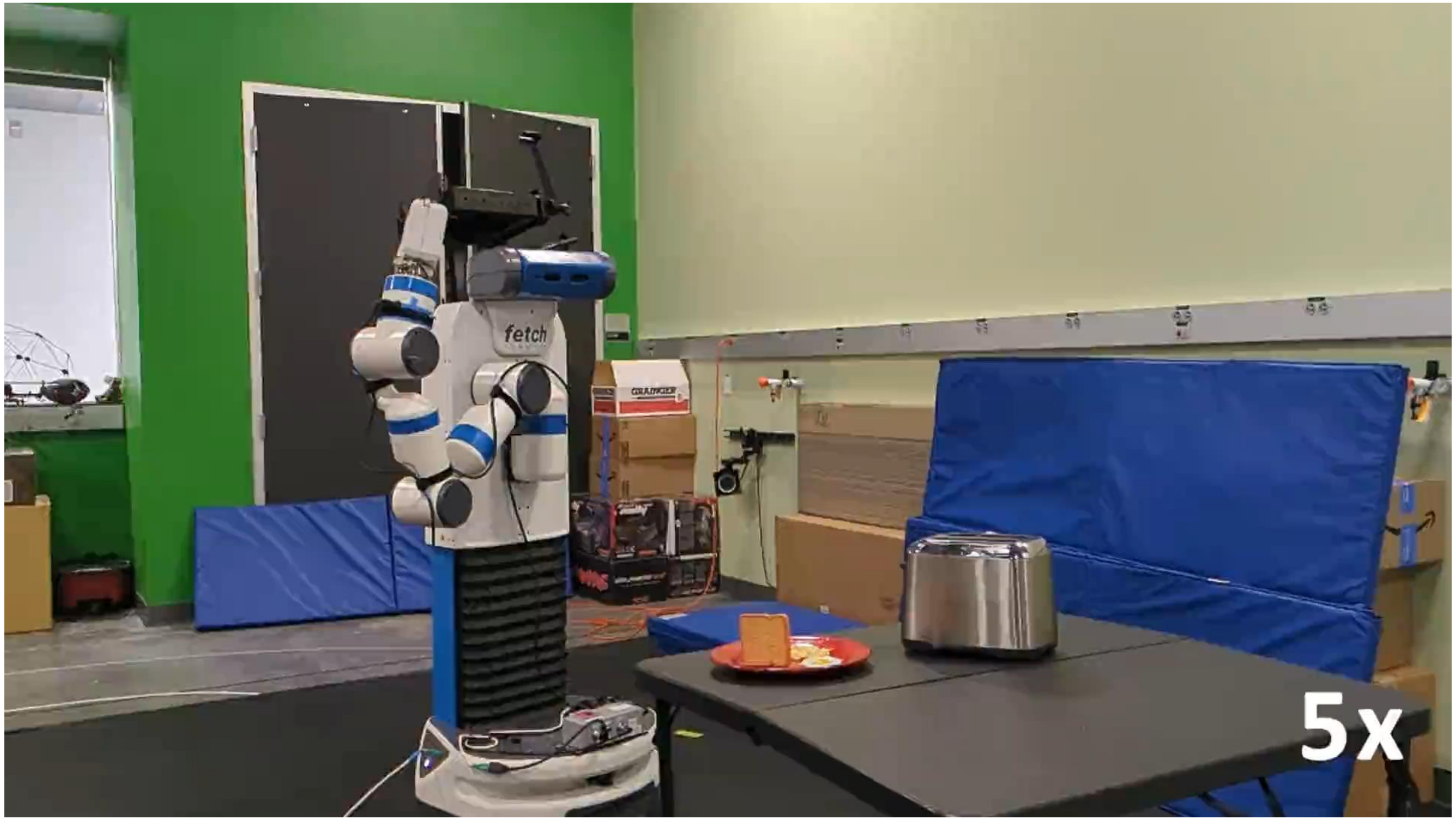
Trajectory Transfer, Heppert et al. University of Freiburg, IROS 2024

Skill	Grasp success		Task completion	
	DITTO [15]	Ours	DITTO [15]	Ours
Single object				
Move the chair	3	3	0	3
Close fire extinguisher door	0	3	0	3
Dual object				
Put toy in the bin	2	3	1	3
Put bread in the toaster	1	3	1	3
Put seasoning on the omelette	3	3	2	3
Put Lays on the red plate	2	2	1	2
Clean plate with brush	1	3	0	3
Clean plate with tissue	0	3	0	3
Clean plate with kitchen towel	2	3	1	3
Remove cap from wall hook	3	3	1	3
Hang cap onto wall hook	0	3	0	2
Take out sugar box from shelf	1	3	0	3
Rearrange sugar box in the shelf	2	3	0	2
Place bottle in the shelf	3	3	0	3
Close jar with a lid	2	3	0	2
Displace cracker box	3	3	3	3
Total	**28/48**	**47/48**	**10/48**	**44/48**

Object pose tracking is not reliable due to occlusions by human hand

One-Shot Human-to-Robot Trajectory Transfer

Put bread in the toaster

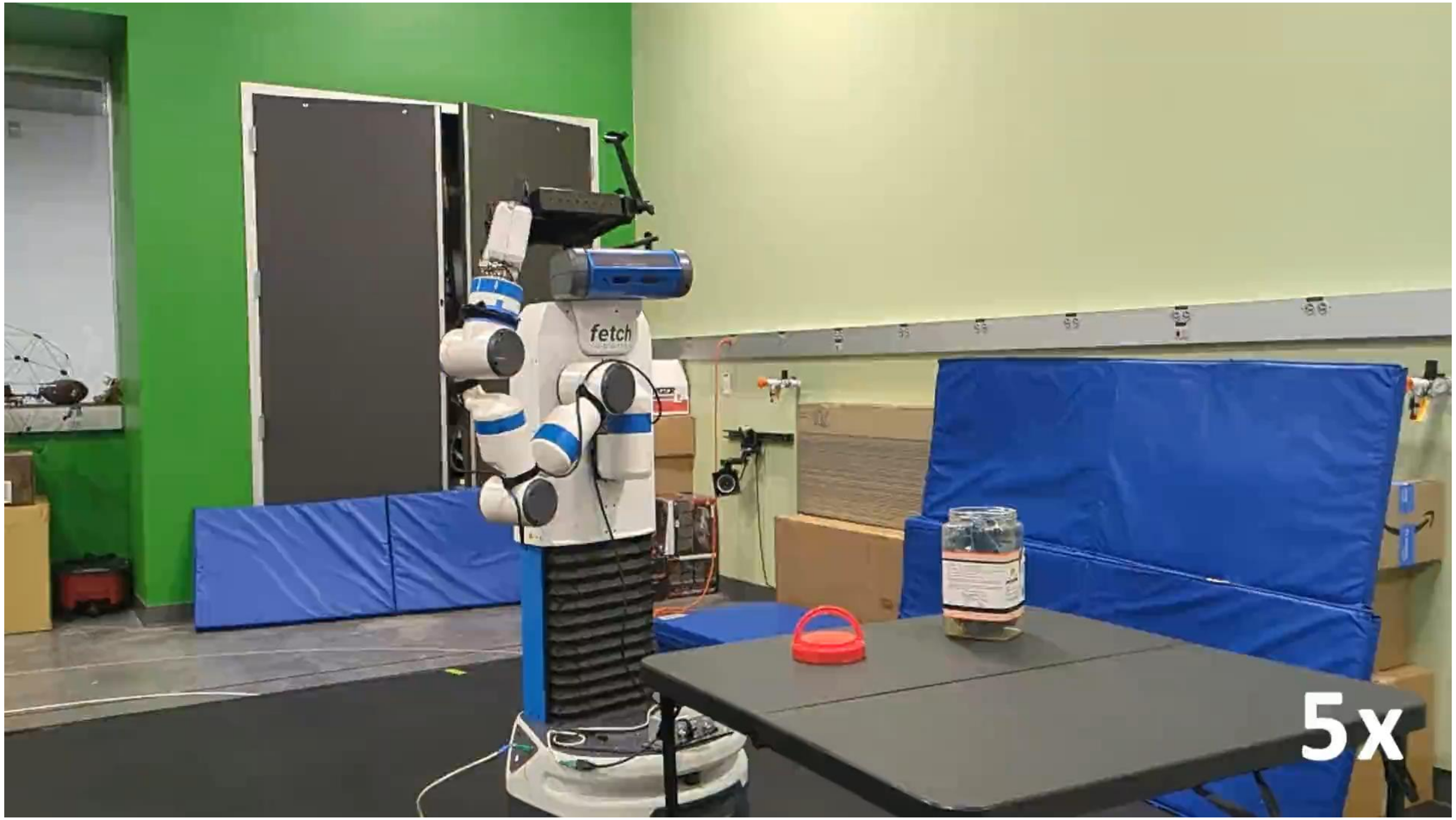


One-Shot Human-to-Robot Trajectory Transfer

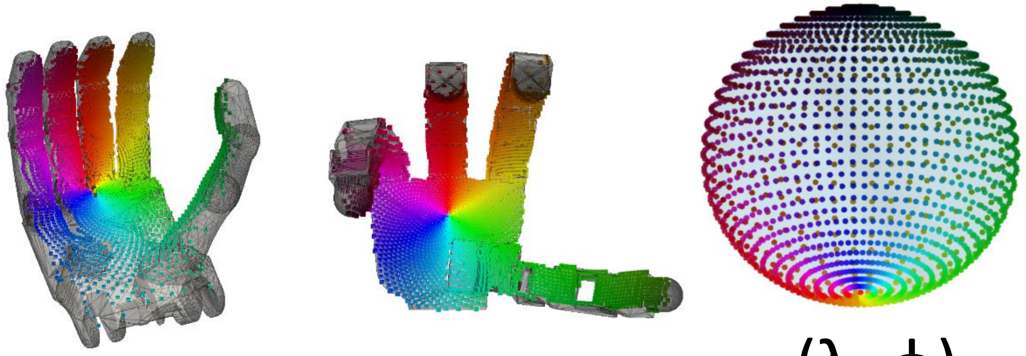
Hang cap onto wall hook

Failure Example

Close jar with a lid



How can we use the UGCS representation for robot manipulation?

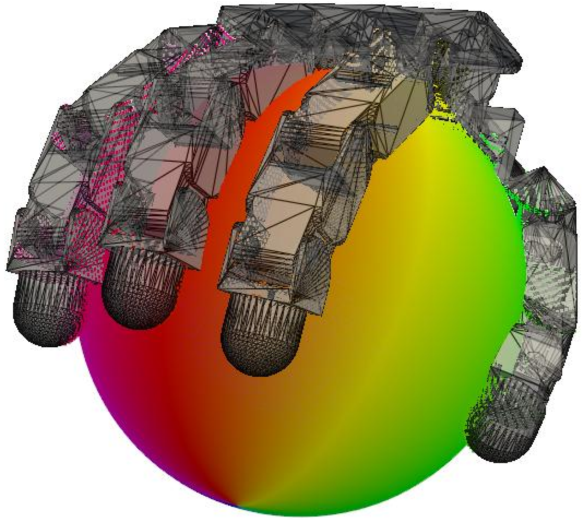


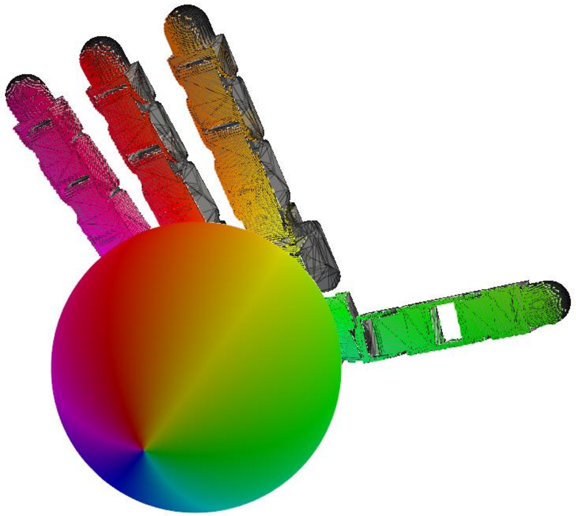
(λ, ϕ)

- Two applications in this talk
- One-shot human-to-robot trajectory transfer
- Cross-embodiment in-hand manipulation (ongoing work)

Unified Gripper Action Space (UGAS)

Luis Felipe Casas



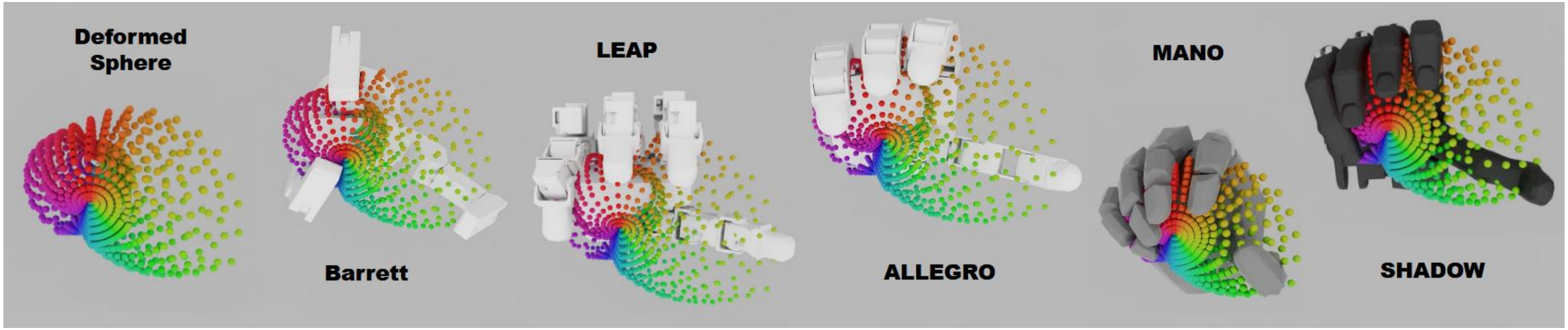


Unified Gripper Coordinate Space

Can we use this sphere to control any robotic gripper/hand?

Unified Gripper Action Space (UGAS)

- **Our Idea:** the deformation of the sphere will drive the movement of the hand (the hand should touch the deformed sphere correctly)
- Action space: deformation of the sphere (shared by any hand!)

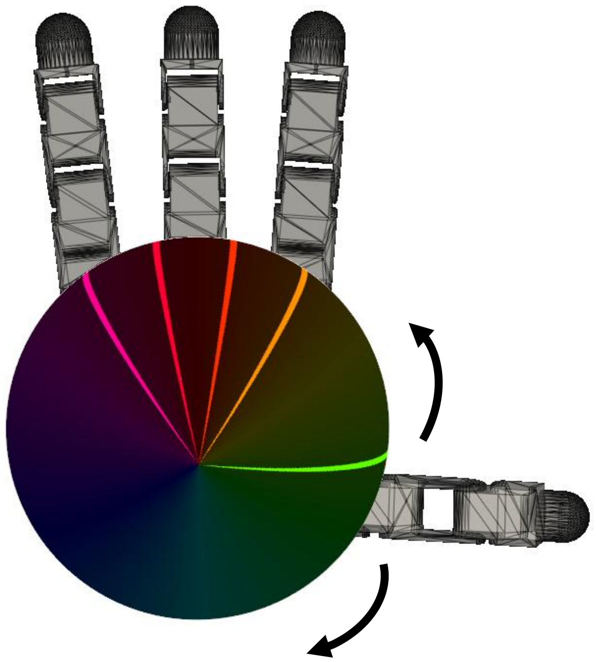


Different Grippers with the same deformed sphere

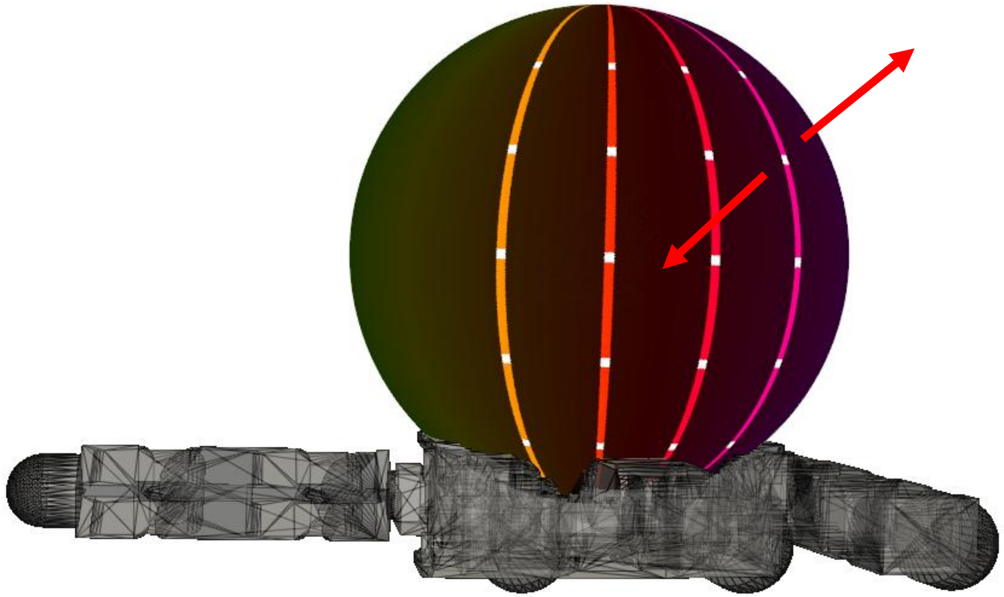
UGAS: deforming the sphere

- Deforming every point on the sphere is too expensive for control

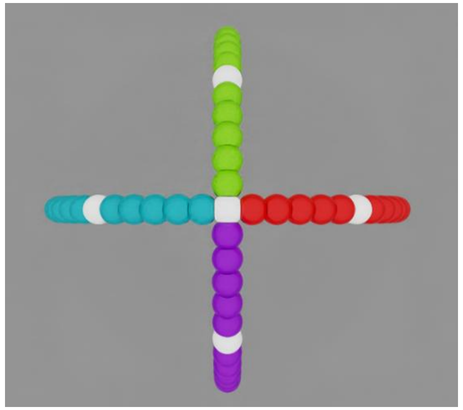
Define several “driving planes”



Define several “driving vectors” on each driving plane

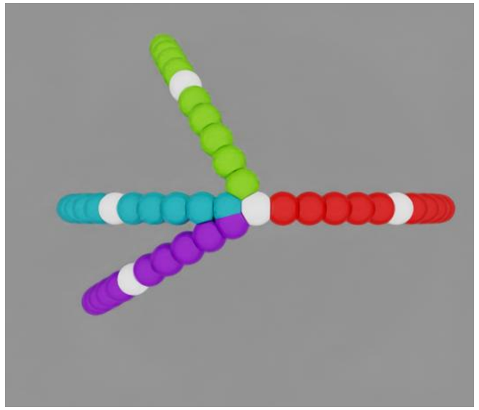


UGAS: deforming the sphere



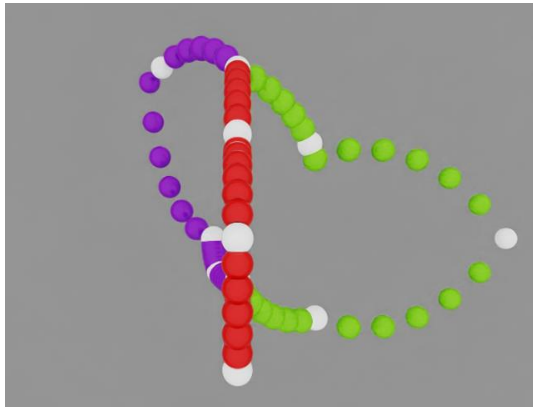
Top View
4 driving planes

Move the driving planes



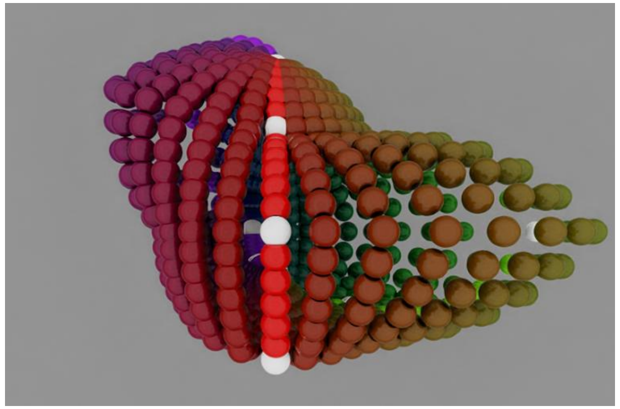
Top View
4 driving planes

Move the driving vectors



Side View
4 driving planes

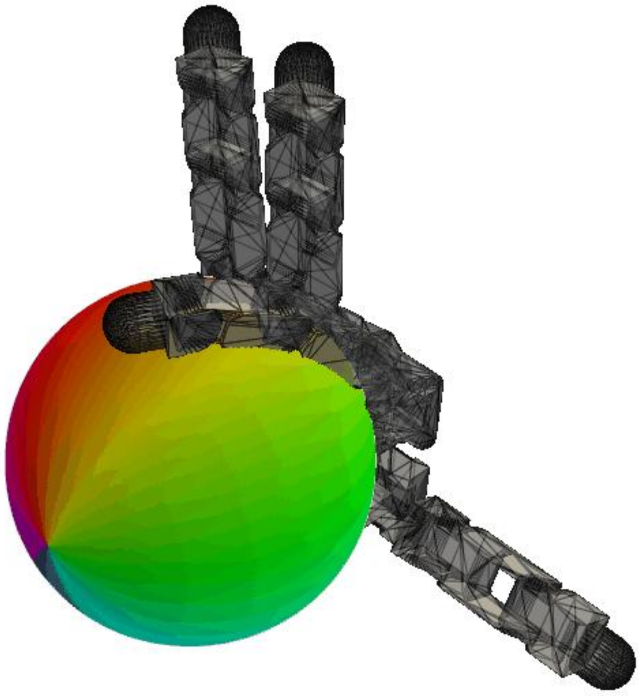
Interpolation to get all the points on the sphere



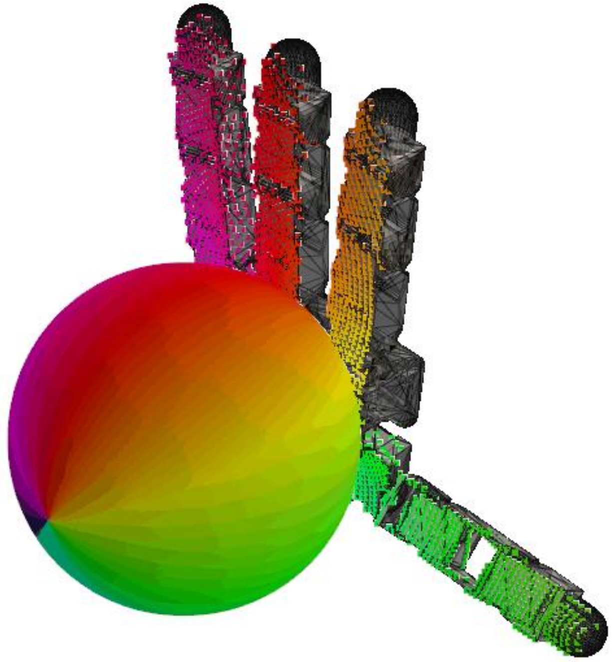
Side View

UGAS: Cascaded Inverse Kinematics (CIK)

- Given a deformed sphere, we solve IK to obtain the hand configuration

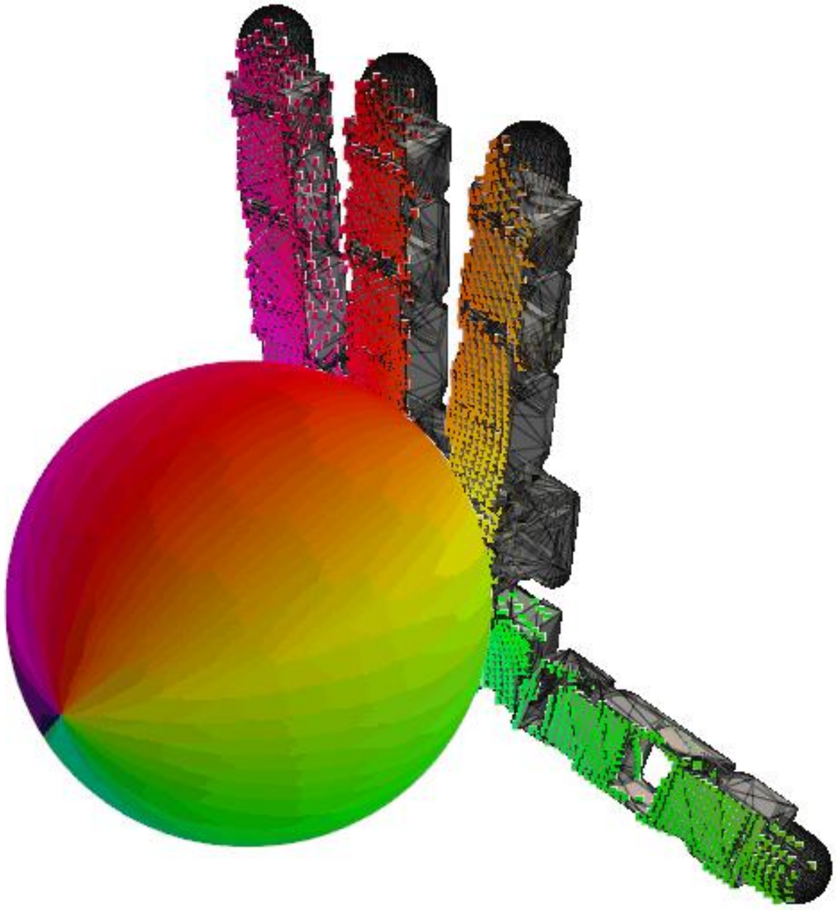


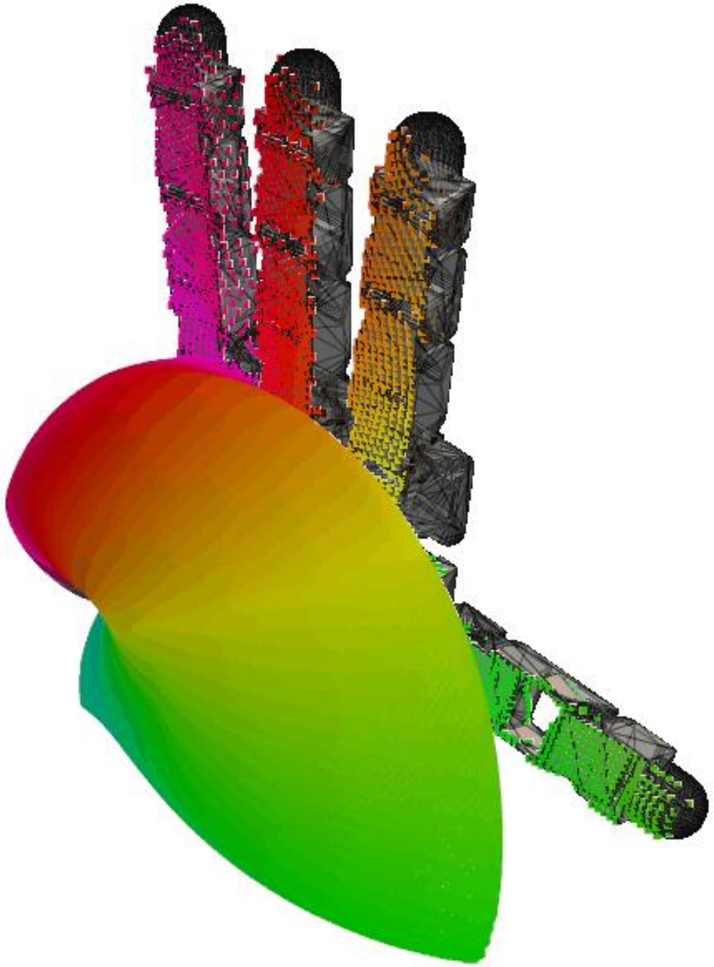
1. Solve lateral joints



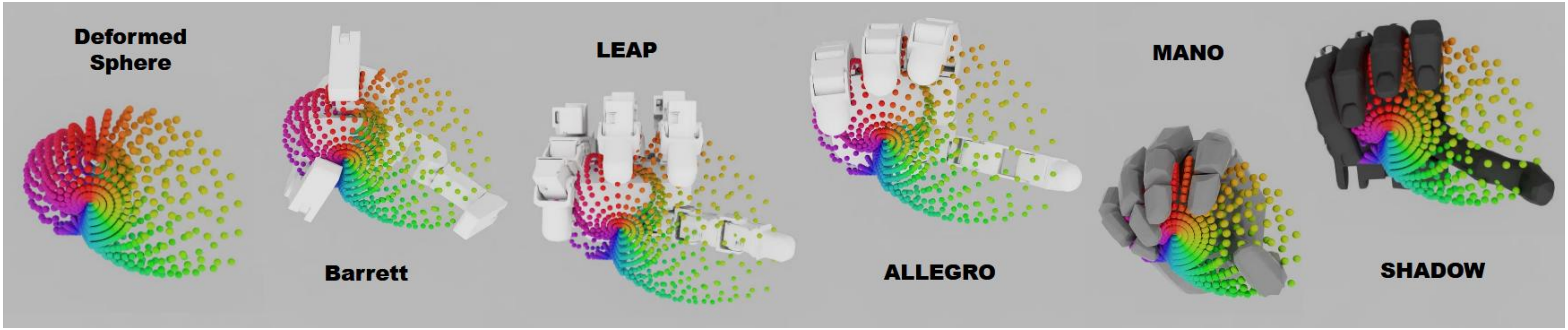
2. Solve encompassing joints
We solve for each joint one at a time,
in the order of the kinematic tree.

UGAS: Cascaded Inverse Kinematics (CIK)





Unified Gripper Action Space (UGAS)

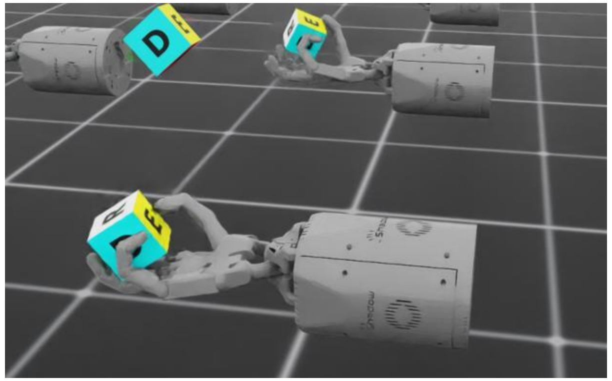


Control actions

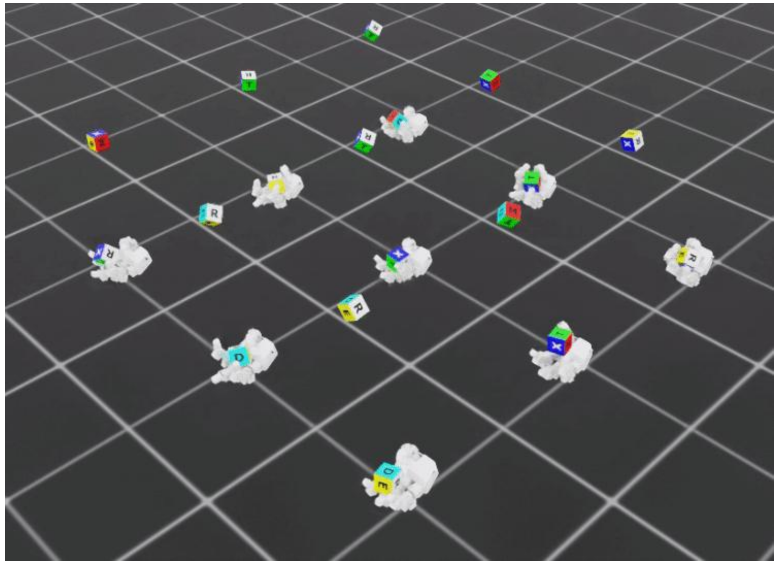
$\boxed{\begin{array}{ll} \text{Driving planes} & \Delta\theta \\ \text{Driving vectors} & \Delta r \end{array}}$ $\xrightarrow{\text{Inverse Kinematics}}$ Hand configuration $\mathbf{q}$

UGAS for In-hand Manipulation

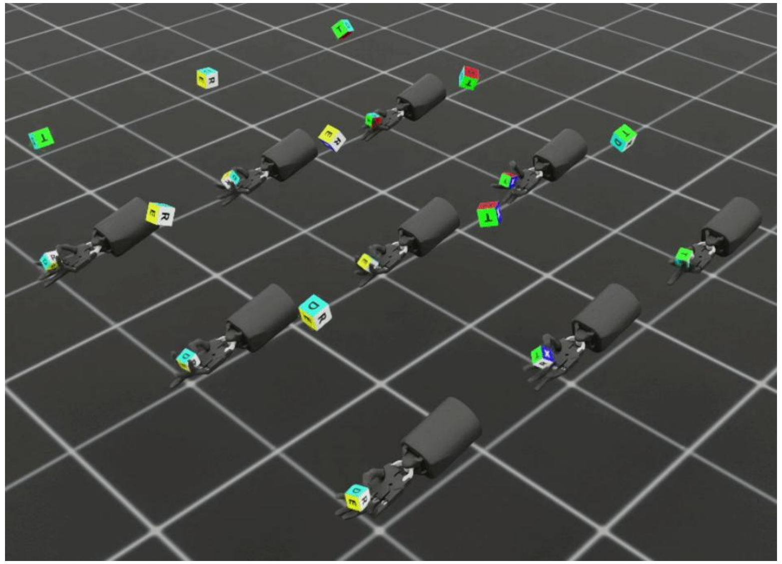
- Task: repose a cube to a target orientation
  - 10 consecutive reposing within 30 seconds
  - RL training in Isaac Lab with PPO and our sphere controller



Allegro (4 fingers)
9.75

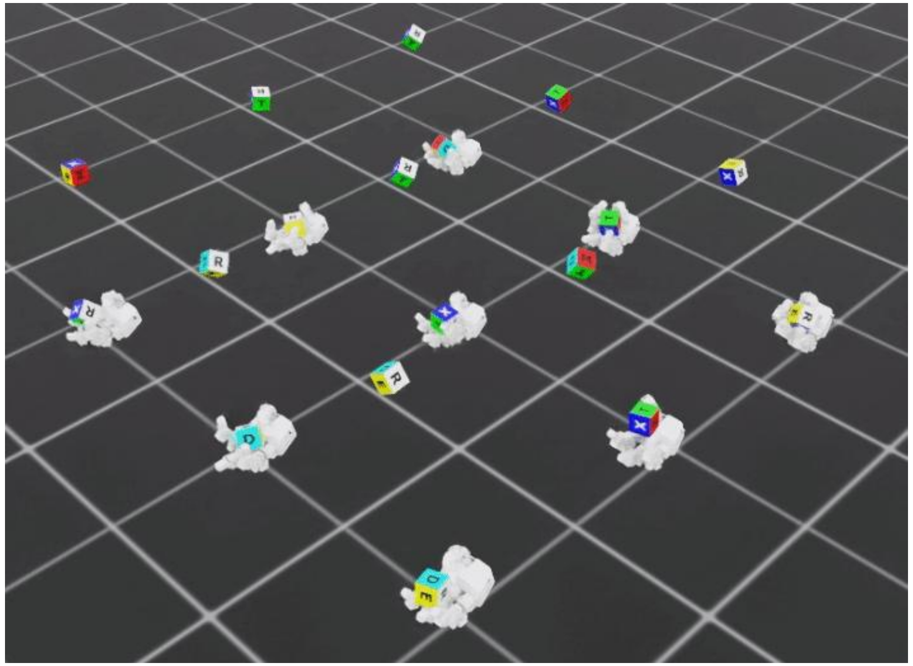


Leap (4 fingers)
9.92



Shadow Hand (5 fingers)
9.938

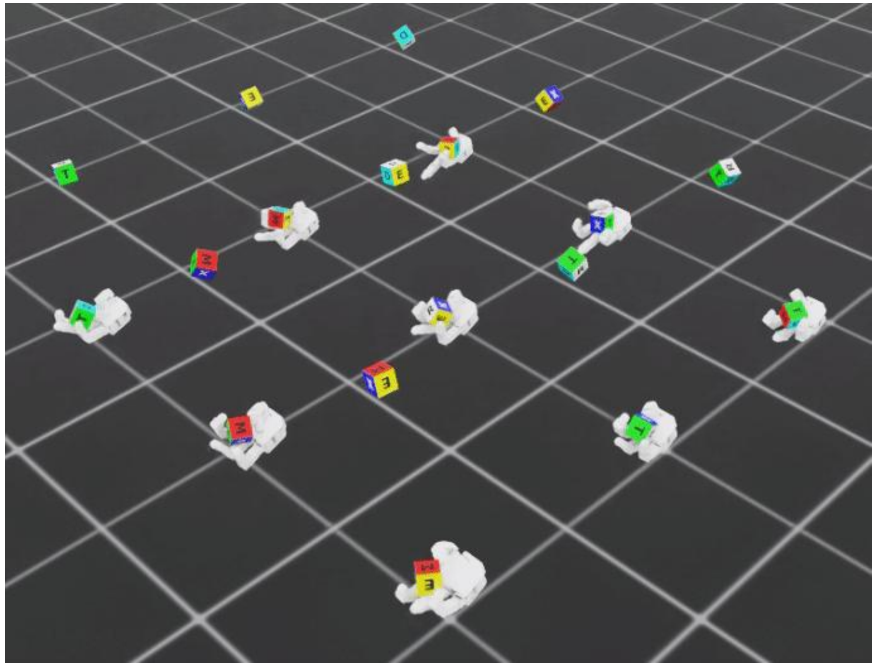
Zero-Shot Policy Transfer (4-finger to 4-finger)



Leap (4-finger)

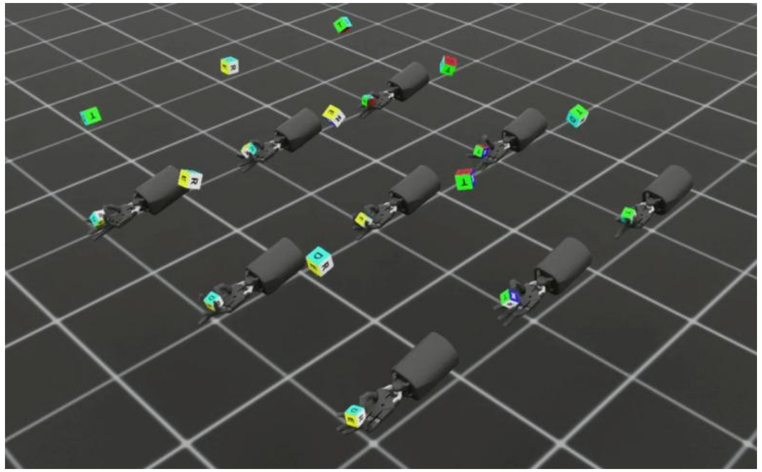
Leap 9.92 to Allegro 9.61 →

← Allegro 9.75 to Leap 6.5

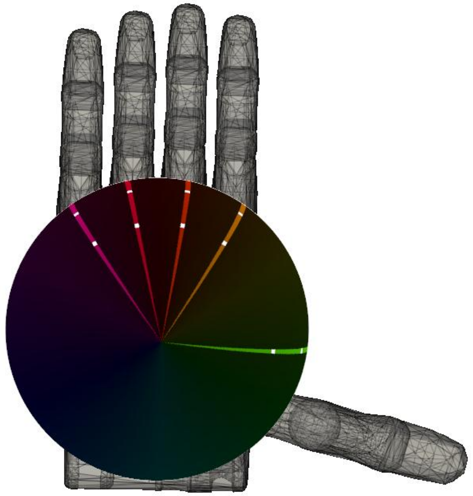


Allegro (4-finger)

Zero-Shot Policy Transfer (5-finger to 4-finger)

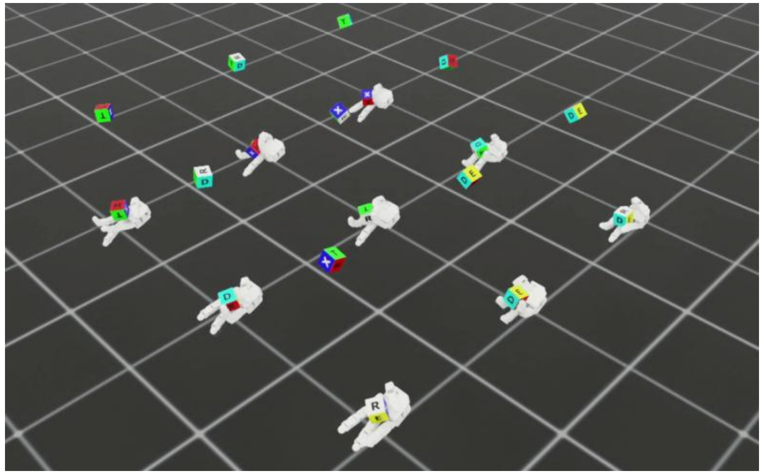


Shadow Hand (5-finger)

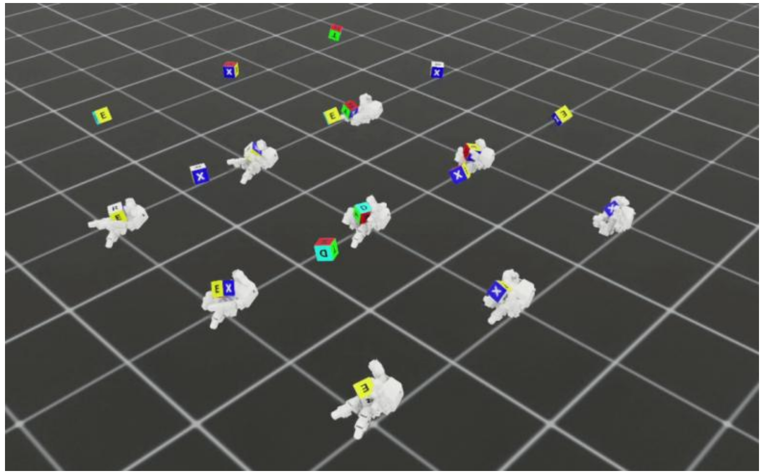


Shadow 9.94 to
Allegro 2.88

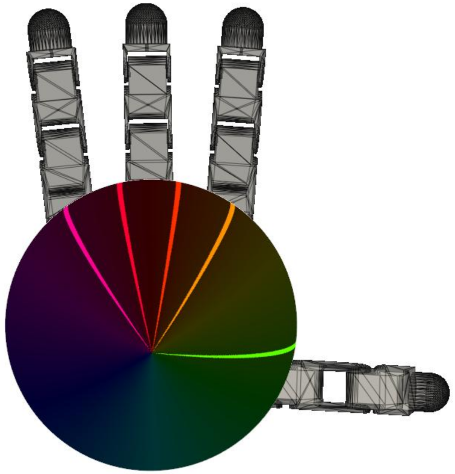
Shadow 9.94 to
Allegro 2.70



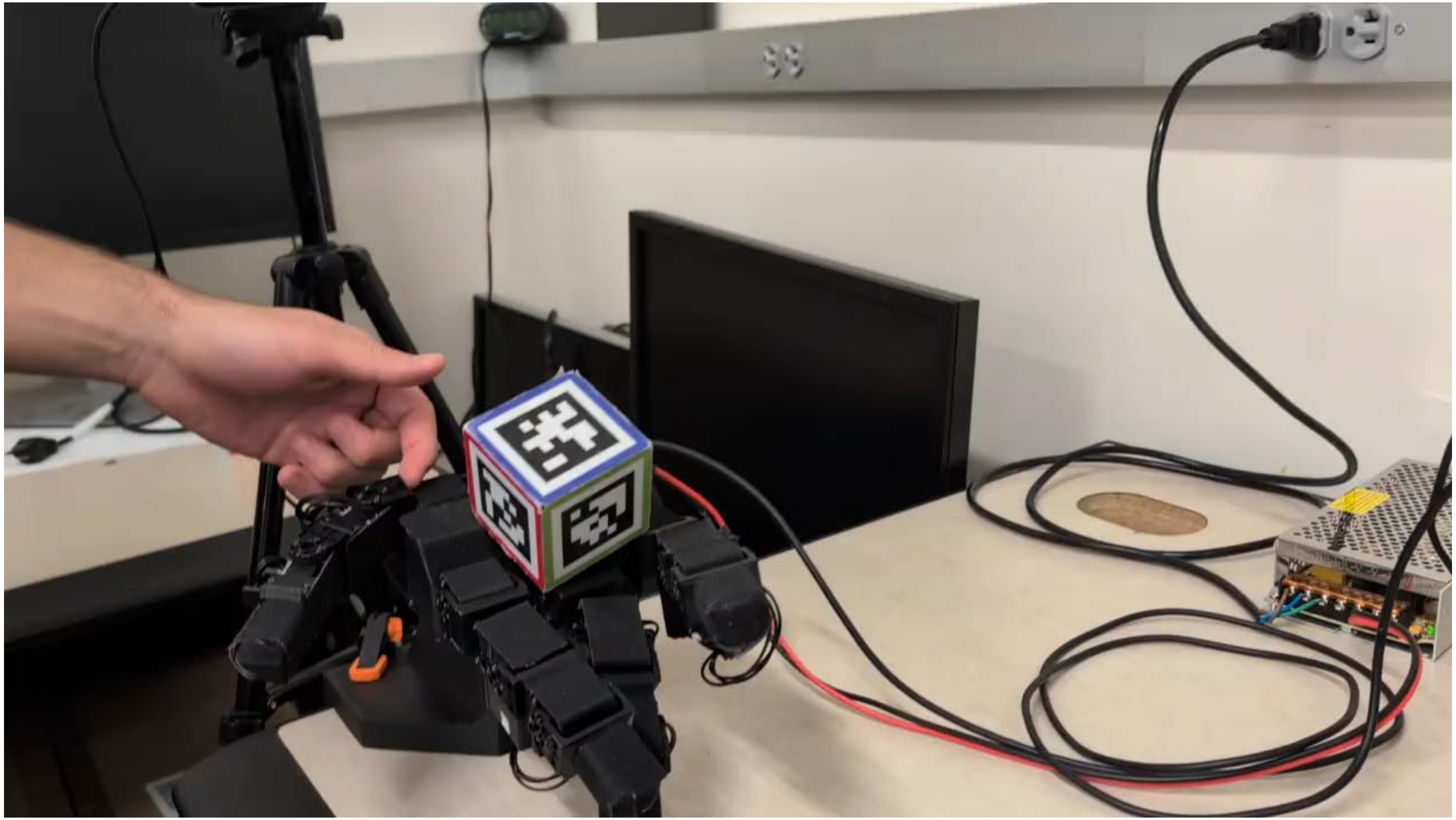
Allegro (4-finger)



Leap Hand (4-finger)



Sim-to-Real Gap

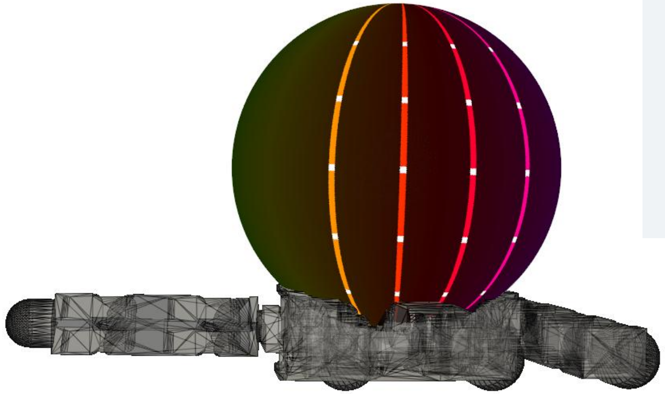


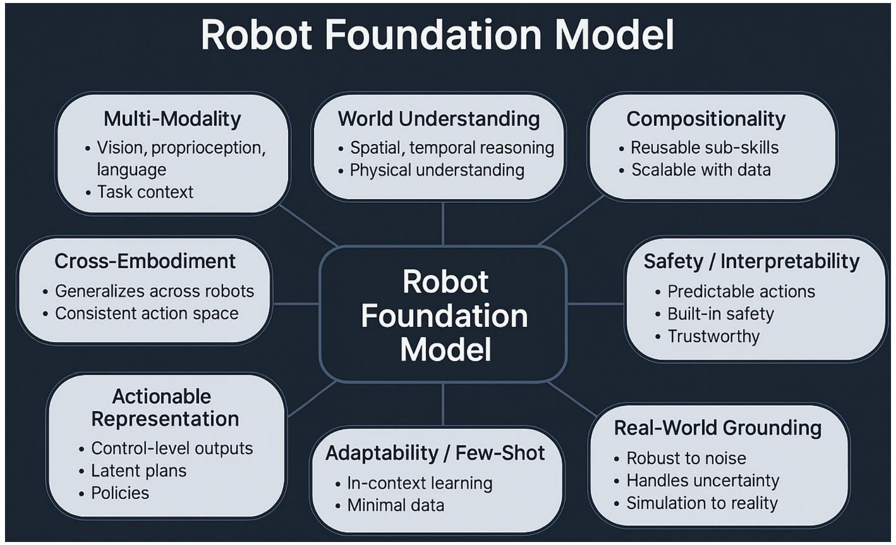
On-going effort

Summary

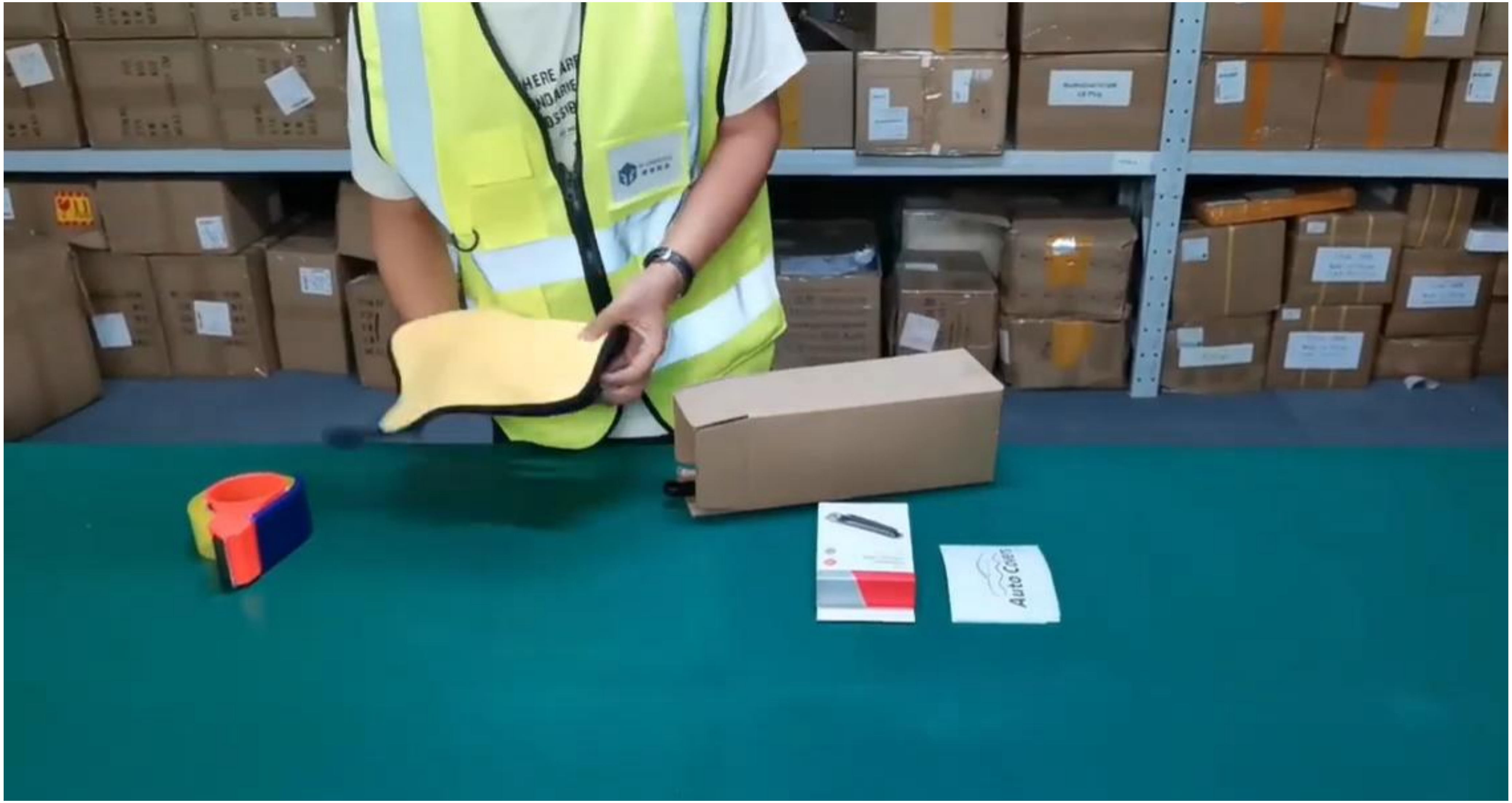
Use data from all robots and human for learning

Unified Gripper Action Space

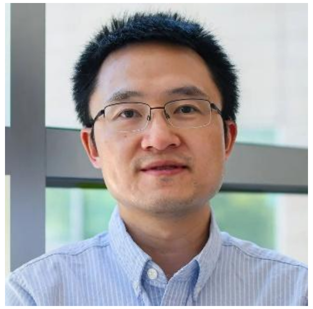




Robot Manipulation is still an Open Challenge



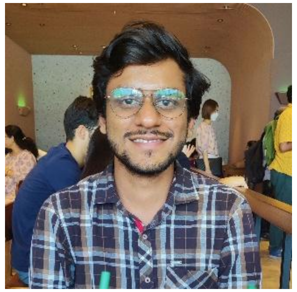
Intelligent Robotics and Vision Lab (IRVL)
















NSF DARPA XPENG

SONY NVIDIA FRONTIER MINING

https://labs.utdallas.edu/irvl/

Assisted by
Ms. Rhonda Walls

Thank you!