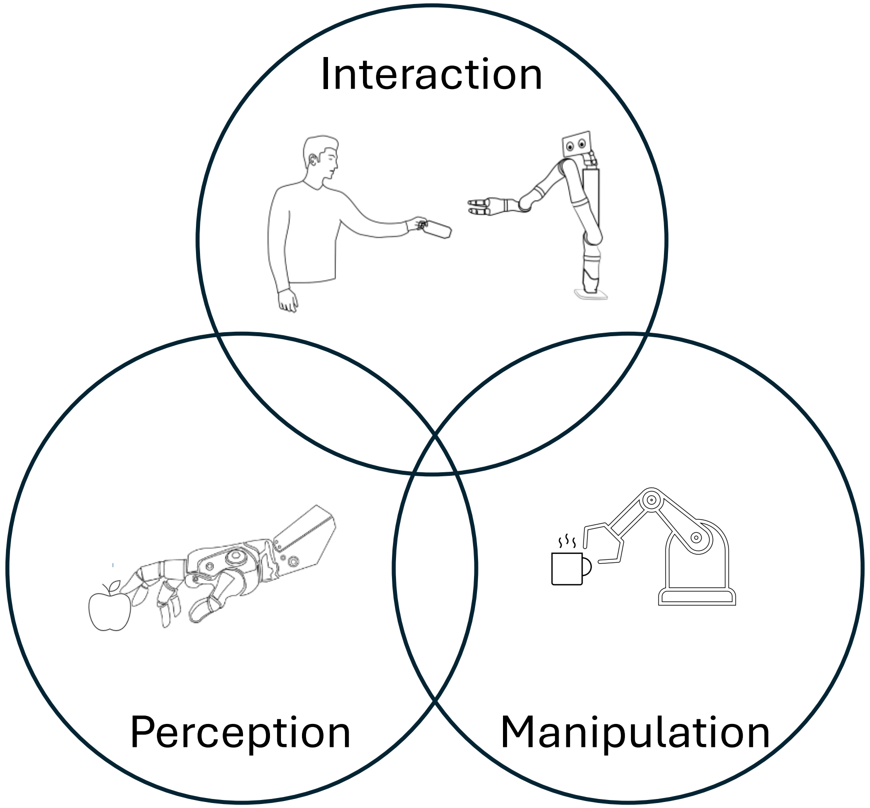
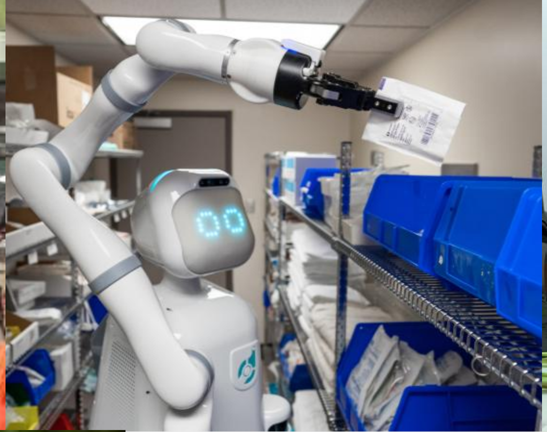
Robot Arms in Action: Interaction, Perception, and Manipulation

Alap Kshirsagar
Postdoctoral Researcher, Intelligent Autonomous Systems Group, TU Darmstadt









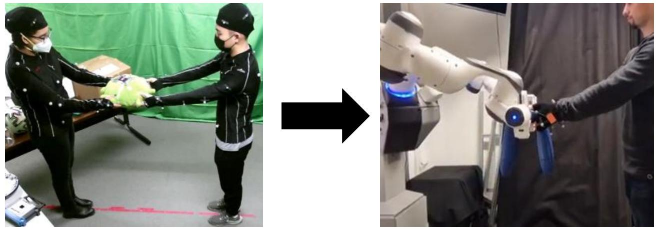


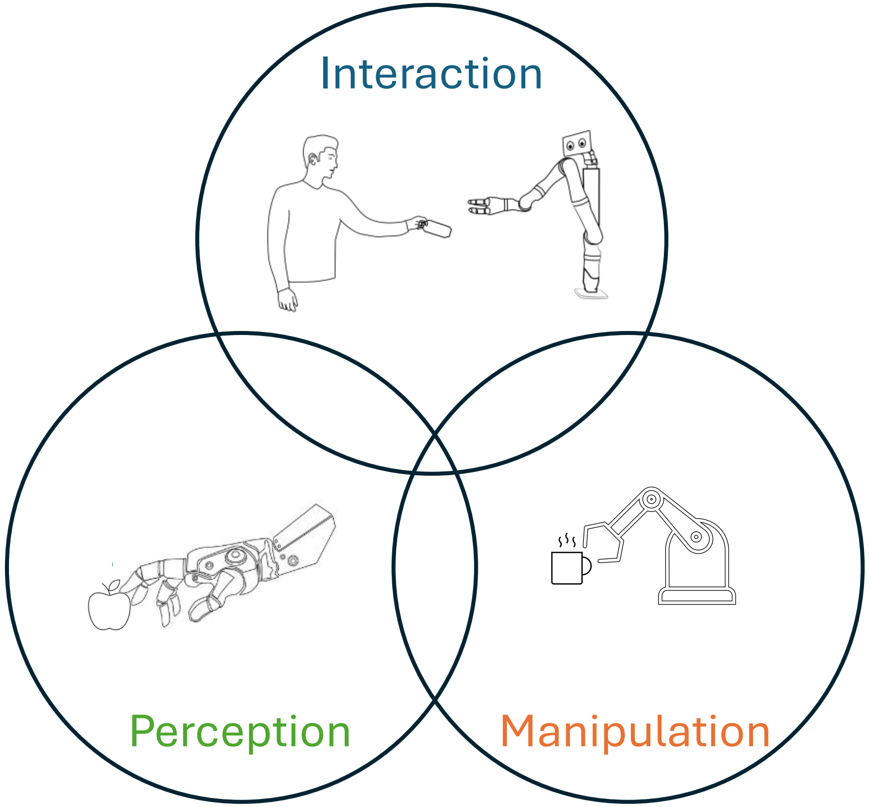


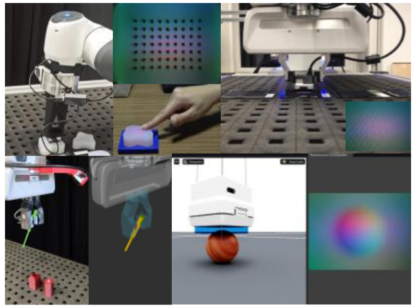
Honda's As

Overview

Learning Human-Robot Interaction from Human Demonstrations





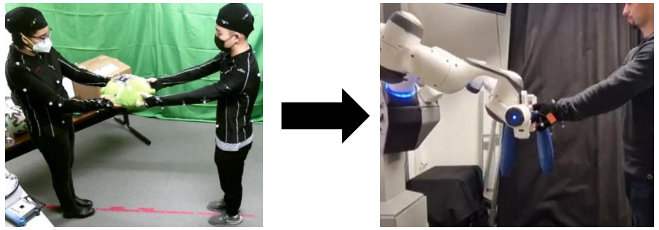


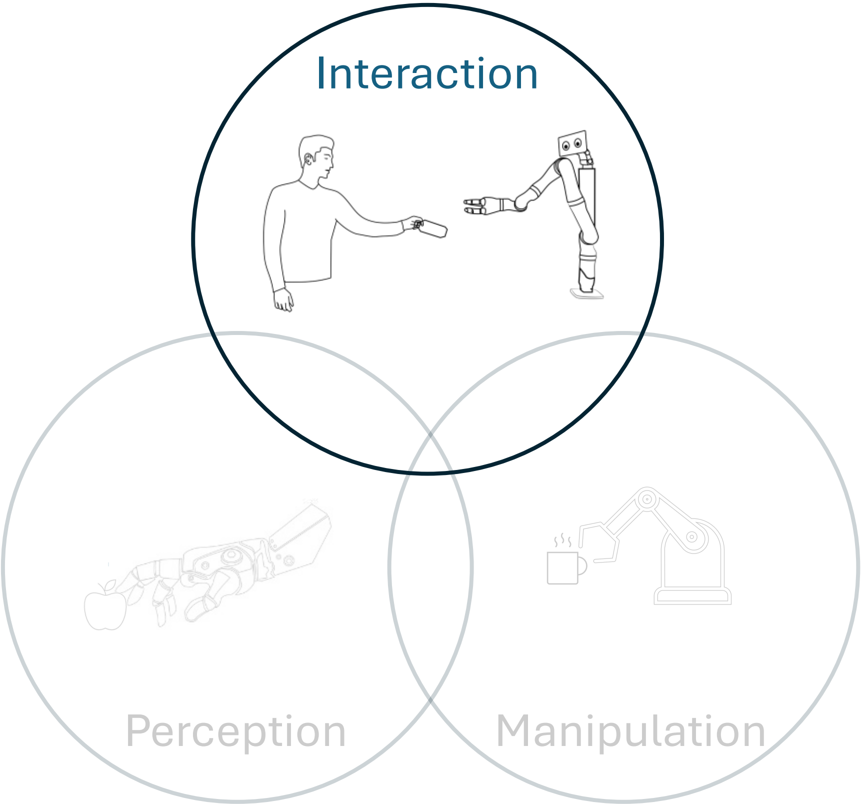
Perceiving Object Properties with Vision-Based Tactile Sensors

Learning Dynamic Manipulations with Reinforcement Learning

Part 1 : Interaction

Learning Human-Robot Interaction from Human Demonstrations






Mango
Strawberry

Handovers – Essential Skill for Collaborative Robots

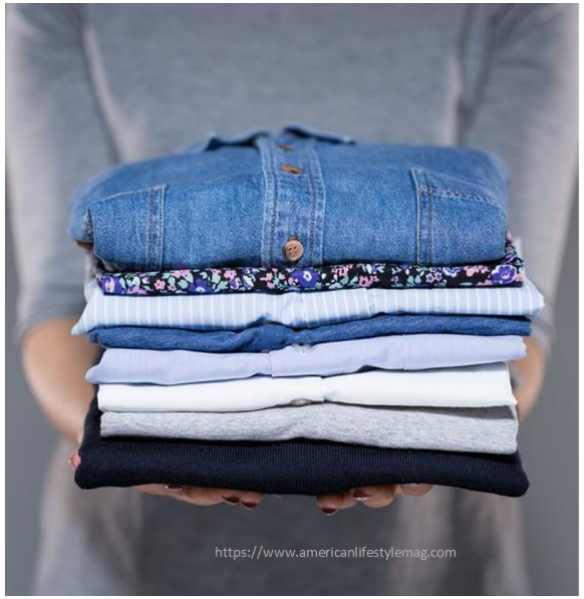
https://www.robotics.org/content-detail.cfm/Industrial-Robotics-Industry-Insights/Collaborative-Robotics-Puts-People-First/content_id/7021

Source: https://www.zdnet.com/pictures/photo-robot-in-

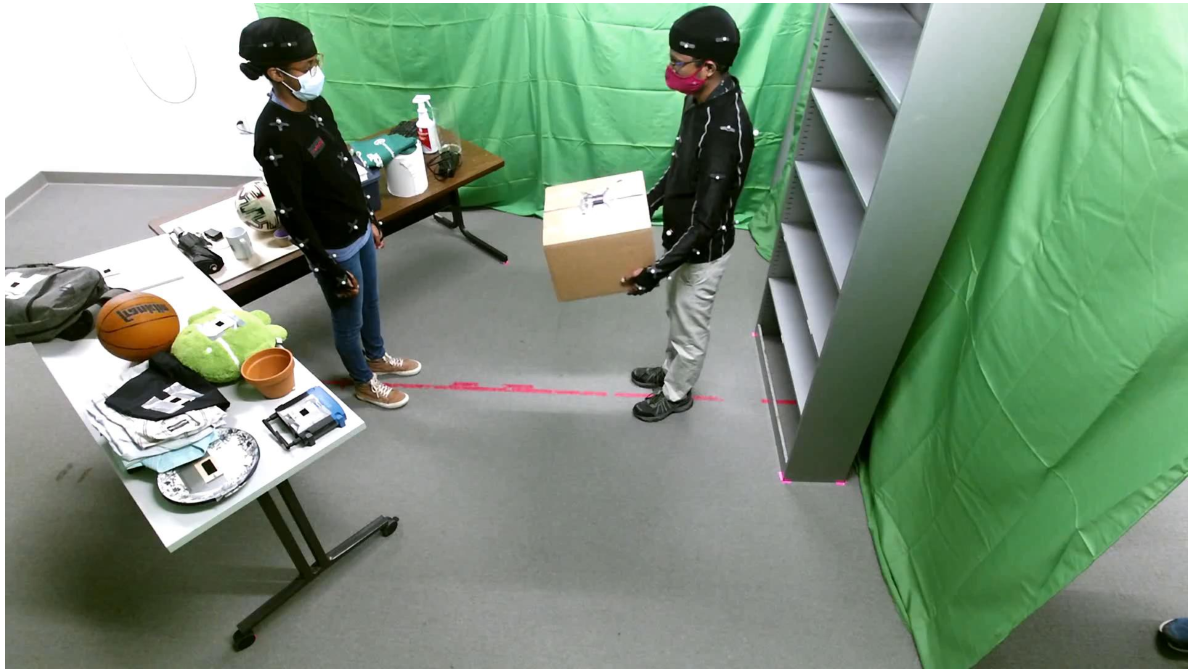
Source: https://www.engineering.com/PLMERP/ArticleID/10519/Can-a-Robot-Help-with-the-

Source: https://sites.google.com/nvidia.com/handovers-of-arbitrary-objects/reactive-human-to-robot-handovers-of-arbitrary-objects

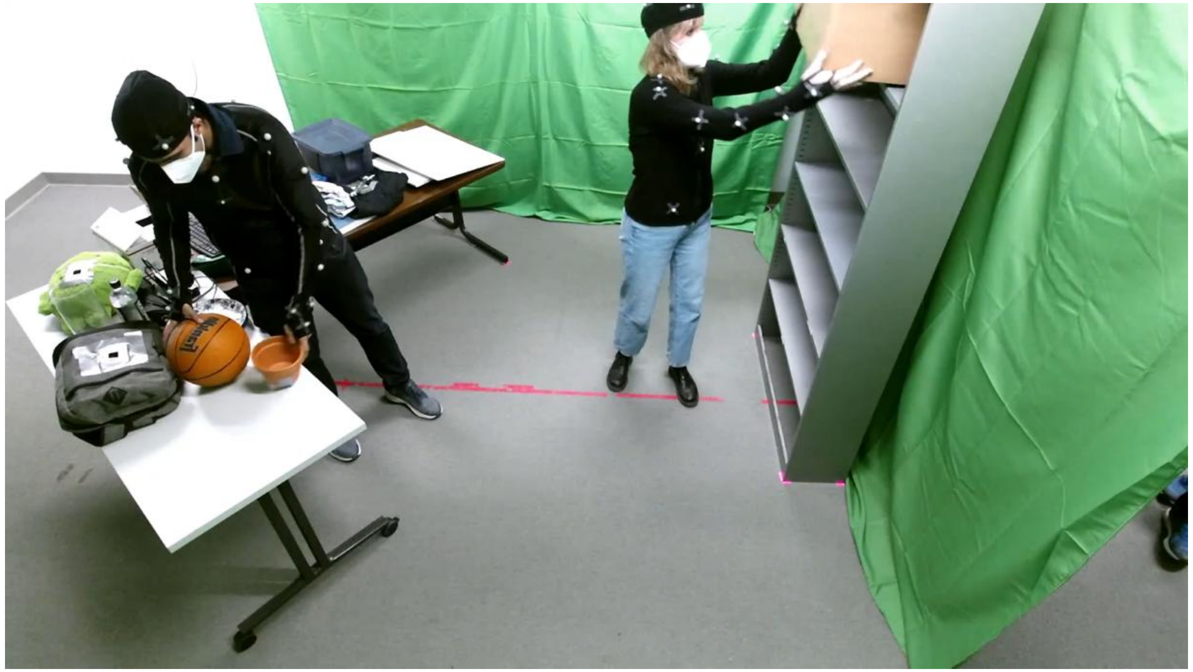
Bimanual Handovers



Multi-Sensor Datasets of Bimanual Human-Human Handovers



Dataset 1: 24 participants, 10 objects, 360 handovers



Dataset 2: 24 participants, 30 objects, 1440 handovers (shelving / un-shelving tasks)

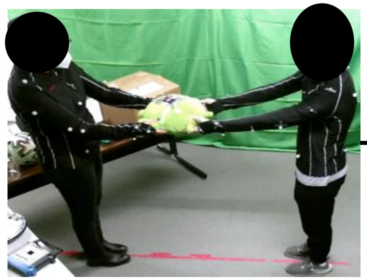
Kshirsagar et al. "Dataset of bimanual human-to-human object handovers." Data in Brief (2023)

Kshirsagar et al. "Multi-sensor Dataset of Multiple Sequential Human-to-Human Object Handovers in Shelving and Un-shelving Tasks.“ Data Descriptions (2025)

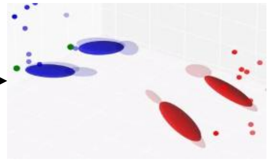
Learning Bimanual Robot-to-Human Handovers

Offline

Human Demonstration



Train Hidden Semi-Markov Model (HSMM)

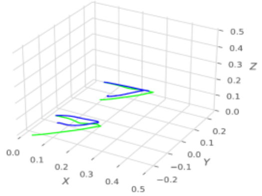


Trained HSMM

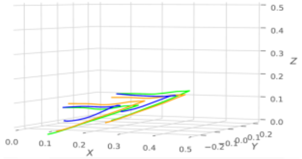
Online

Observe Human's Trajectory upto Current Time

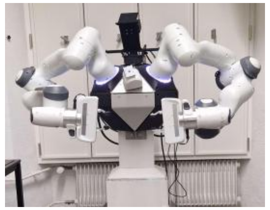
Compute Robot's Position with HSMM conditioning



Optimize Robot's Position with Grip-width Constraint



Apply Inverse Kinematics and Send Joint Angles



Goeksu et al. "Kinematically Constrained Human-like Bimanual Robot-to-Human Handovers" HRI-LBR (2024)

Learning Bimanual Robot-to-Human Handovers

Preliminary Study (4 participants, 3 objects)

Metric Scale 1-5	Baseline Median	Proposed Median
Humanlike	1	**2**
Sensible	2	**3**

p-value < 0.05

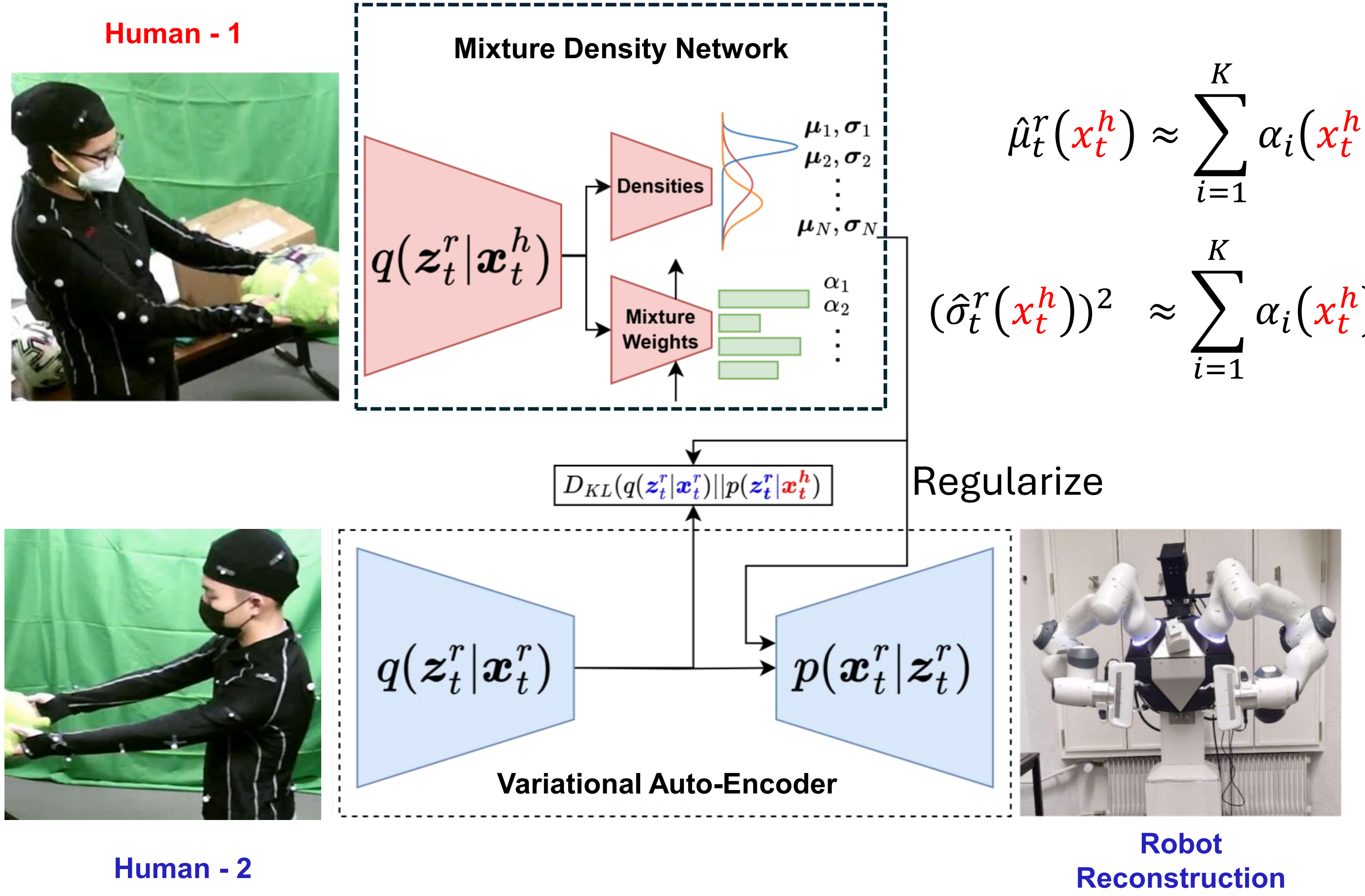
Goeksu et al. "Kinematically Constrained Human-like Bimanual Robot-to-Human Handovers" HRI-LBR (2024)

Typical Interactions






Mixture of Variational Experts for Interaction

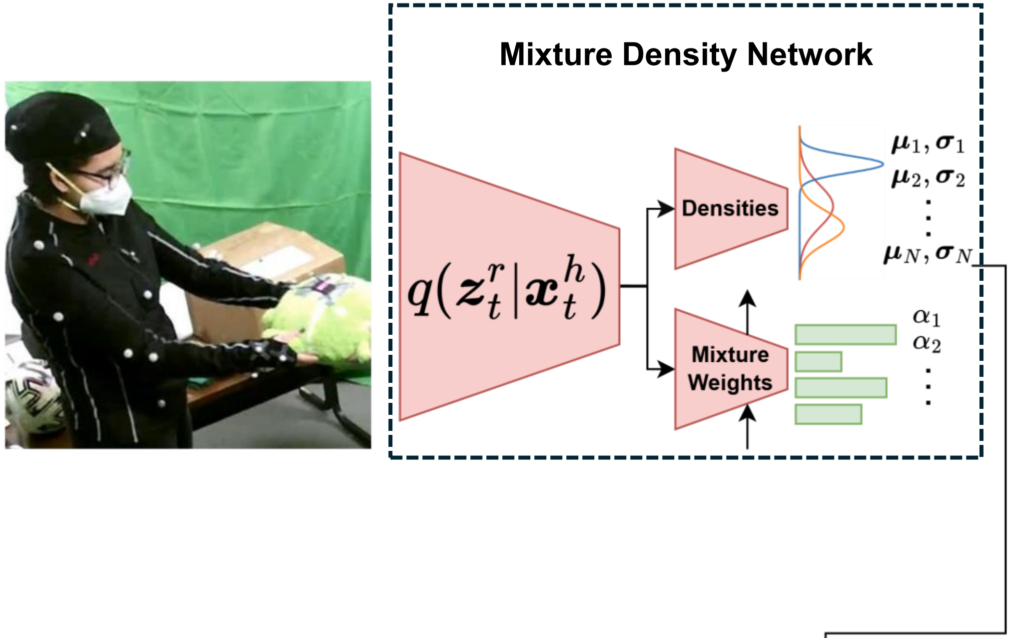


$$\hat{\mu}_t^r(x_t^h) \approx \sum_{i=1}^{K} \alpha_i(x_t^h)\mu_i^r(x_t^h)$$

$$(\hat{\sigma}_t^r(x_t^h))^2 \approx \sum_{i=1}^{K} \alpha_i(x_t^h)\,(\sigma_i^r(x_t^h))^2$$

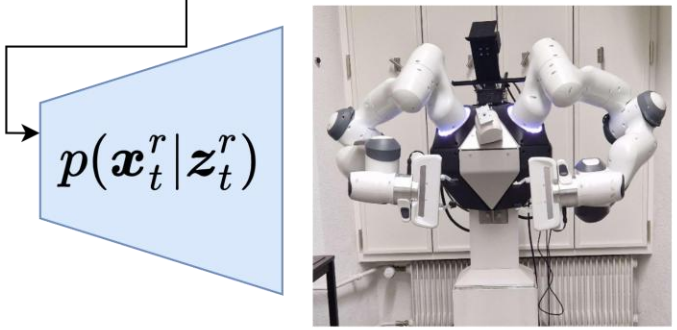
Prasad et al. “MoVEInt: Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations”, RA-L (2024)

Mixture of Variational Experts for Interaction



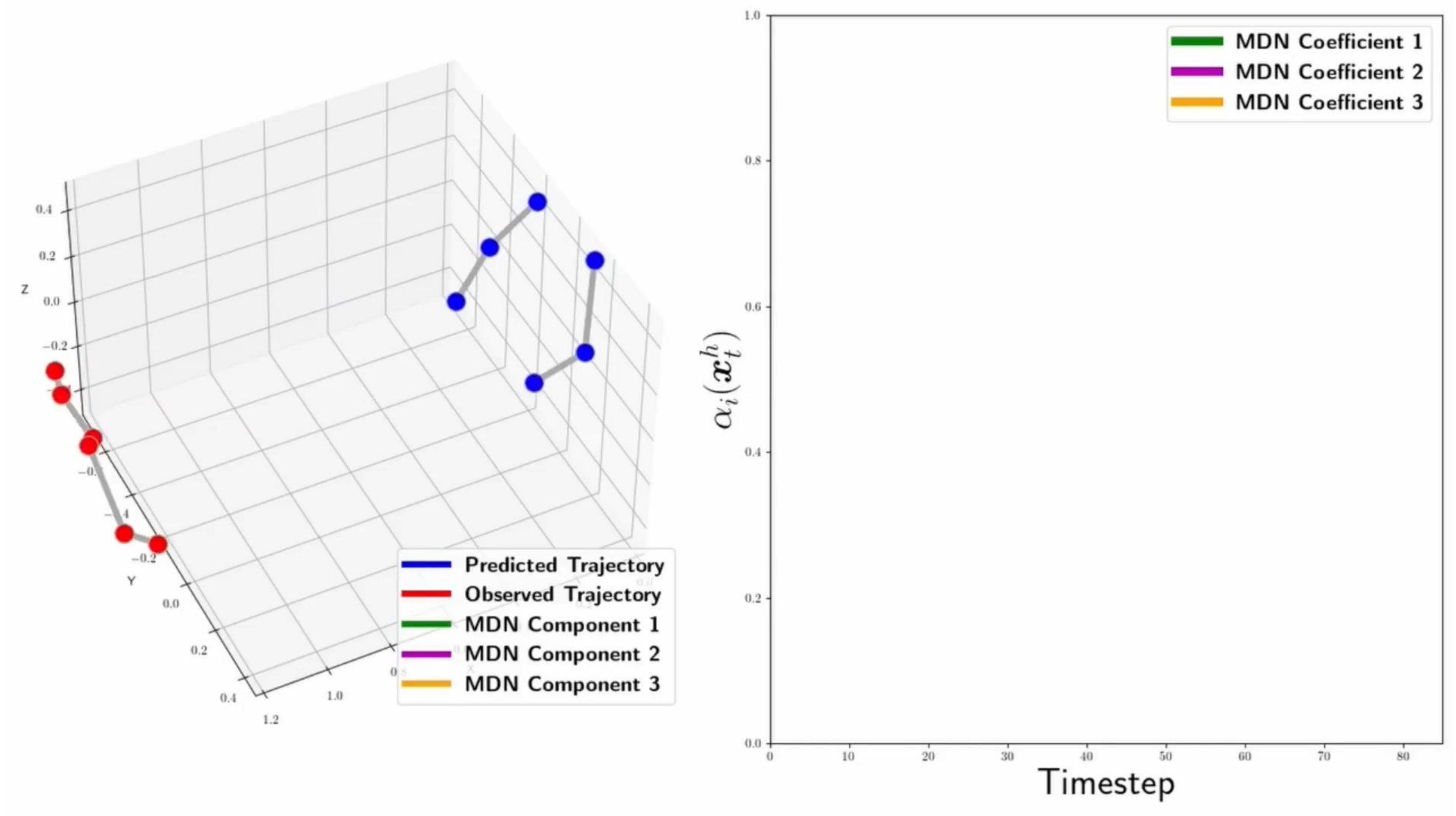
$$\hat{\mu}_t^r(x_t^h) \approx \sum_{i=1}^{K} \alpha_i(x_t^h)\mu_i^r(x_t^h)$$

$$(\hat{\sigma}_t^r(x_t^h))^2 \approx \sum_{i=1}^{K} \alpha_i(x_t^h)\,(\sigma_i^r(x_t^h))^2$$



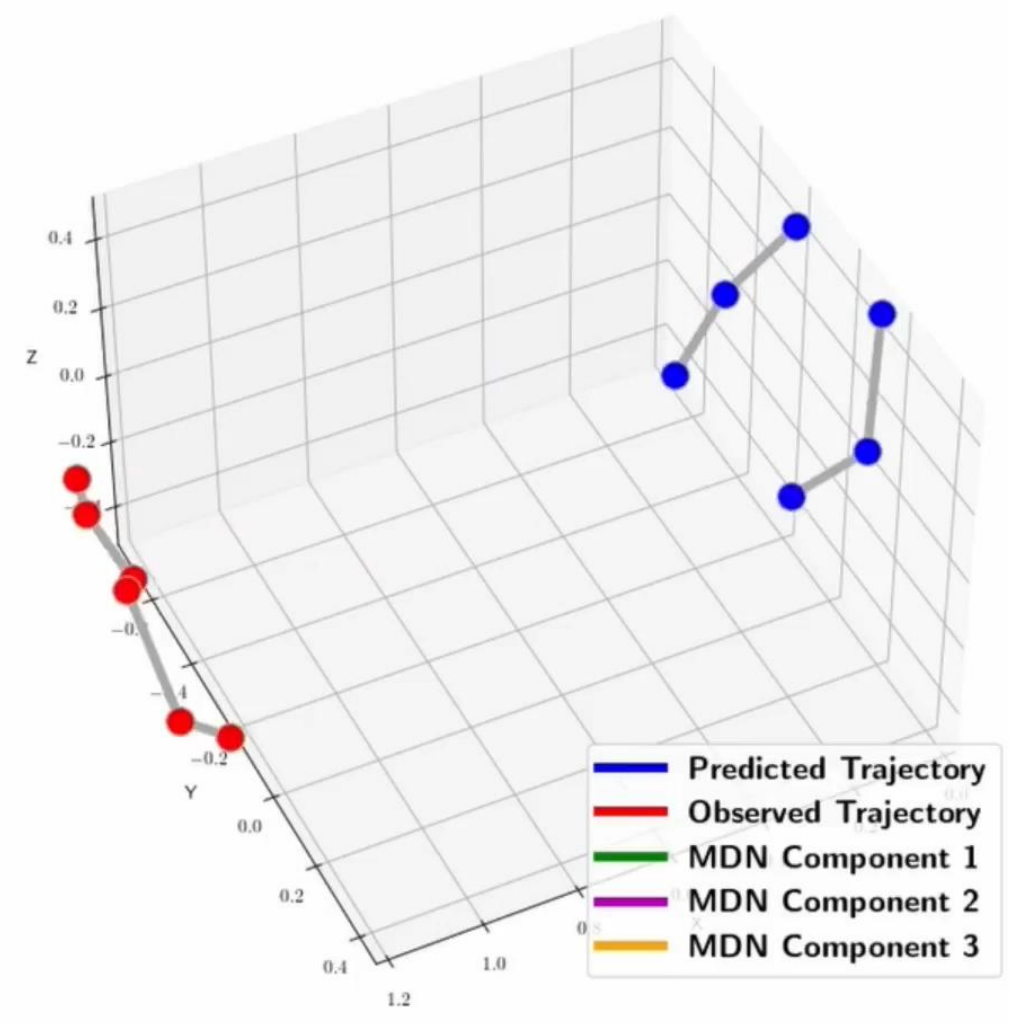
Prasad et al. "MoVEInt: Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations", RA-L (2024)

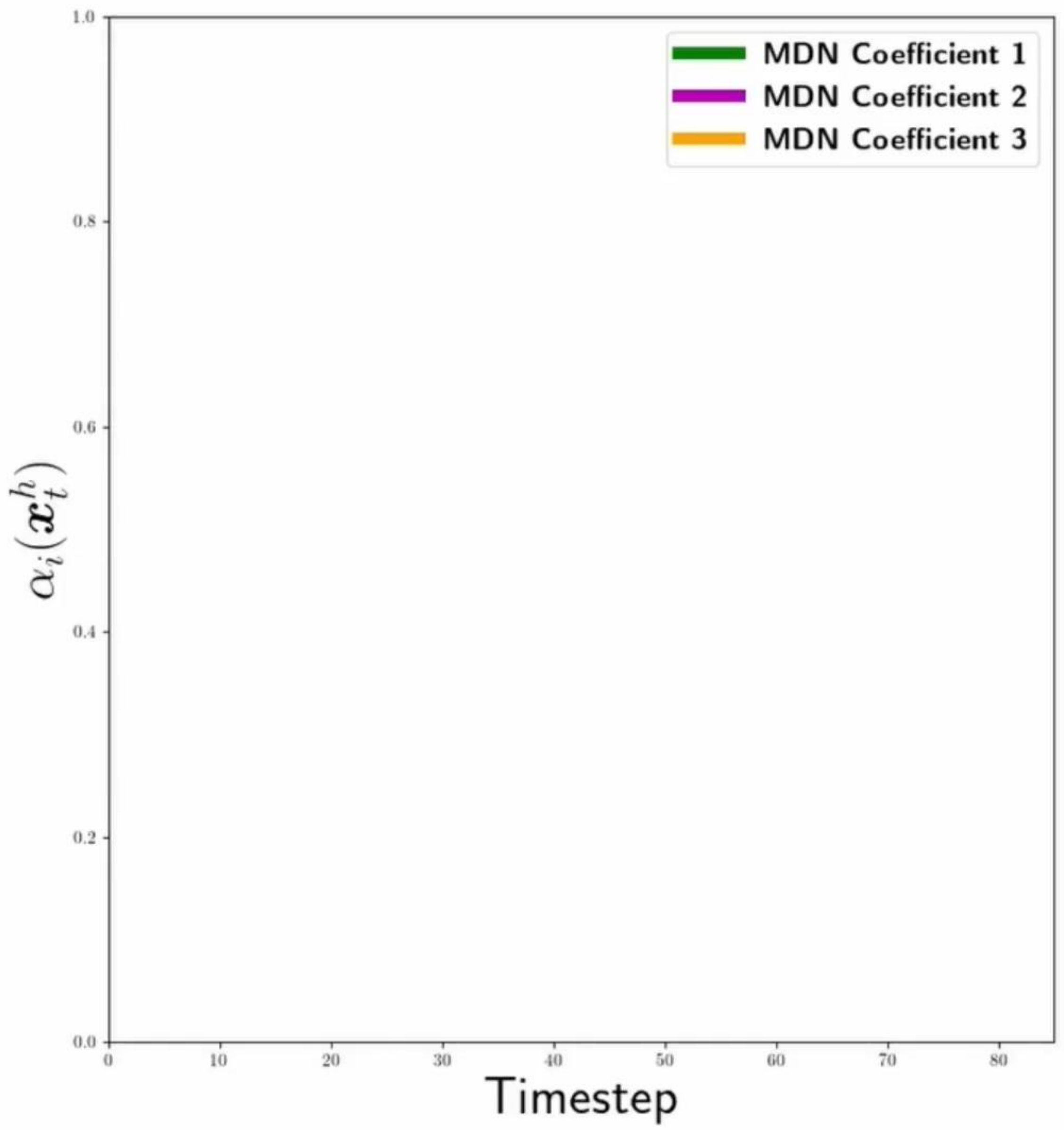
Demonstration : Reactiveness



Prasad et al. "MoVEInt: Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations", RA-L (2024)

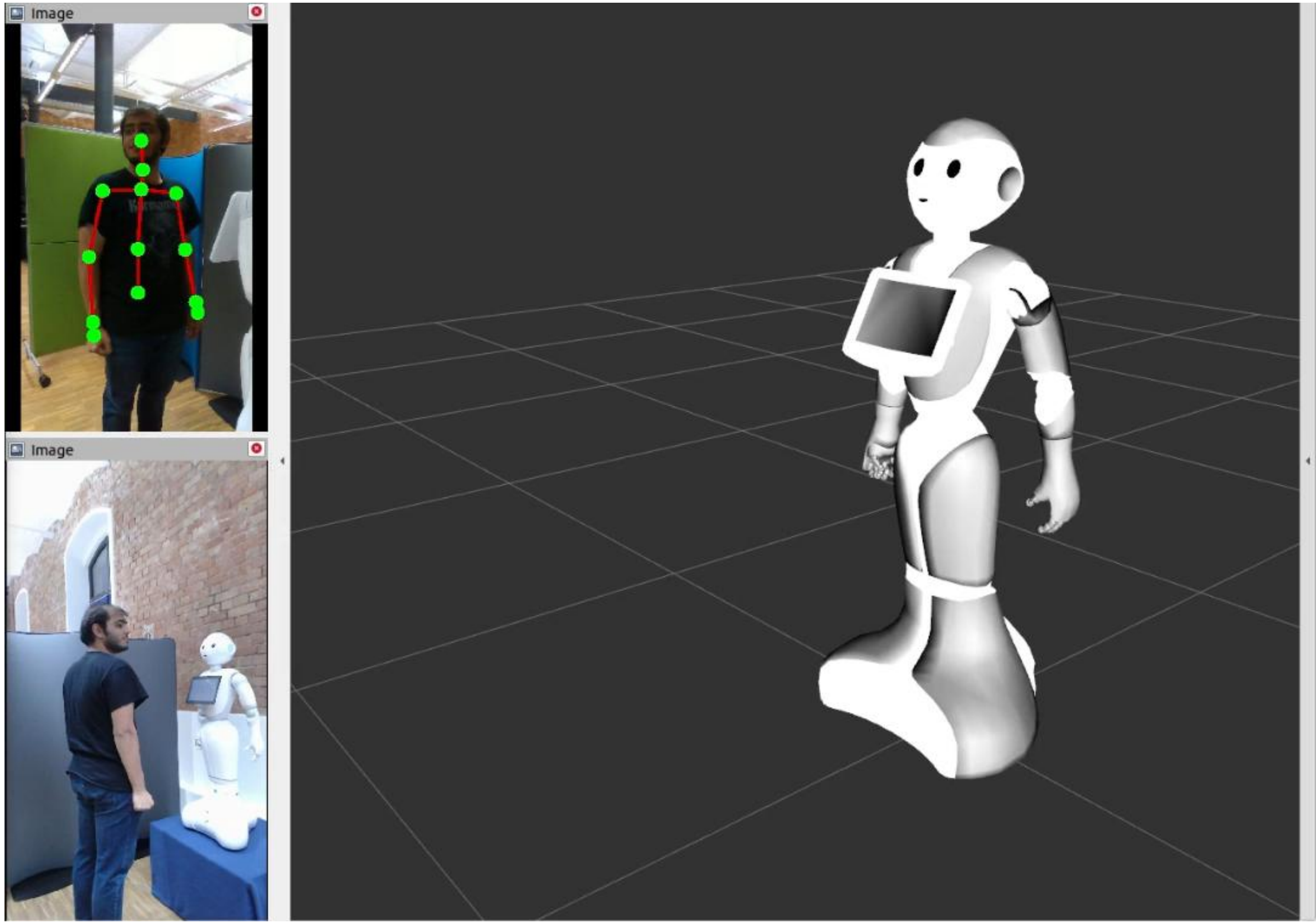
Demonstration : Reactiveness





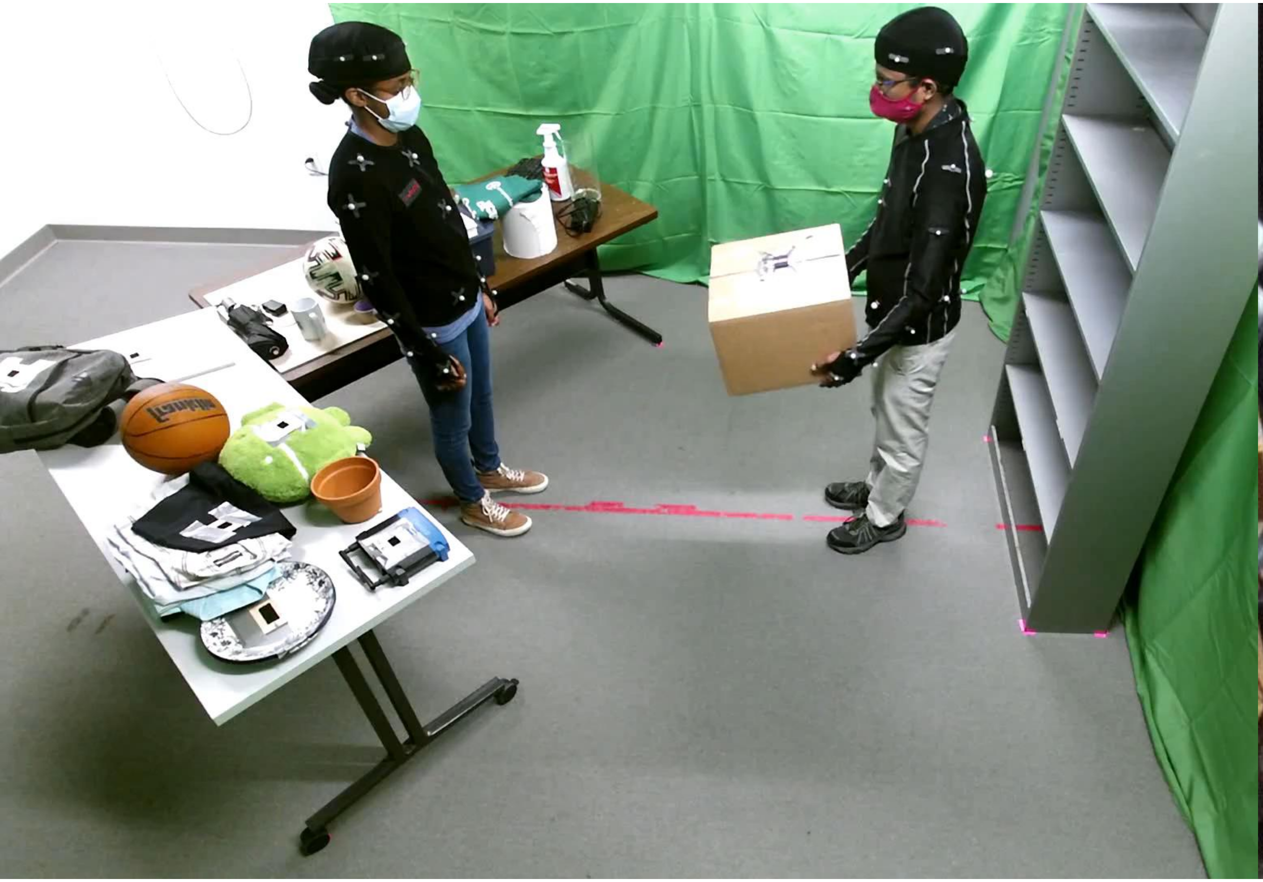
Prasad et al. "MoVEInt: Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations", RA-L (2024)

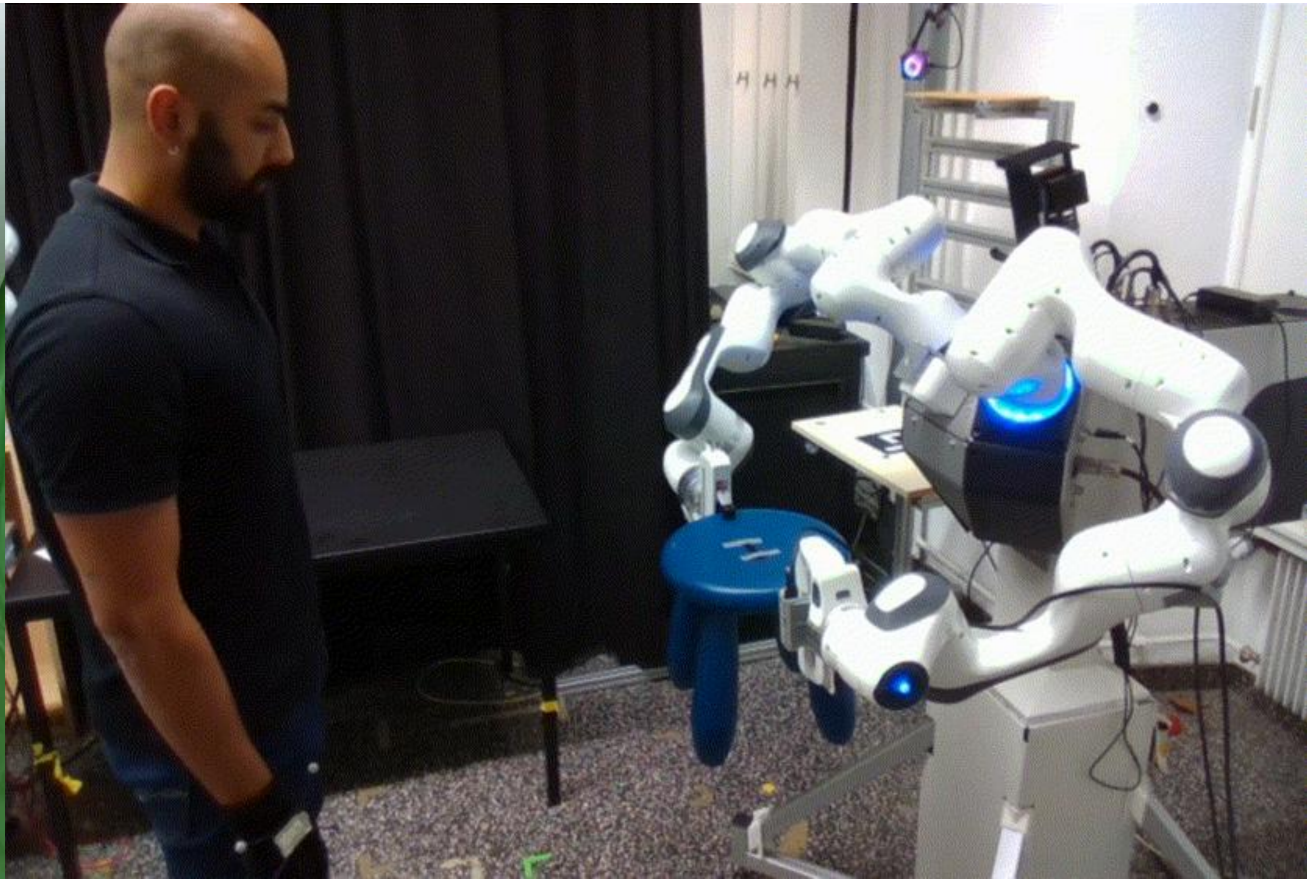
Demonstration : Rocket Fist-bump



Prasad et al. “MoVEInt: Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations”, RA-L (2024)

Demonstration : Bimanual Handovers





Prasad et al. “MoVEInt: Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations”, RA-L (2024)

Mean-squared Errors in Reconstructed Trajectories

Datasets


Bütepage et al. 2020


Prasad et al. 2023


Kshirsagar et al. 2023

Dataset (units)	Action	[Prasad et al. 2023]	[Bütepage et al. 2020]	Our Method
HHI [Bütepage et al. 2020] (cm)	Hand Wave	0.788 ± 1.226	4.121 ± 2.252	**0.448 ± 0.630**
	Handshake	1.654 ± 1.549	1.181 ± 0.859	**0.196 ± 0.153**
	Rocket Fistbump	0.370 ± 0.682	0.544 ± 1.249	**0.123 ± 0.175**
	Parachute Fistbump	0.537 ± 0.579	0.977 ± 1.141	**0.314 ± 0.348**
HRI-Pepper [Bütepage et al. 2020] (rad)	Hand Wave	0.103 ± 0.103	0.664 ± 0.277	**0.087 ± 0.089**
	Handshake	0.056 ± 0.041	0.184 ± 0.141	**0.015 ± 0.014**
	Rocket Fistbump	0.018 ± 0.035	0.033 ± 0.045	**0.007 ± 0.015**
	Parachute Fistbump	0.088 ± 0.148	0.189 ± 0.196	**0.048 ± 0.112**
HRI-Yumi [Bütepage et al. 2020] (rad)	Hand Wave	1.033 ± 1.204	0.225 ± 0.302	**0.147 ± 0.072**
	Handshake	0.068 ± 0.052	0.133 ± 0.214	**0.057 ± 0.044**
	Rocket Fistbump	0.128 ± 0.071	0.147 ± 0.119	**0.093 ± 0.045**
	Parachute Fistbump	**0.028 ± 0.034**	0.181 ± 0.155	0.081 ± 0.082
HHI [Prasad et al. 2023] (cm)	Hand Wave	0.408 ± 0.538	3.168 ± 3.392	**0.298 ± 0.274**
	Handshake	0.311 ± 0.259	1.489 ± 3.327	**0.149 ± 0.120**
	Rocket Fistbump	1.142 ± 1.375	3.576 ± 3.082	**0.673 ± 0.679**
	Parachute Fistbump	0.453 ± 0.578	2.008 ± 2.024	**0.291 ± 0.199**
HRI-Pepper [Prasad et al. 2023] (rad)	Hand Wave	0.046 ± 0.059	0.057 ± 0.093	**0.044 ± 0.048**
	Handshake	0.020 ± 0.014	0.083 ± 0.075	**0.011 ± 0.008**
	Rocket Fistbump	0.077 ± 0.067	0.101 ± 0.086	**0.045 ± 0.045**
	Parachute Fistbump	0.022 ± 0.027	0.049 ± 0.040	**0.017 ± 0.014**
HHI-Handovers [Kshirsagar et al. 2023] (cm)	Unimanual	0.441 ± 0.280	1.133 ± 0.721	**0.441 ± 0.221**
	Bimanual	0.869 ± 0.964	0.990 ± 0.764	**0.685 ± 0.643**

[Prasad et al. 2023] – VAE + HMM latent space
[Bütepage et al. 2020] – Unimodal recurrent VAE

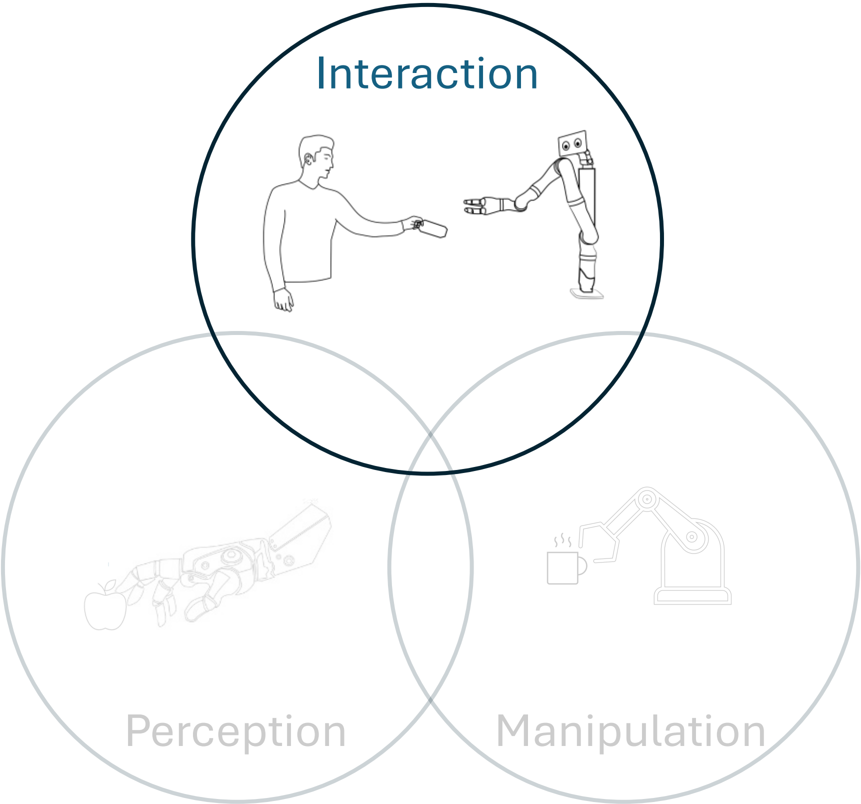
Prasad et al. "MoVEInt: Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations", RA-L (2024)

Part 1 : Interaction

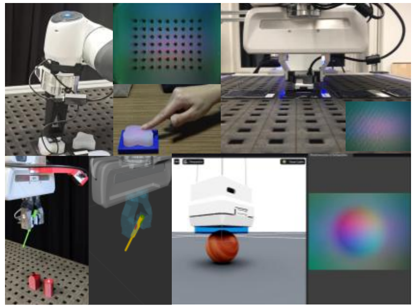
Learning Human-Robot Interaction from Human Demonstrations

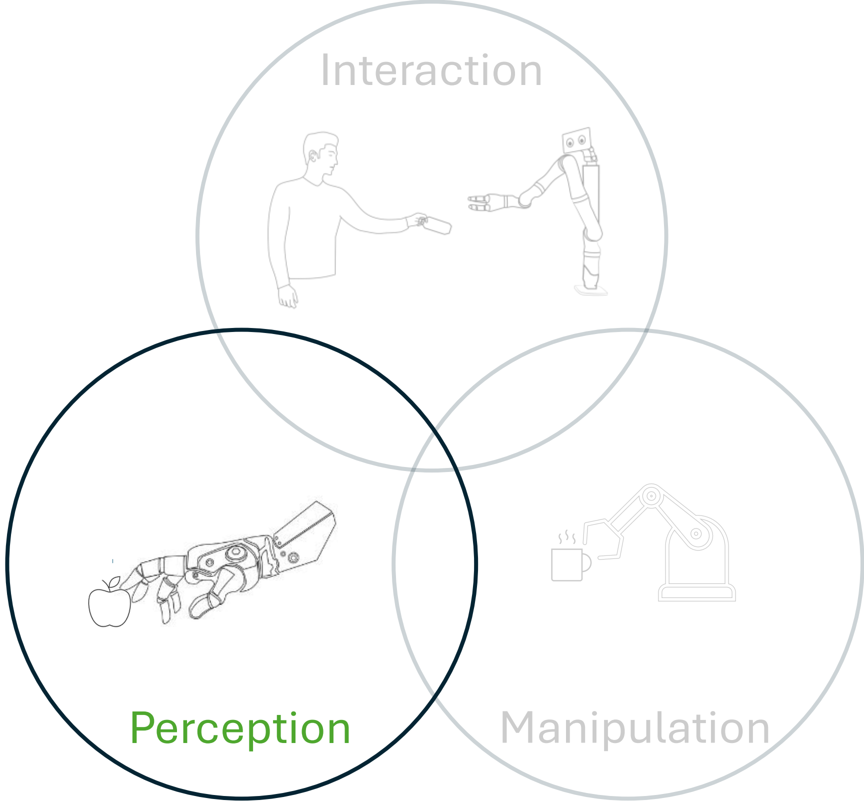
HSMM + Constrained Optimization

Mixture Density Network + VAE



Part 2 : Perception





Perceiving Object Properties with Vision-Based Tactile Sensors

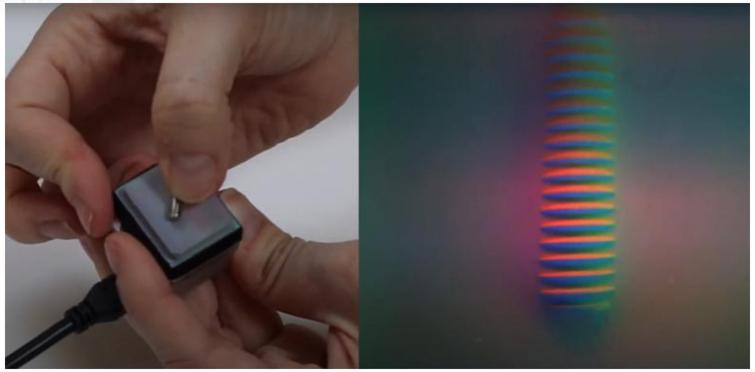
Vision-Based Tactile Sensors

Digit (Meta AI Research)

Source: https://digit.ml/

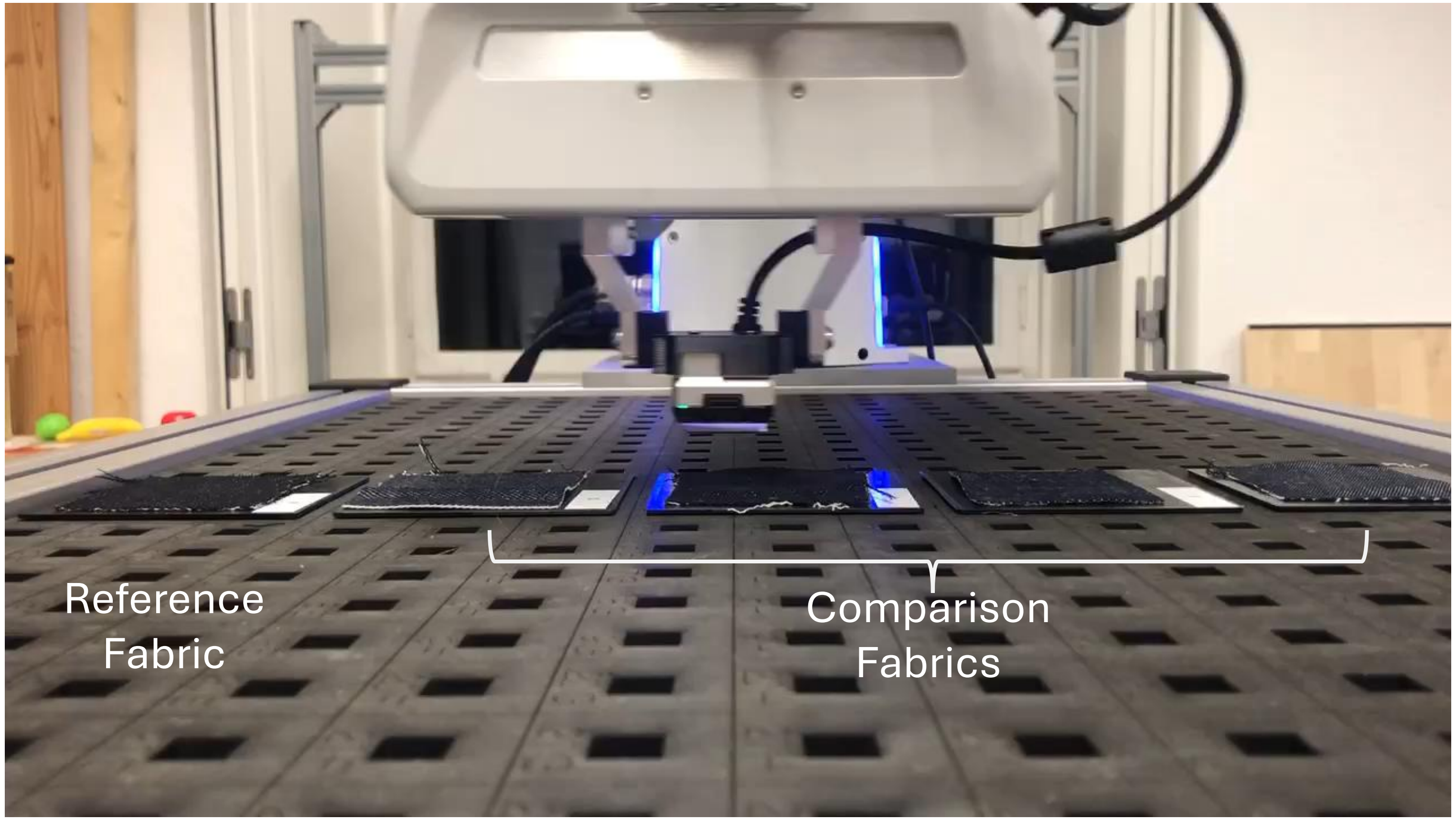
Gelsight Mini

Source: https://www.gelsight.com



Texture Recognition With Vision-Based Tactile Sensors

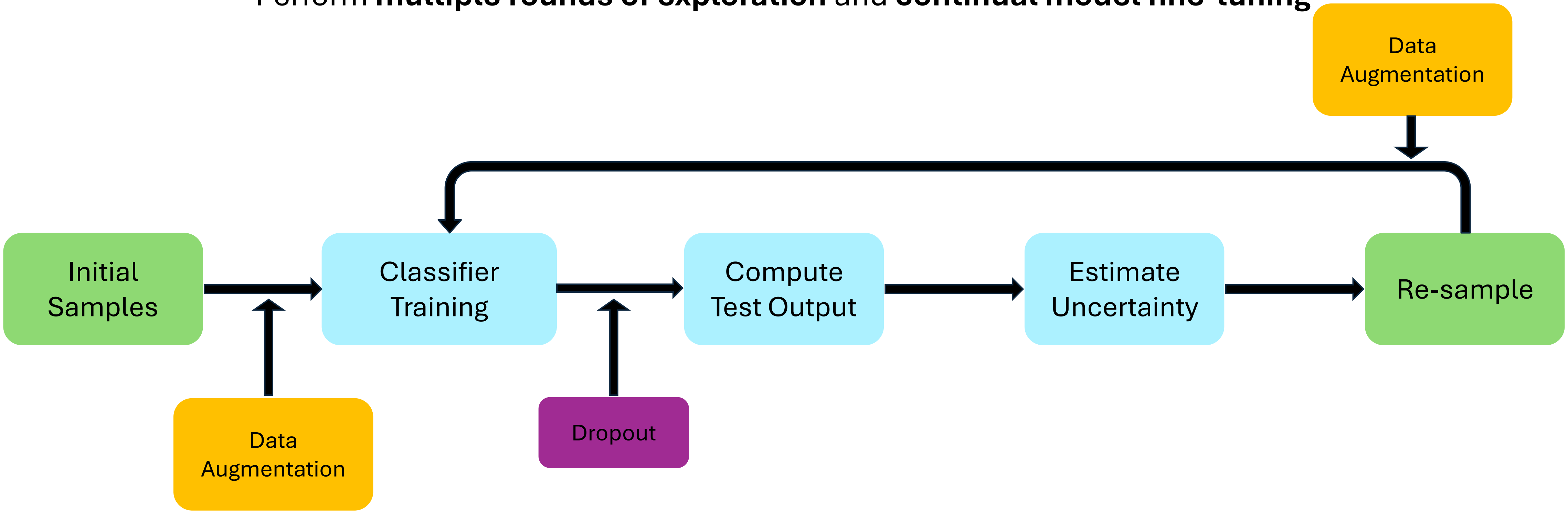
Task: Find reference fabric among comparison fabrics in **as few touches as possible**



Boehm et al. “What Matters for Active Texture Recognition with Vision-Based Tactile Sensors”, ICRA (2024)

Tactile Active Recognition of Textures (TART)

Perform **multiple rounds of exploration** and **continual model fine-tuning**



Re-sampling Strategies: Touch the next fabric based on the **model uncertainty**

1. Variance Strategy

2. Entropy Strategy

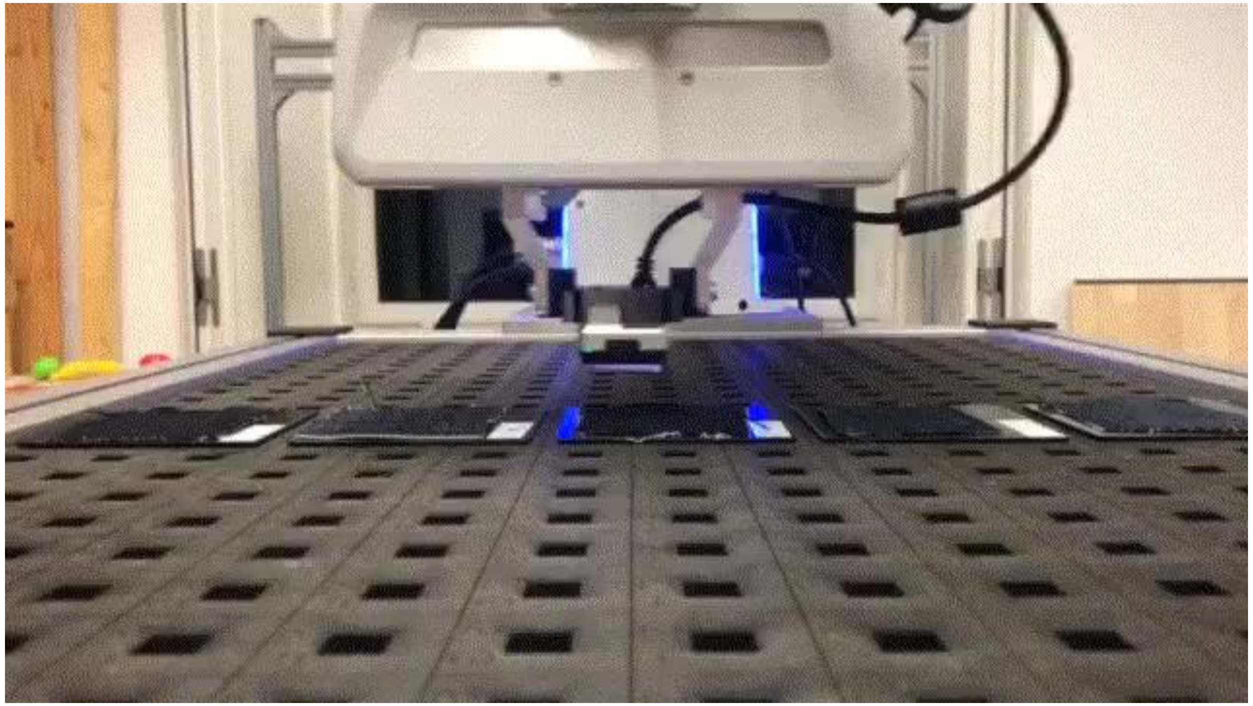
3. Random Strategy

4. You Only Touch Once (YOTO) Strategy

Boehm et al. “What Matters for Active Texture Recognition with Vision-Based Tactile Sensors”, ICRA (2024)

Tactile Active Recognition of Textures (TART)

Perform **multiple rounds of exploration** and **continual model fine-tuning**



Re-sampling Strategies: Touch the next fabric based on the **model uncertainty**

1. Variance Strategy

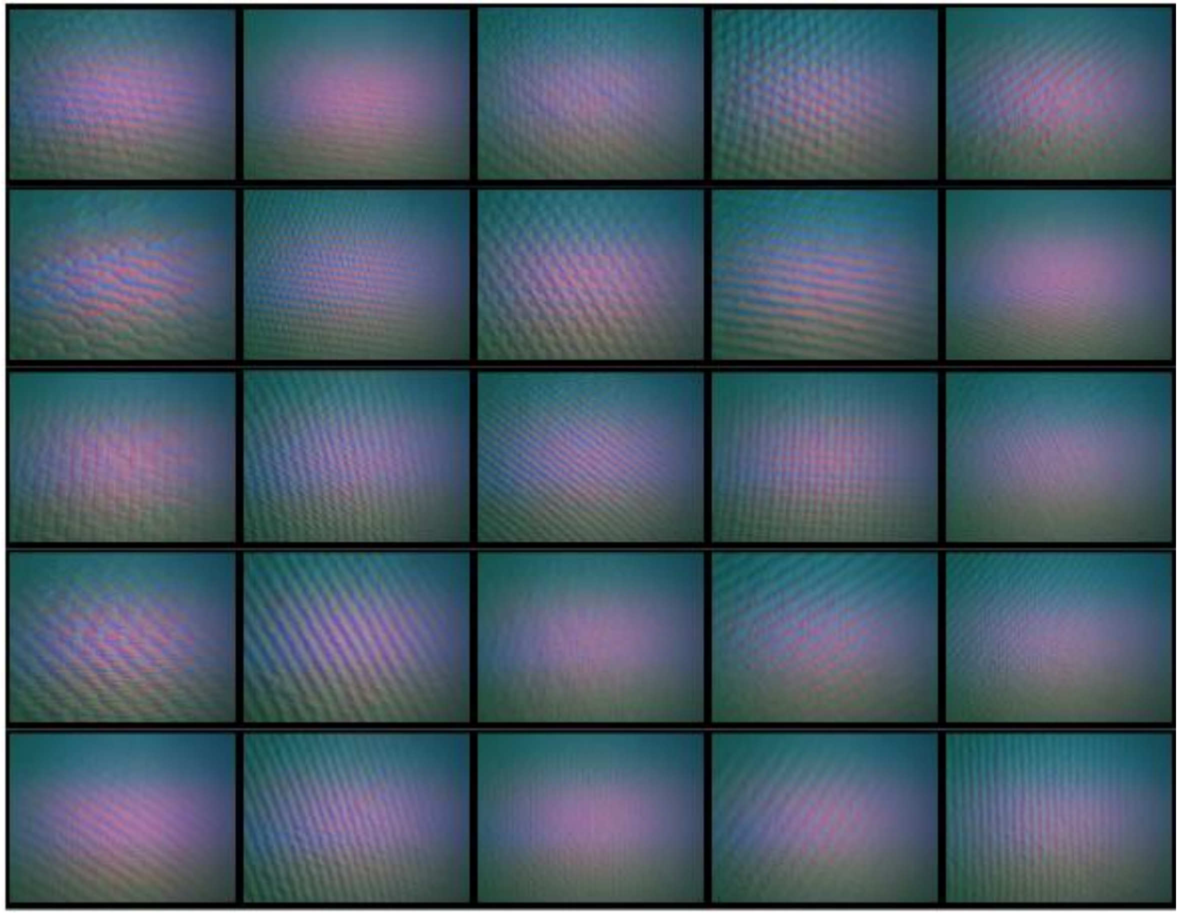
2. Entropy Strategy

3. Random Strategy

4. You Only Touch Once (YOTO) Strategy

Boehm et al. "What Matters for Active Texture Recognition with Vision-Based Tactile Sensors", ICRA (2024)

Texture Dataset



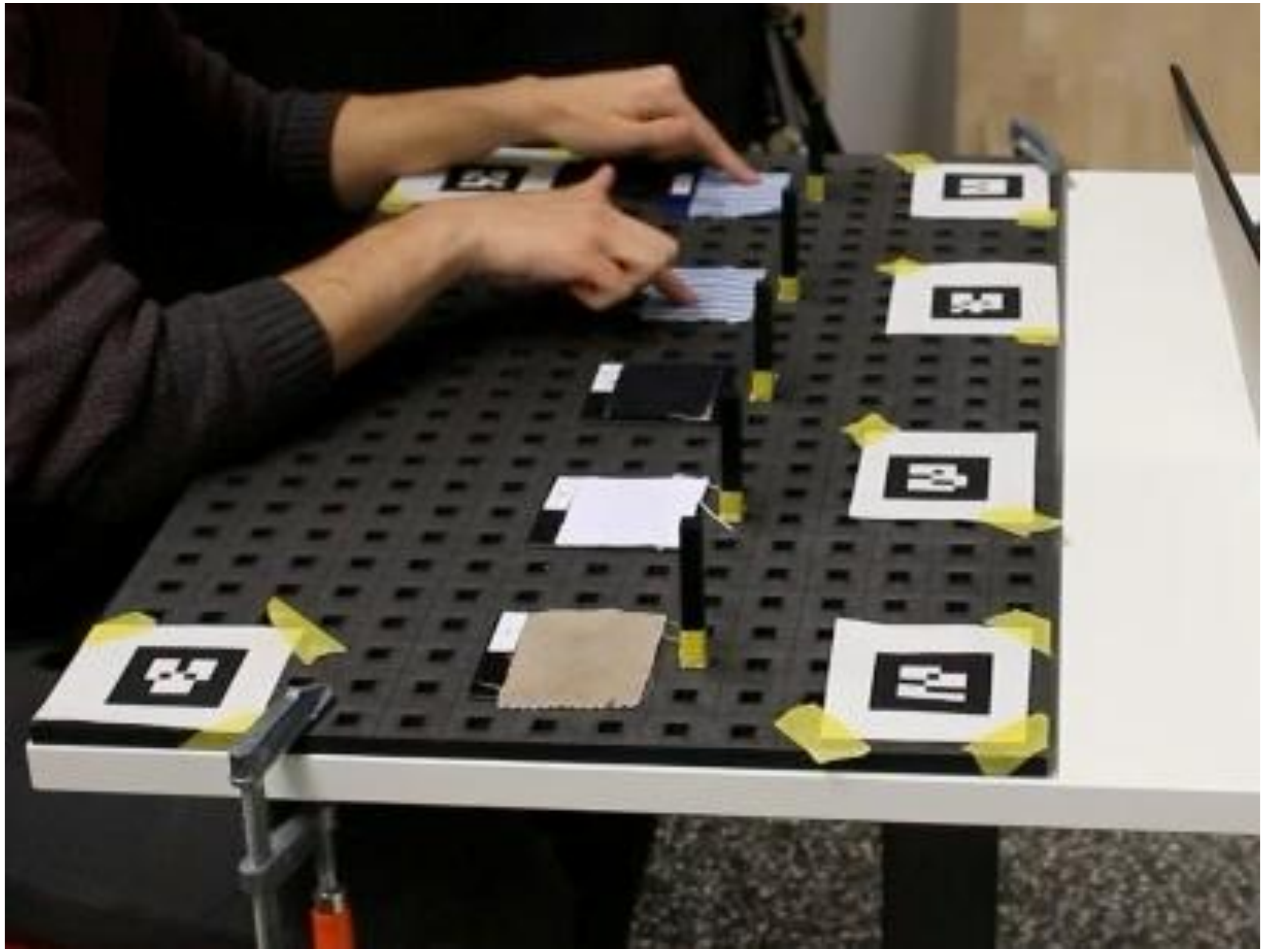
200 texture images each for 25 denim and cotton fabrics

(hard to distinguish by touch)

Boehm et al. “What Matters for Active Texture Recognition with Vision-Based Tactile Sensors”, ICRA (2024)

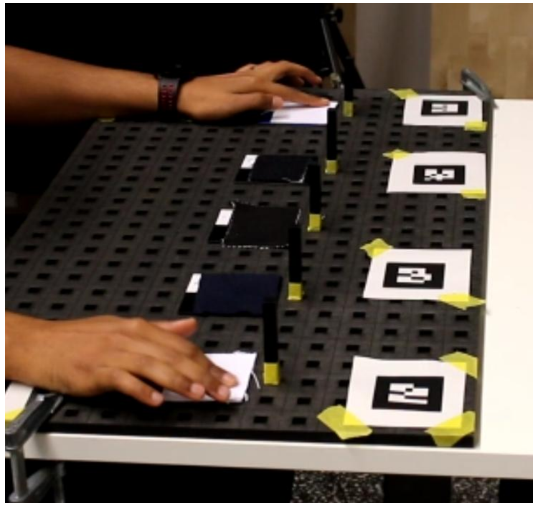
Human Study – Texture Recognition

Blindfolded participants (n=10) did the texture recognition task



Boehm et al. "What Matters for Active Texture Recognition with Vision-Based Tactile Sensors", ICRA (2024)

Texture Recognition Accuracy



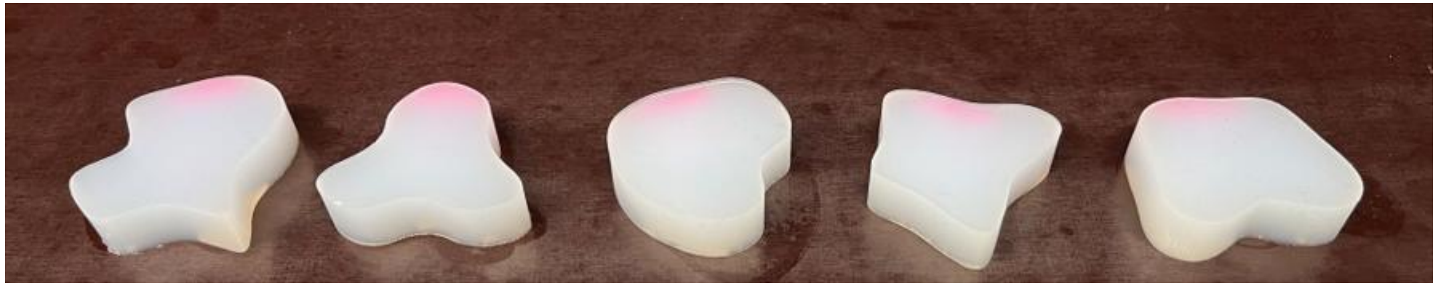
Humans	*Variance*	*Entropy*	*Random*	*YOTO*
66.88% ±16.93%	**90.00%** ±15.24%	88.13% ±14.24%	89.38% ±14.35%	80.63% ±22.42%

Ablation: Dropout rate and data augmentation matter more than active sampling strategy

Boehm et al. "What Matters for Active Texture Recognition with Vision-Based Tactile Sensors", ICRA (2024)

Hardness Recognition With Vision-Based Tactile Sensors

Task: Find the comparison object with the same hardness as the test object in **as few touches as possible**



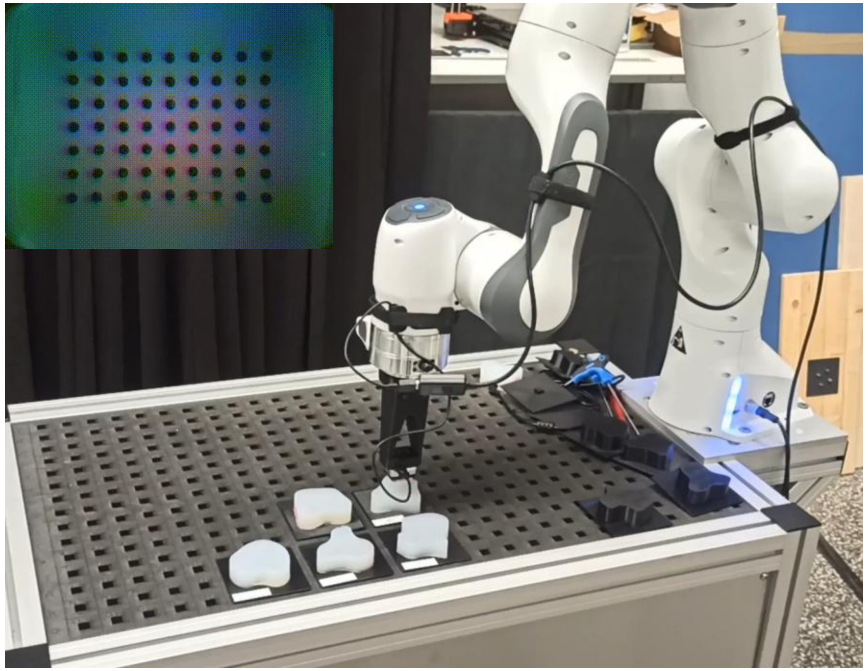
Comparison objects

Test Object

Chen et al. "Investigating Active Sampling for Hardness Classification with Vision-Based Tactile Sensors", arXiv (2025)

Active Sampling for Hardness Recognition With Vision-Based Tactile Sensors

Perform **multiple rounds of exploration** and **continual model fine-tuning**



Re-sampling Strategies: Touch the next object based on the **model uncertainty**

1. Variance Strategy

2. Entropy Strategy

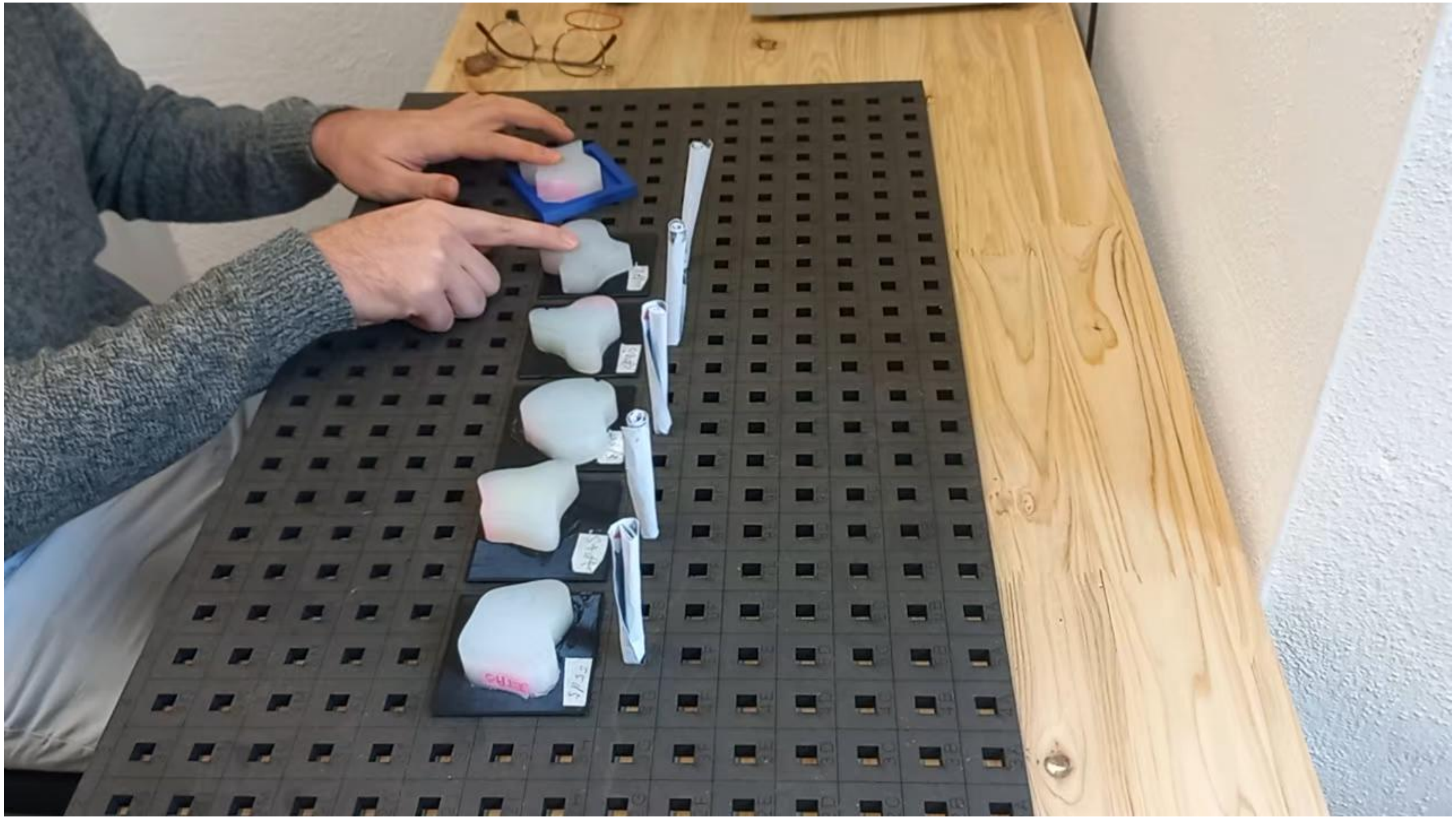
3. Random Strategy

4. No-resampling Strategy

Chen et al. "Investigating Active Sampling for Hardness Classification with Vision-Based Tactile Sensors", arXiv (2025)

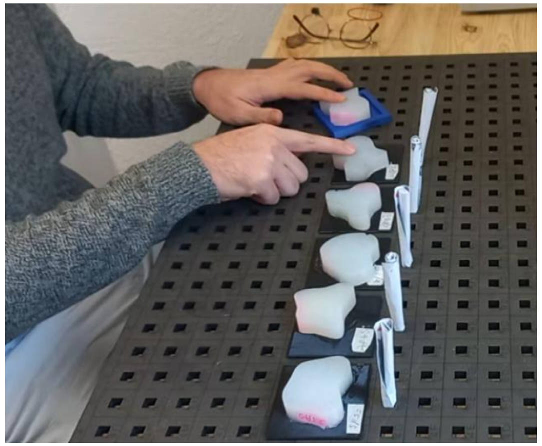
Human Study – Hardness Recognition

Blindfolded participants (n=10) did the hardness recognition task

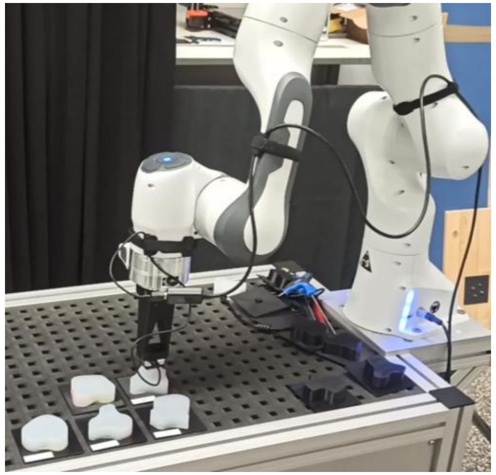


Chen et al. "Investigating Active Sampling for Hardness Classification with Vision-Based Tactile Sensors", arXiv (2025)

Hardness Recognition Accuracy



Humans	No Resampling	Variance	Entropy	Random
48.00% ±22.27%	57.20% ±37.25%	**88.78%** ±26.85%	85.79% ±25.85%	83.26% ±30.84%

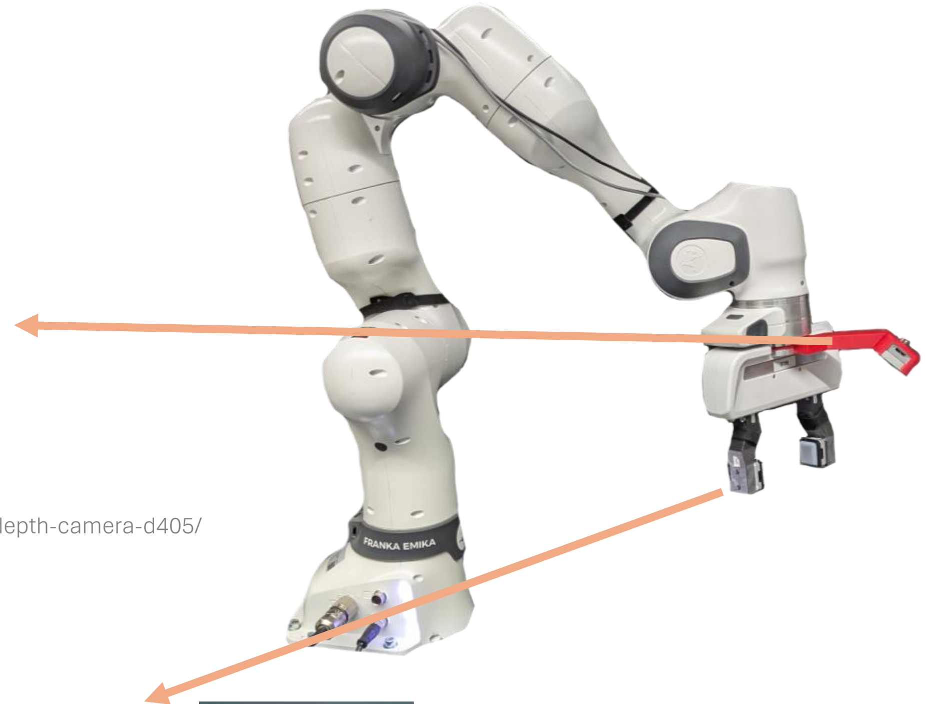


Ablation : Dropout rate and classifier architecture matter more than active sampling strategy

Chen et al. “Investigating Active Sampling for Hardness Classification with Vision-Based Tactile Sensors”, arXiv (2025)

Visuo-tactile In-hand Pose Estimation

https://www.intelrealsense.com/depth-camera-d405/

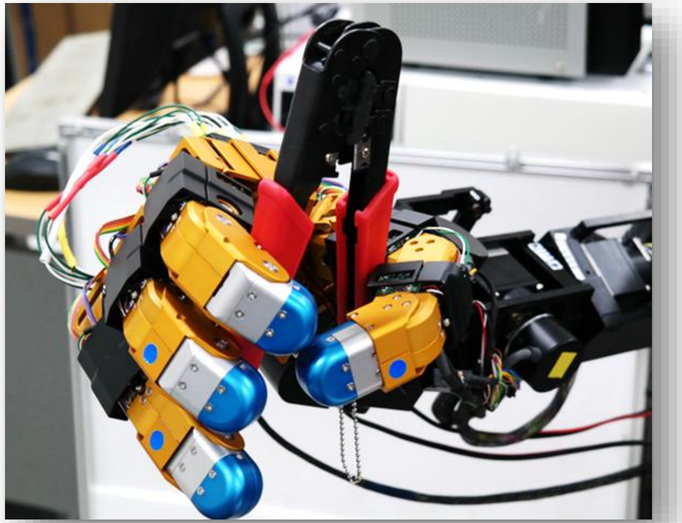


https://www.gelsight.com/gelsightmini/

https://www.motoman.com/en-us/flexible,-high-speed-robotic-parcel-induction

Applications

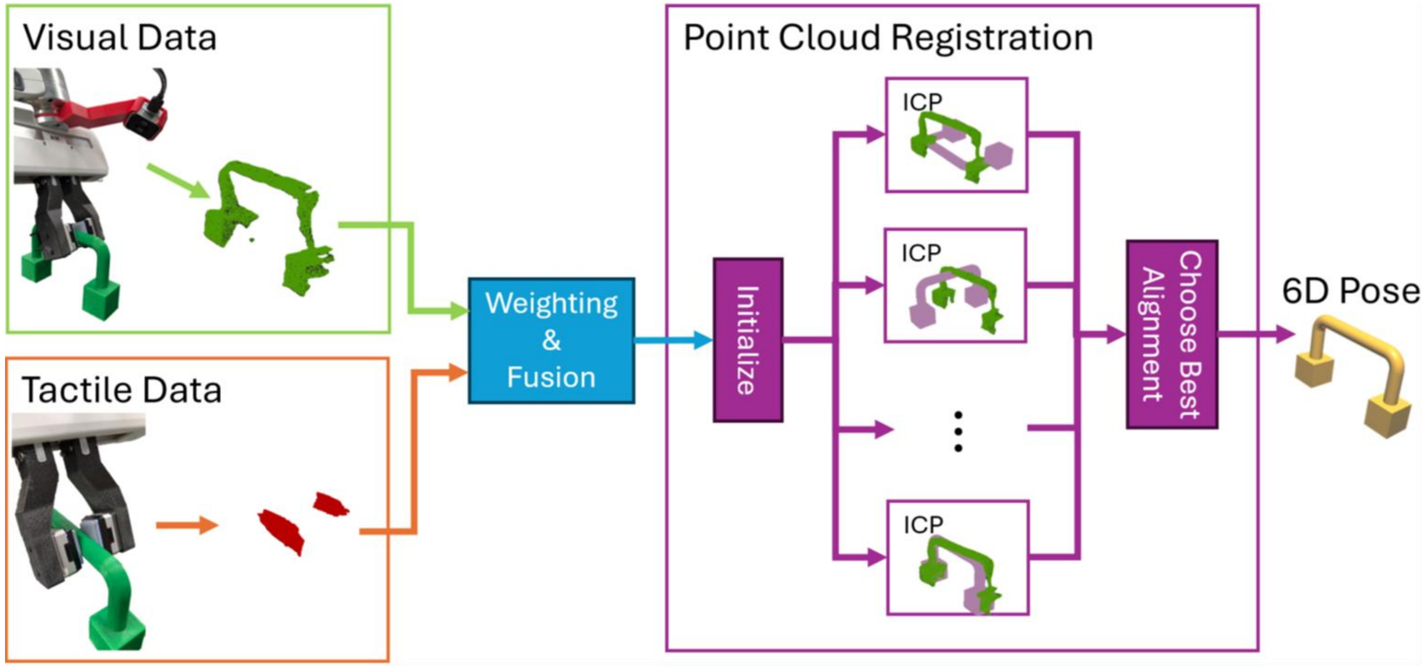
- Pick and place
- Assembly
- Human-robot collaboration

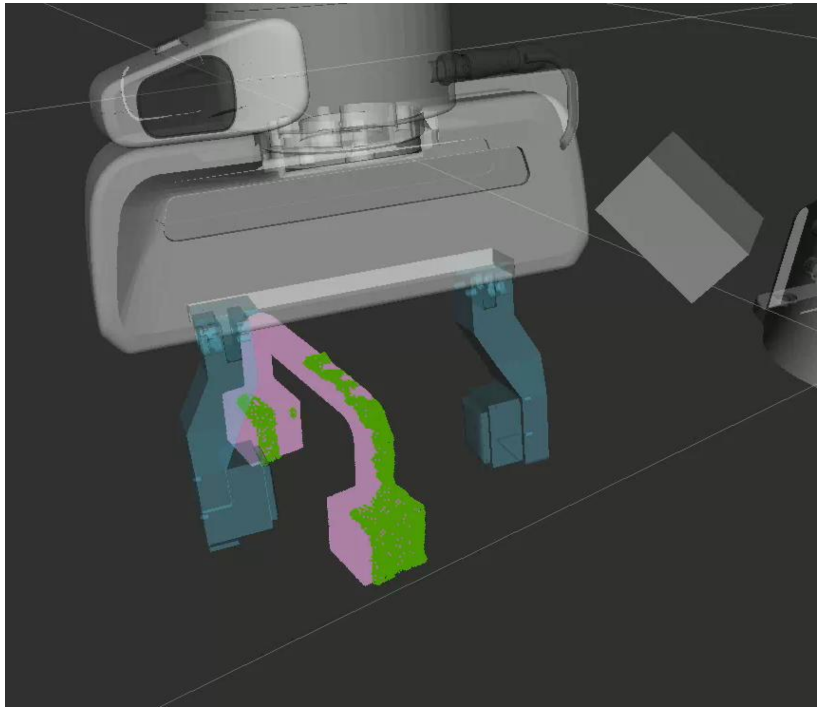


https://www.ieee-ras.org/robotic-hands-grasping-and-manipulation

Nonnengießer et al. “Visuotactile In-Hand Pose Estimation”, arXiv (2025)

Visuo-tactile In-hand Pose Estimation

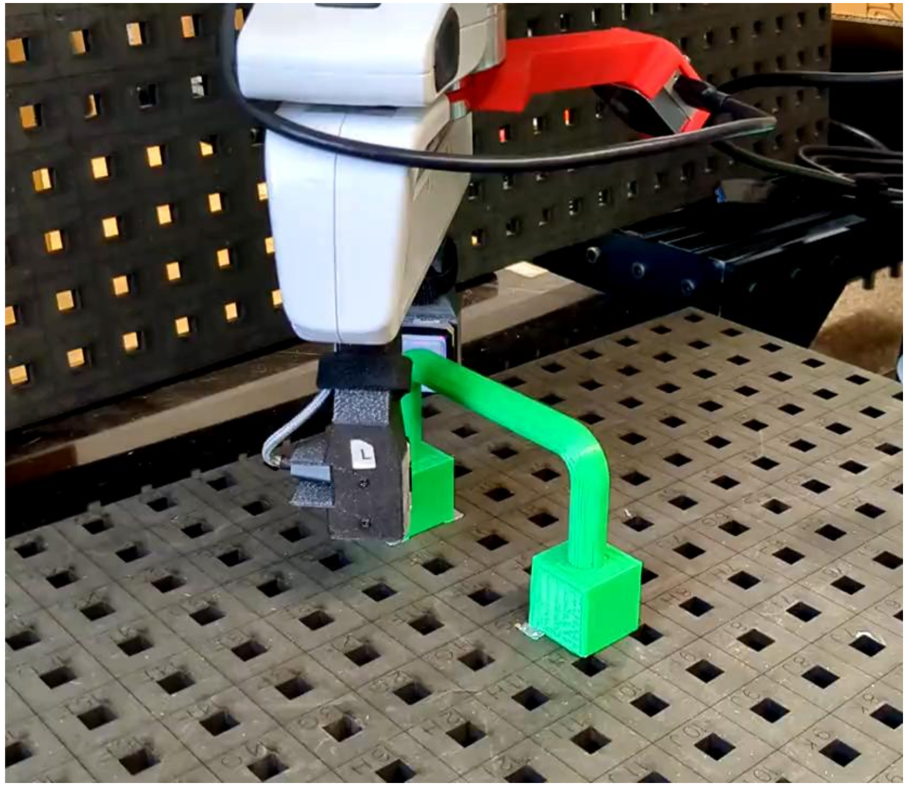


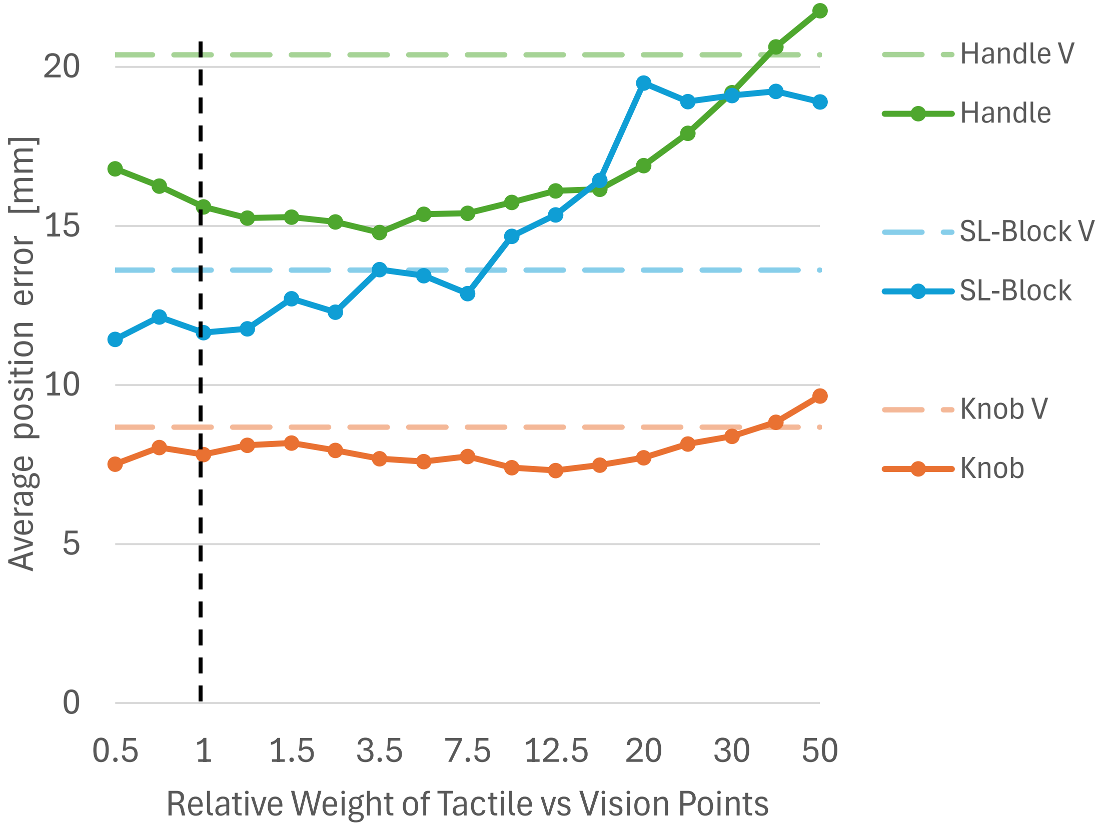


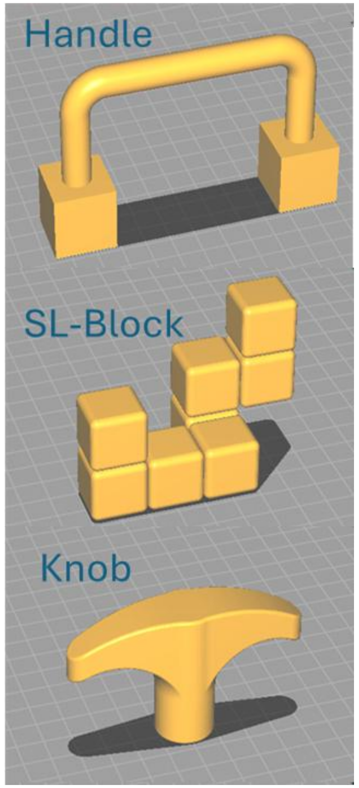
Nonnengießer et al. "Visuotactile In-Hand Pose Estimation", arXiv (2025)

Visuo-tactile In-hand Pose Estimation

Effect of Tactile Modality on Pose Estimation Error



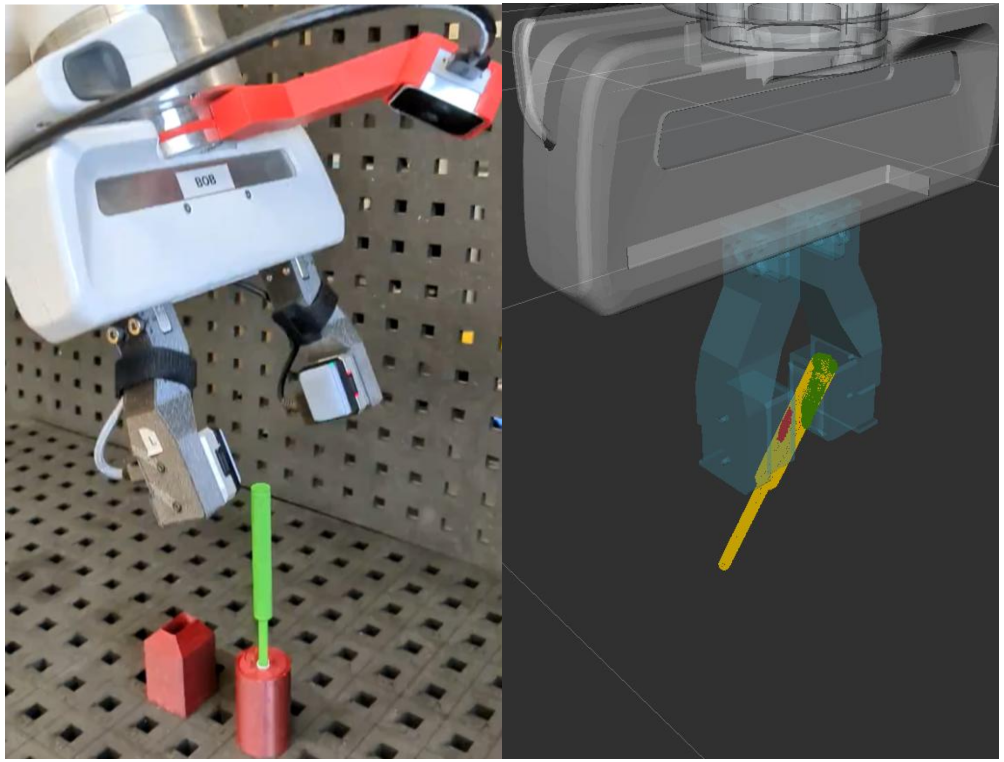




50 poses per object

Nonnengießer et al. “Visuotactile In-Hand Pose Estimation”, arXiv (2025)

Visuo-tactile In-hand Pose Estimation



Insertion Task

Camera z-Rotation: 0°

Angle	Vision	Vision + Touch	Occlusion
0	✓ ✓	✓ ✓	80%
-5	✓ ✓	✓ ✓	82%
-10	✓ ✓	✓ ✓	85%
-15	✓ ✓	○ ✓	86%
-20	○ ○	✓ ✓	86%
-25	○ ✓	✓ ✓	88%
-30	○ ○	✓ ○	89%
-35	○ ✕	✓ ○	93%
-40	✕ ✕	✕ ✓	98%
-45	○ ✕	✓ ✓	98%
-50	✕ ✕	✓ ✓	98%

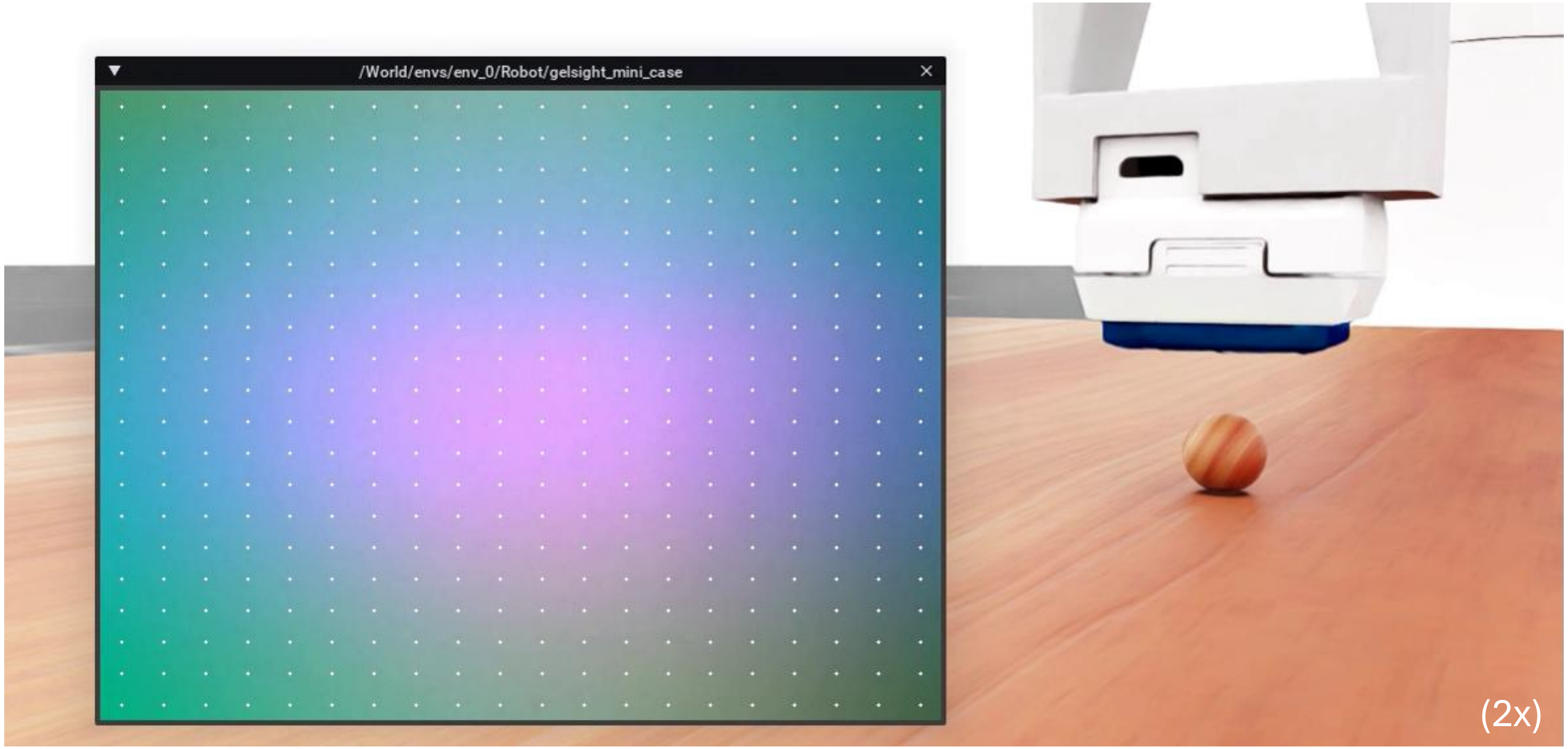
Camera z-Rotation: 45°

Angle	Vision	Vision + Touch	Occlusion
0	✓ ✓	✓ ✓	96%
-5	✓ ✓	✓ ✓	95%
-10	✓ ✕	✓ ✓	94%
-15	✕ ✕	✕ ✓	94%
-20	✕ ✕	✓ ✓	94%
-25	✕ ✕	✓ ✓	94%
-30	✕ ✕	✓ ✓	94%
-35	✕ ✕	✓ ✓	94%
-40	✕ ✕	✓ ✓	94%
-45	✕ ✕	✓ ✓	94%
-50	✕ ✕	○ ✓	94%

✓ **Perfect insertion** (tool does not touch hole) | ○ **Successful insertion** (tool touches inner sides of hole) | ✕ **Failed insertion** (tool misses hole)

Nonnengießer et al. "Visuotactile In-Hand Pose Estimation", arXiv (2025)

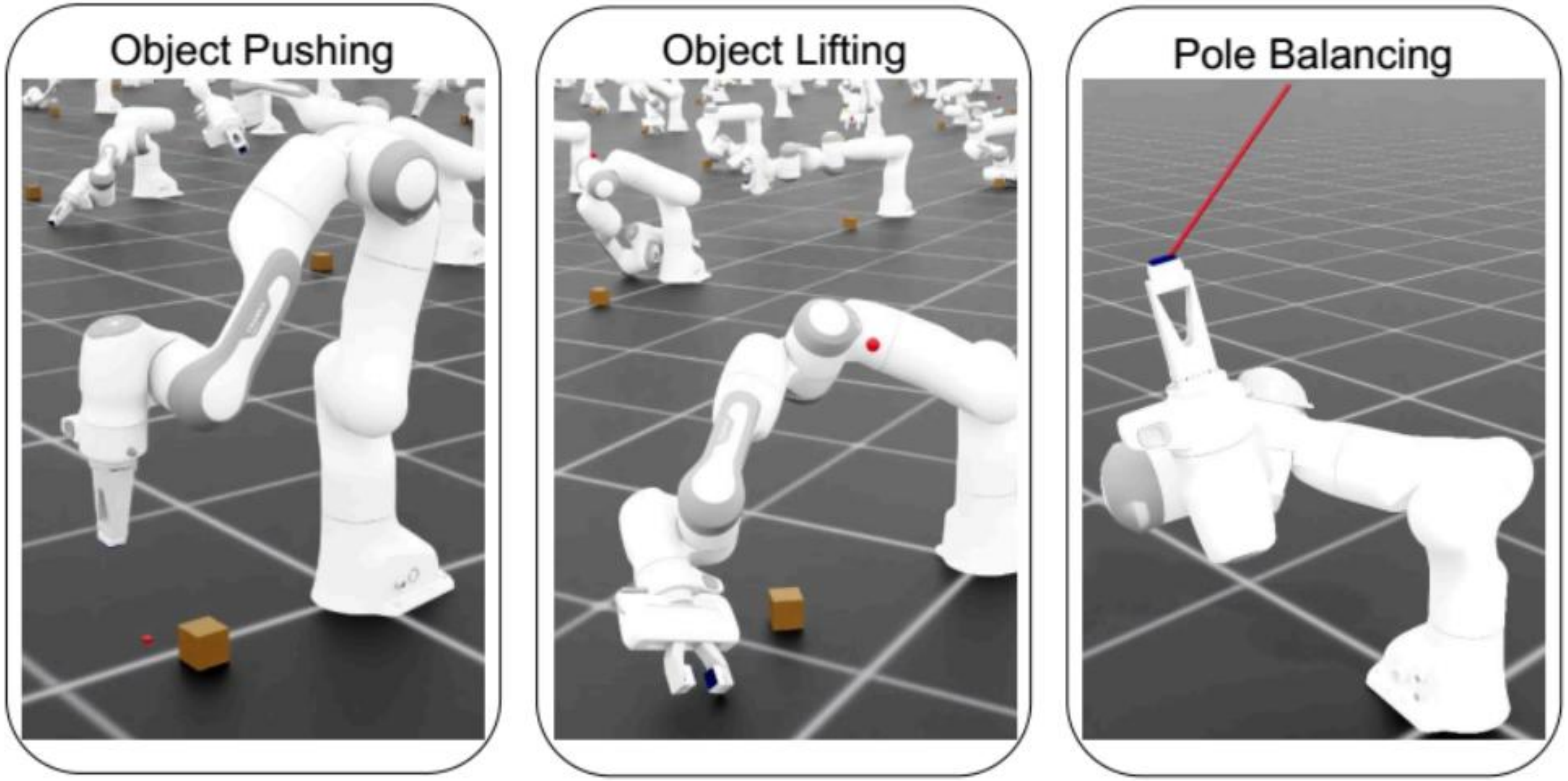
Ongoing Work: Simulation of Vision-based Tactile Sensors



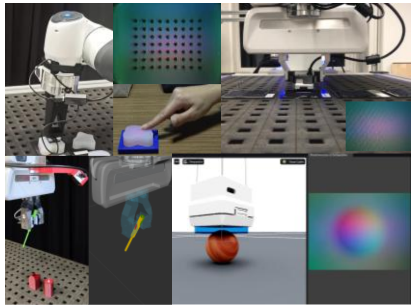
Deformable Body Simulation (IPC) + Optical Simulation (Taxim) + Robot Simulation (Isaac Sim)

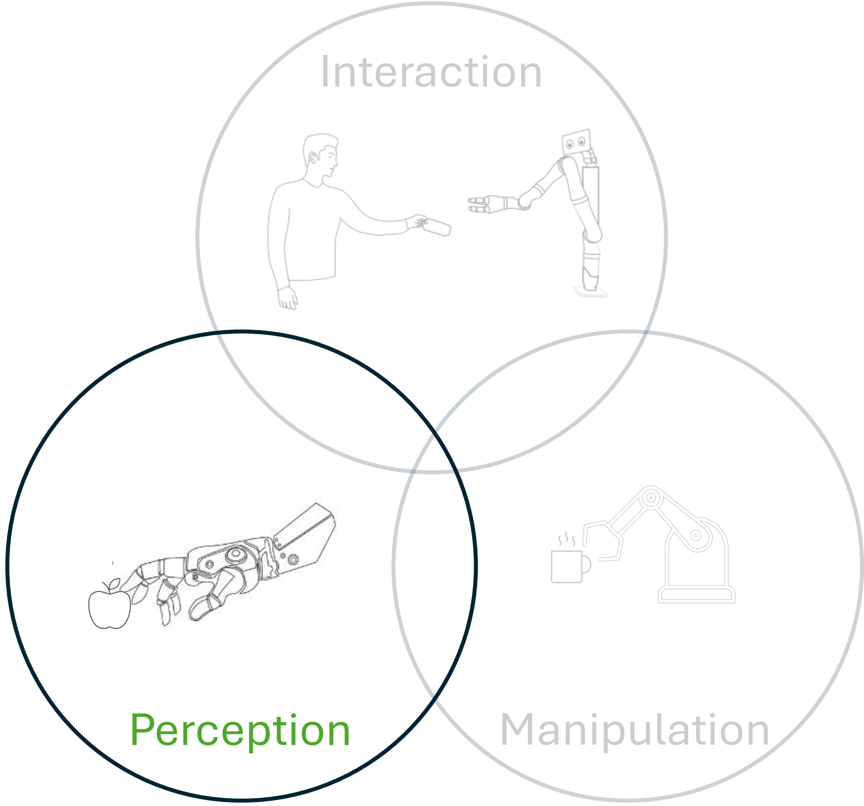
Nguyen et al. “TacEx: GelSight Tactile Simulation in Isaac Sim – Combining Soft-Body and Visuotactile Simulators”, GRC (2025)

Ongoing Work: Simulation of Vision-based Tactile Sensors



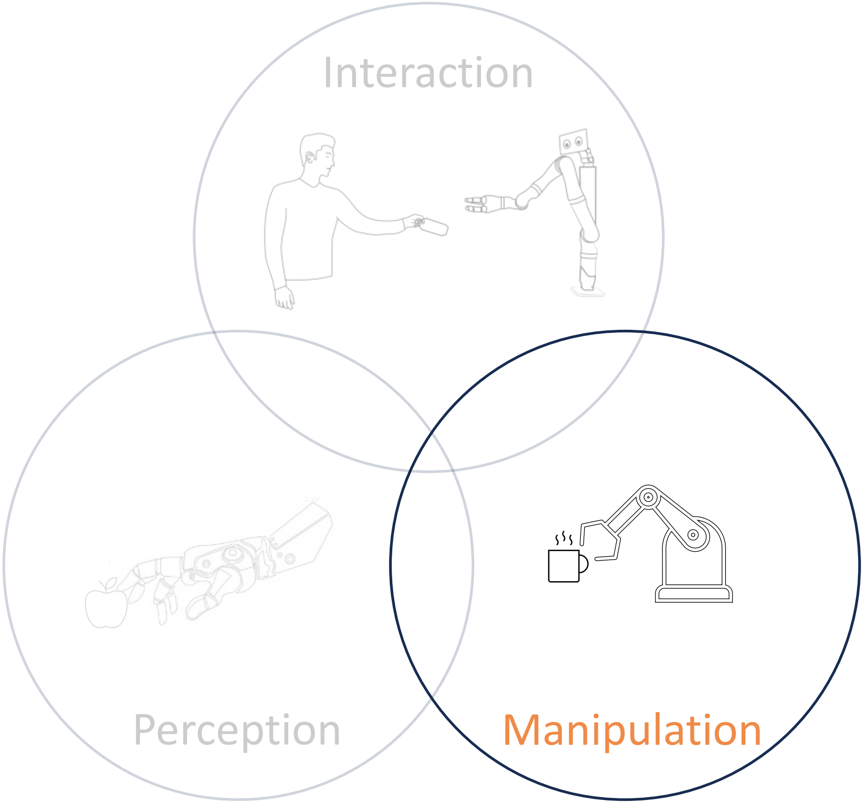
Part 2 : Perception





Perceiving Object Properties with Vision-Based Tactile Sensors

Part 3 : Manipulation



Learning Dynamic Manipulations with Reinforcement Learning

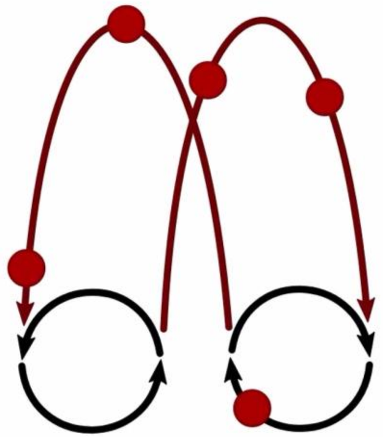
Toss-Juggling

Robotic Toss Juggling

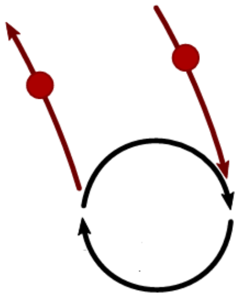
Push the limits of dynamic robotic manipulation

Investigate human adaptation in dyadic human-robot juggling

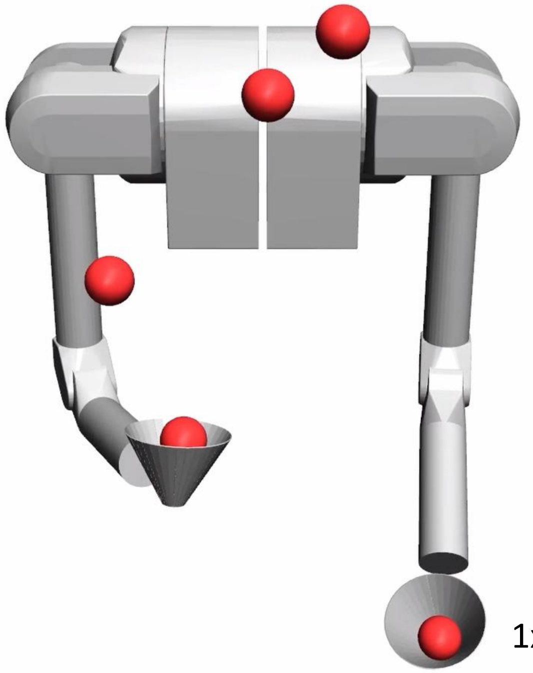
Previous Work: Kinematic Planning for Robotic Toss-Juggling



5-ball Cascade



Hand-cycle



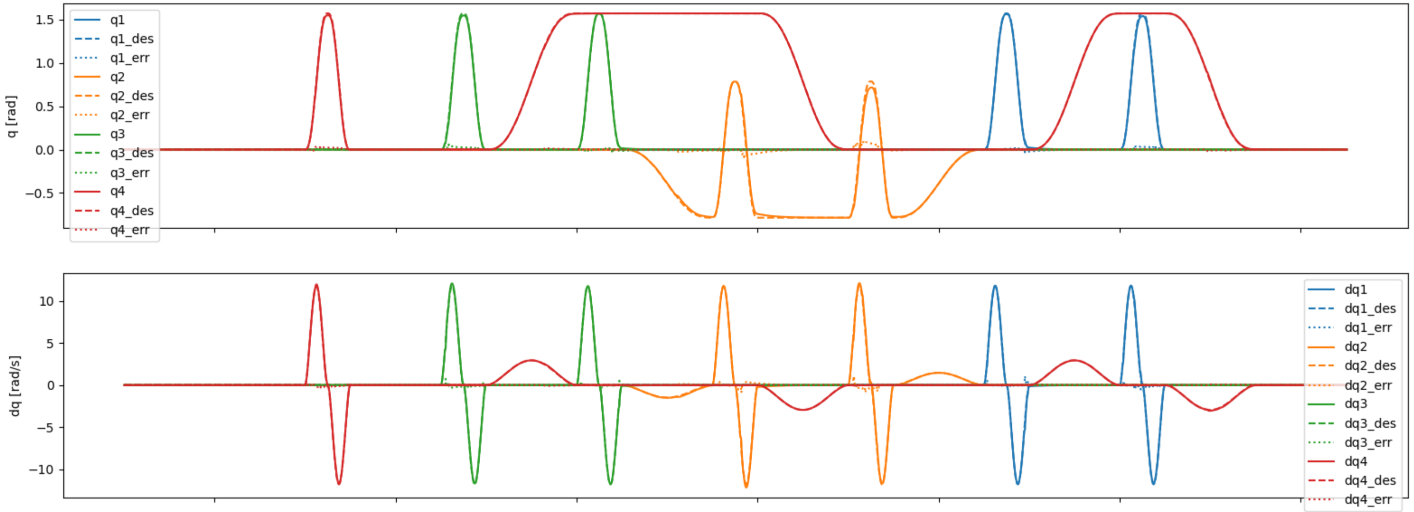
$$
\begin{aligned}
&\min_{\{\ddot{\mathbf{q}}_0,\dots,\ddot{\mathbf{q}}_{K_c-1}\}} \sum_{k=0}^{K_c} \ddot{\mathbf{q}}_k^T \ddot{\mathbf{q}}_k \text{ s.t.} \\
&(\mathbf{q}_0, \dot{\mathbf{q}}_0, \ddot{\mathbf{q}}_0) = (\mathbf{q}_{\mathrm{TO}-1}, \dot{\mathbf{q}}_{\mathrm{TO}-1}, \ddot{\mathbf{q}}_{\mathrm{TO}-1}) \\
&\mathbf{h}(\mathbf{q}_k, \dot{\mathbf{q}}_k, \ddot{\mathbf{q}}_k, \mathbf{q}_k) = 0 \\
&-\ddot{q}_{i,\max} \leq \ddot{q}_{k,i} \leq \ddot{q}_{i,\max}
\end{aligned}
$$

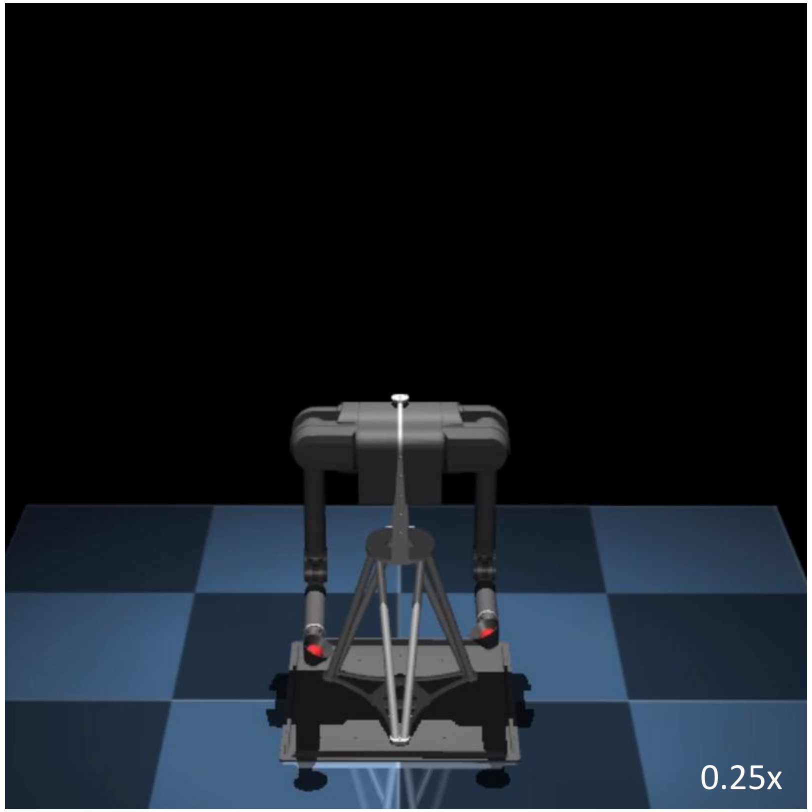
Kinematic Planning based on take-off and touch-down constraints

Kai Ploeger and Jan Peters, “Controlling the Cascade: Kinematic Planning for N-ball Toss Juggling”, IROS (2022)

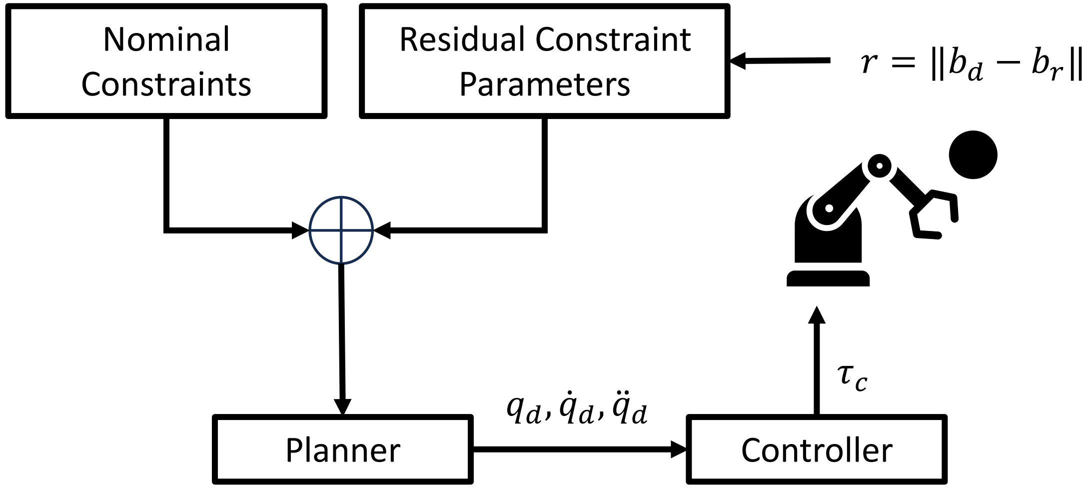
Previous Work: Kinematic Planning for Robotic Toss-Juggling

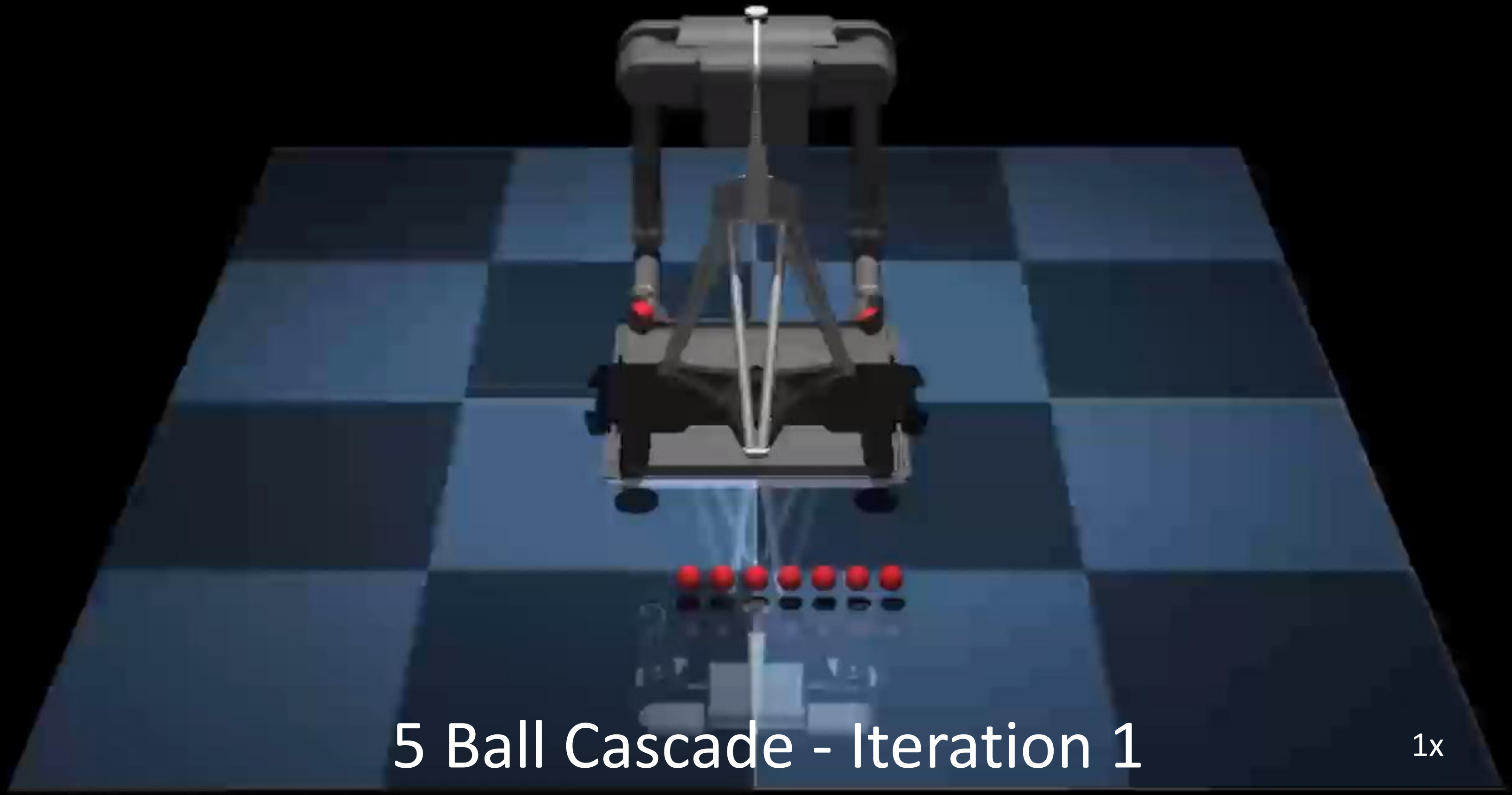
Problem: Inaccurate Tracking Controllers + Contact Dynamics



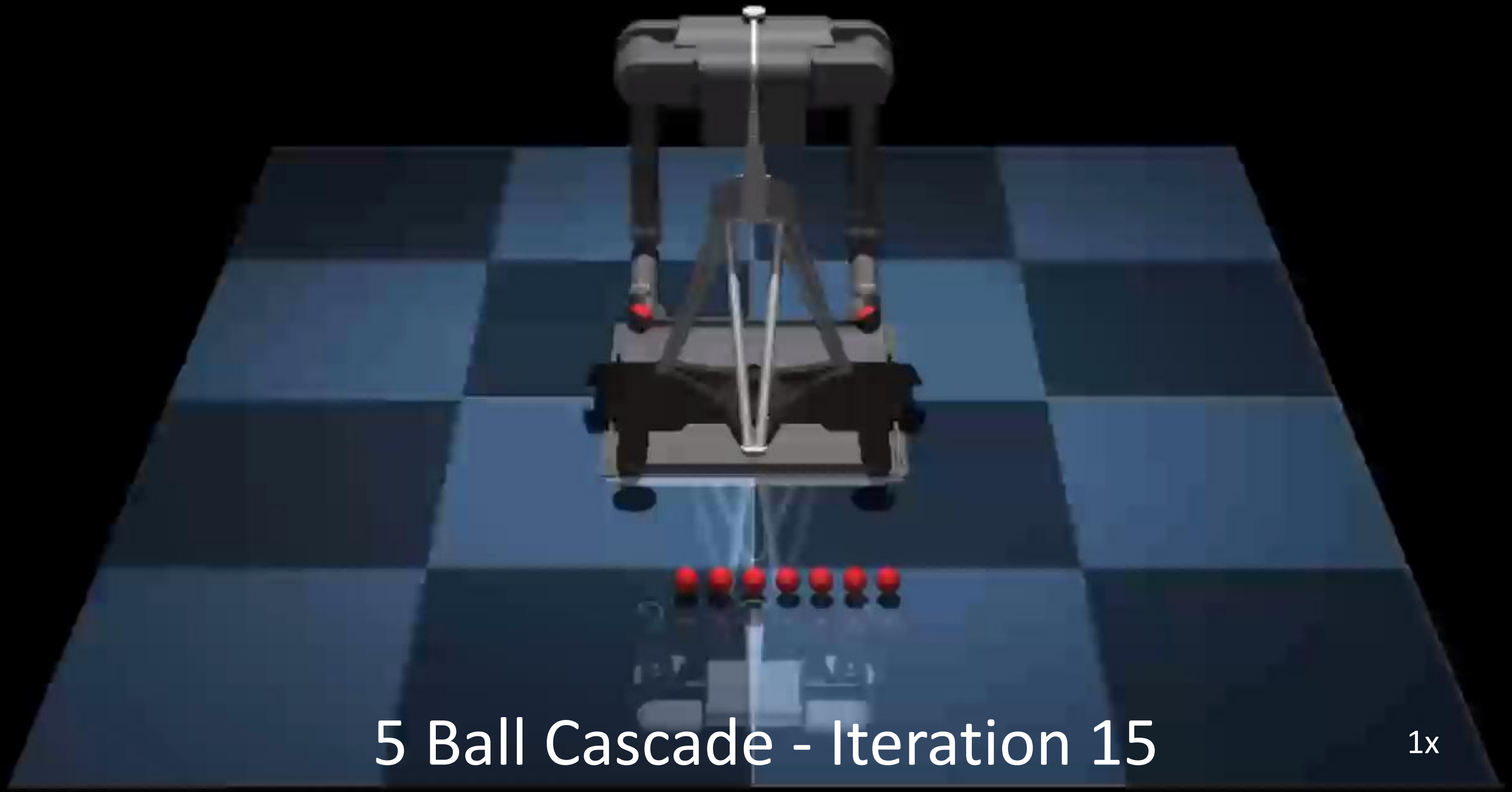


Ongoing Work: Learning Robotic Toss-Juggling with Residual Reinforcement Learning

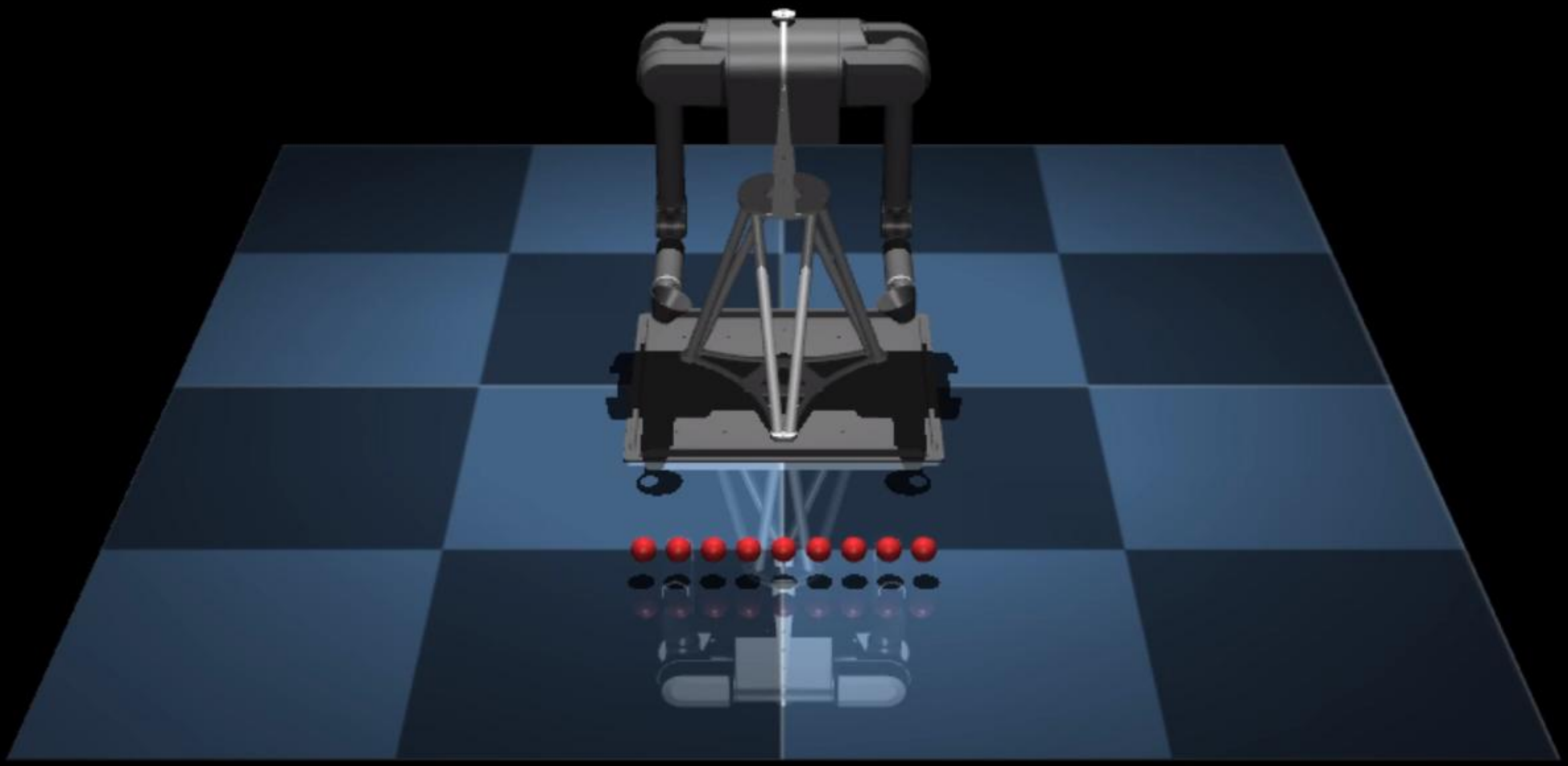




5 Ball Cascade - Iteration 1

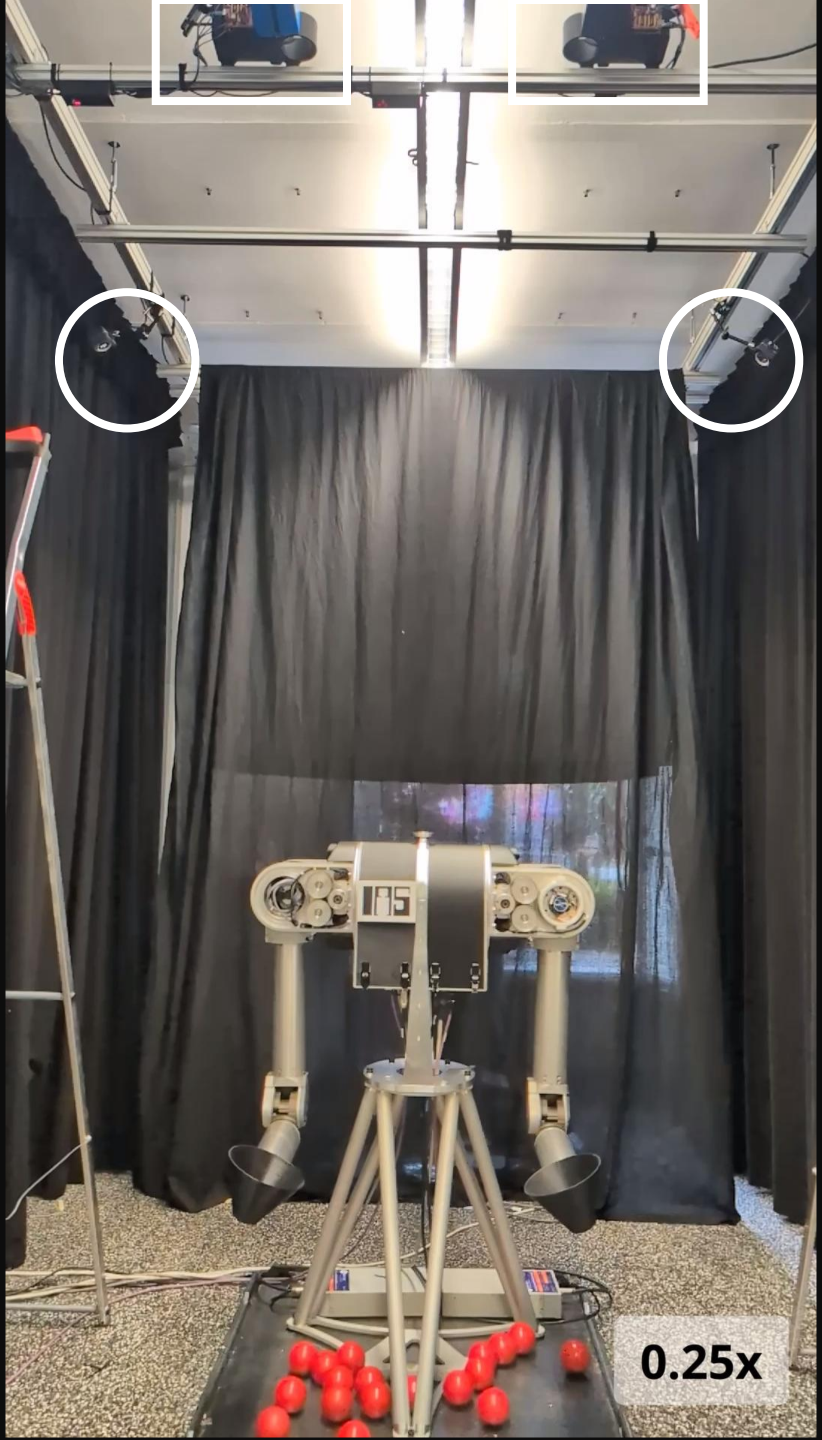


5 Ball Cascade - Iteration 15



7 Ball Cascade - Iteration 25

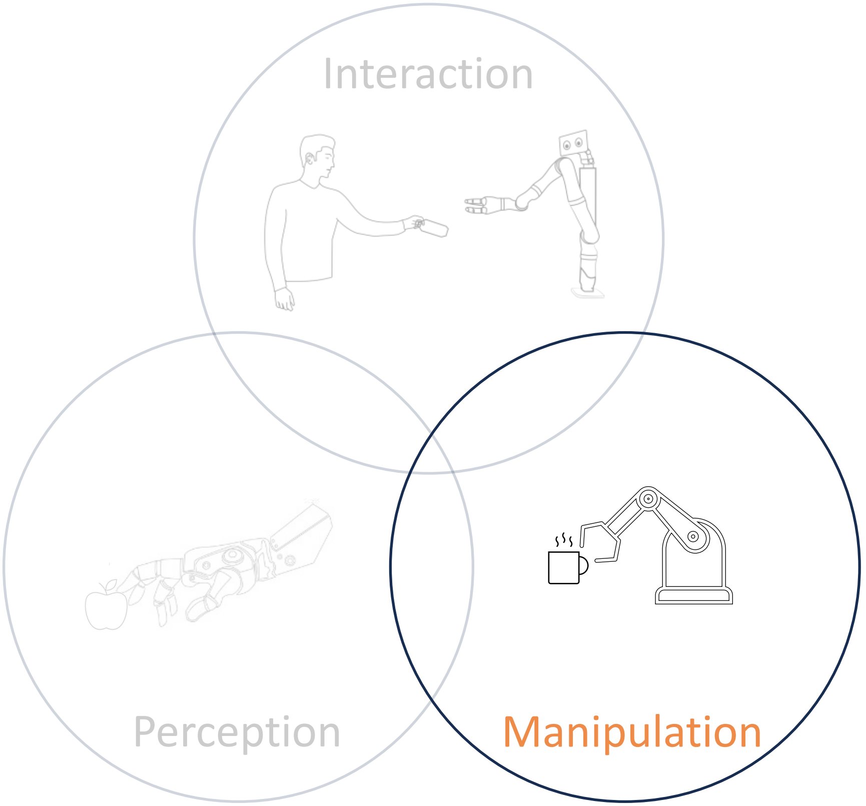
1x



Ball Launchers

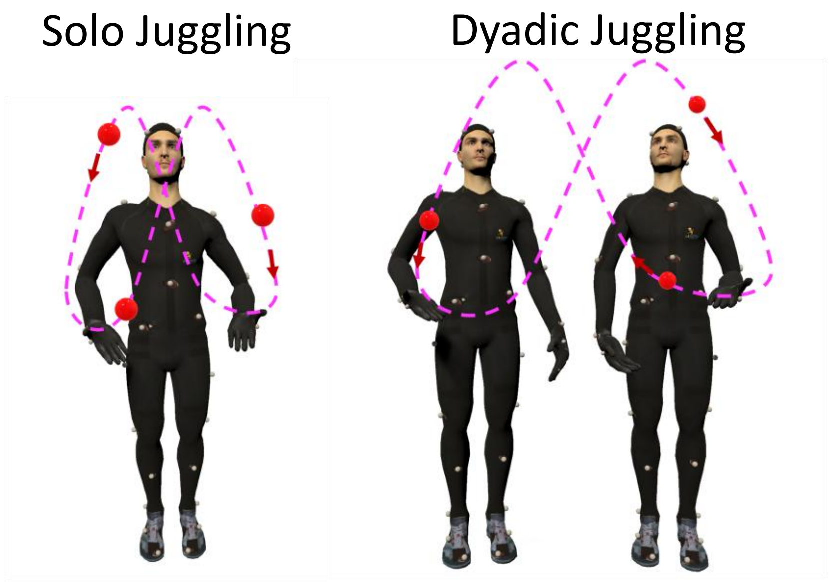
OptiTrack

Part 3 : Manipulation



Learning Dynamic Manipulations with Reinforcement Learning

Hand Movements in Juggling



- Jugglers executed goal-directed movements while catching

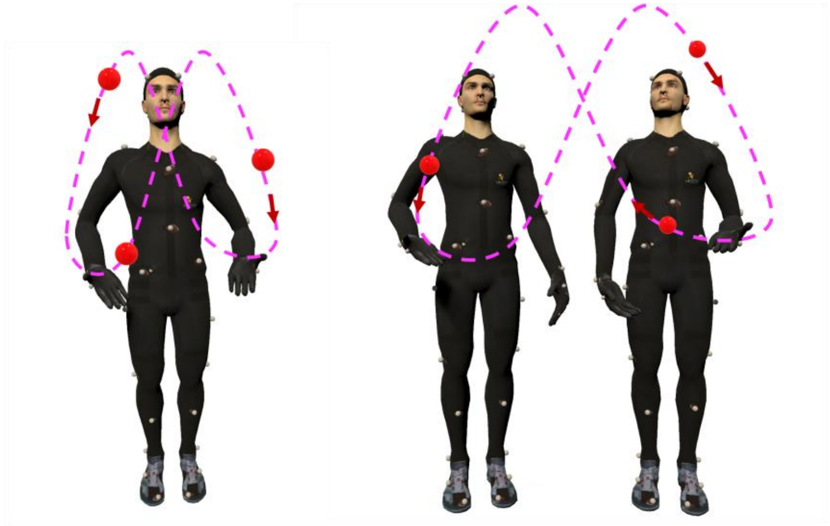
- Onset of goal directed movement delayed in dyadic juggling

Chowdhury et al. “The role of trajectory prediction in juggling”, NCM, 2025

Hand Movements in Juggling

Solo Juggling

Dyadic Juggling



- Jugglers executed goal-directed movements while catching
- Onset of goal directed movement delayed in dyadic juggling

Chowdhury et al. “The role of trajectory prediction in juggling”, NCM, 2025

Finger Movements in Tactile Exploration

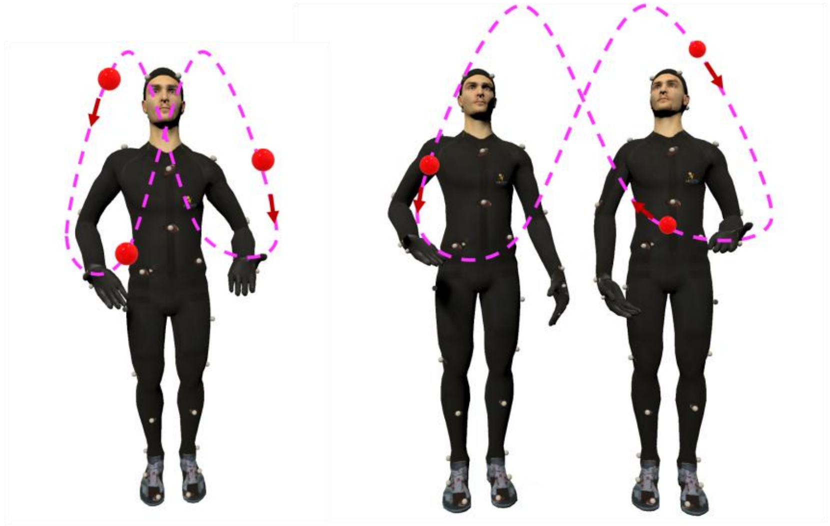
- Nine exploratory procedures (EPs) observed during shape and deformability judgments
- EPs varied as a function of material/object properties unrelated to the primary task

Lin et al. “Task-Adapted Single-Finger Explorations of Complex Objects”, Eurohaptics, 2024

Hand Movements in Juggling	Finger Movements in Tactile Exploration	Postural Responses to Virtual Heights

Hand Movements in Juggling

Solo Juggling   Dyadic Juggling



- Jugglers executed goal-directed movements while catching
- Onset of goal directed movement delayed in dyadic juggling

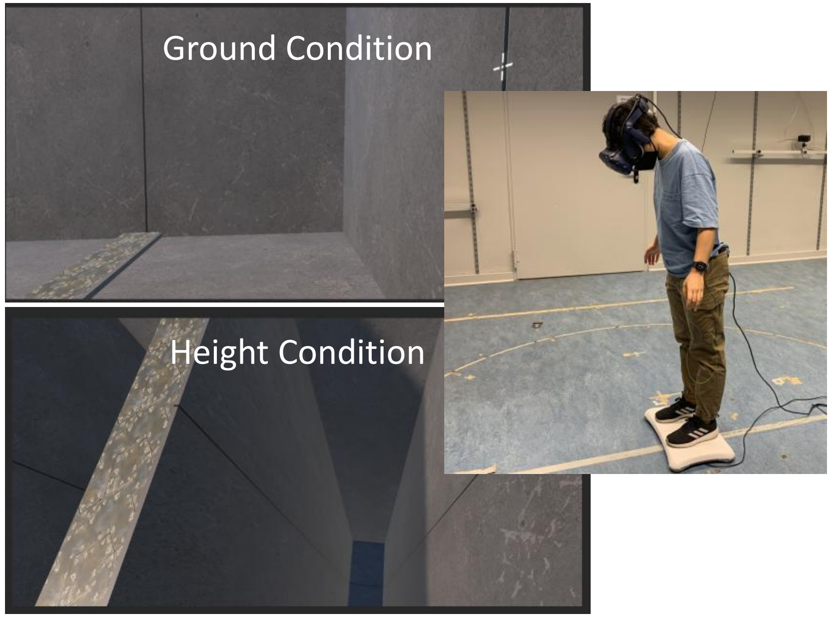
Chowdhury et al. “The role of trajectory prediction in juggling”, NCM, 2025

Finger Movements in Tactile Exploration

- Nine exploratory procedures (EPs) observed during shape and deformability judgments
- EPs varied as a function of material/object properties unrelated to the primary task

Lin et al. “Task-Adapted Single-Finger Explorations of Complex Objects”, Eurohaptics, 2024

Postural Responses to Virtual Heights

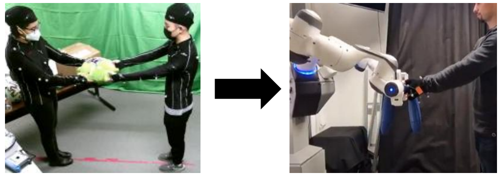


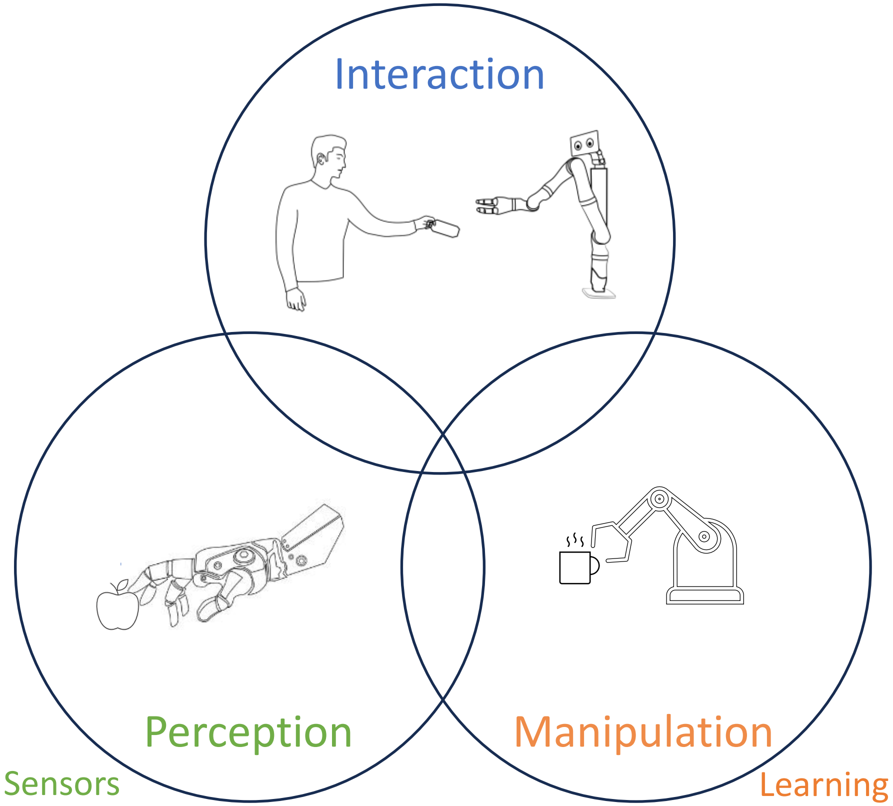
- Height exposure increases postural sway frequency and reduces amplitude
- Arm joints show strongest reduction in sway range

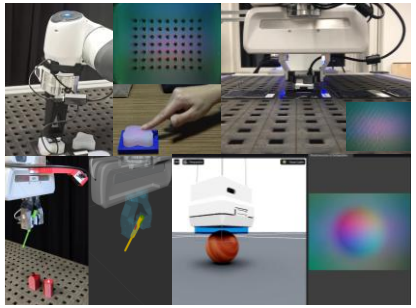
Koosha et al. “Staring Down the Elevator Shaft”, CogSci, 2025

Thank You

Learning Human-Robot Interaction from Human Demonstrations







Perceiving Object Properties with Vision-Based Tactile Sensors

Learning Dynamic Manipulations with Reinforcement Learning

Collaborators

Abir Chowdhury, Alina Boehm, Antonio De Almeida-Correia, Boris Belousov, Constantin Rothkopf, Dorothea Koert, Duc Huy Nguyen, Fabian Hahne, Felix Nonnengießer, Frederik Heller, Georgia Chalvatzaki, Guy Hoffman, Jan Peters, Junyi Chen, Kai Ploeger, Katja Doerschner, Knut Drewing, Lisa Lin, Mario Gomez, Raphael Fortuna, Ruth Stock-Homburg, Shemar Christian, Tim Schneider, Vignesh Prasad, Yasemin Goeksu, Zhiming Xie