TAMS Introduction

04.2025

Shang-Ching Liu (Sam)

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics

Technical Aspects of Multimodal Systems

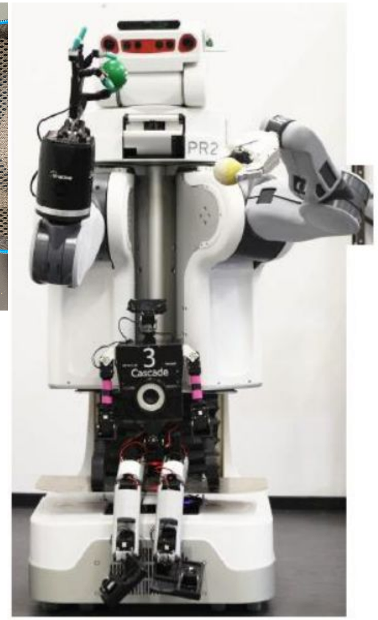
TAMS

- Technical Aspects of Multimodal Systems（Prof. Dr. Jianwei Zhang）
- Robot perception, machine learning, multimodal fusion

TAMS - University of Hamburg

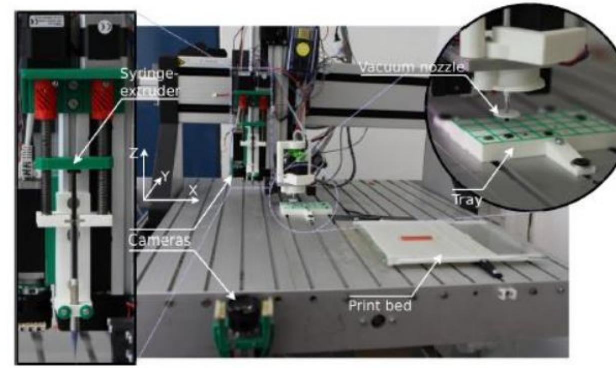
TAMS Hardware

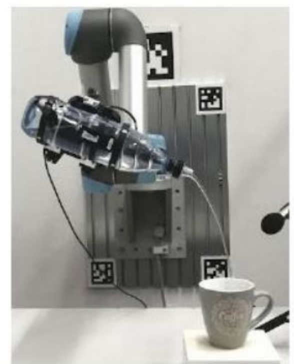












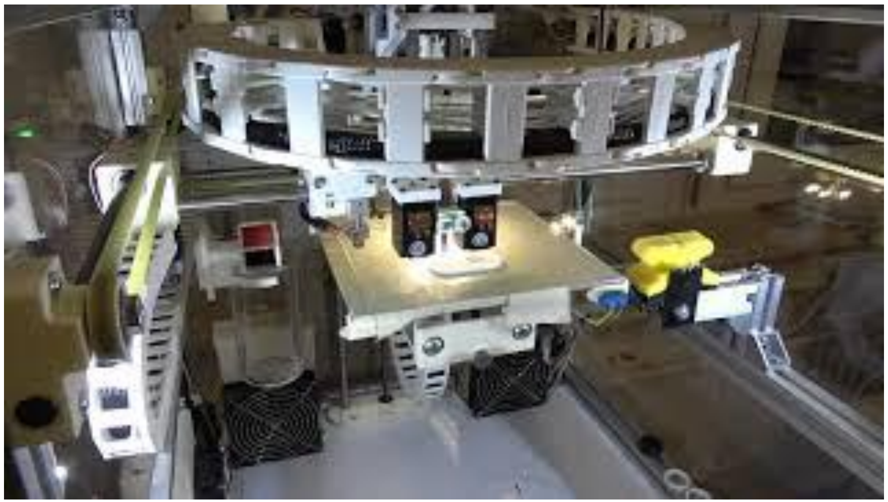


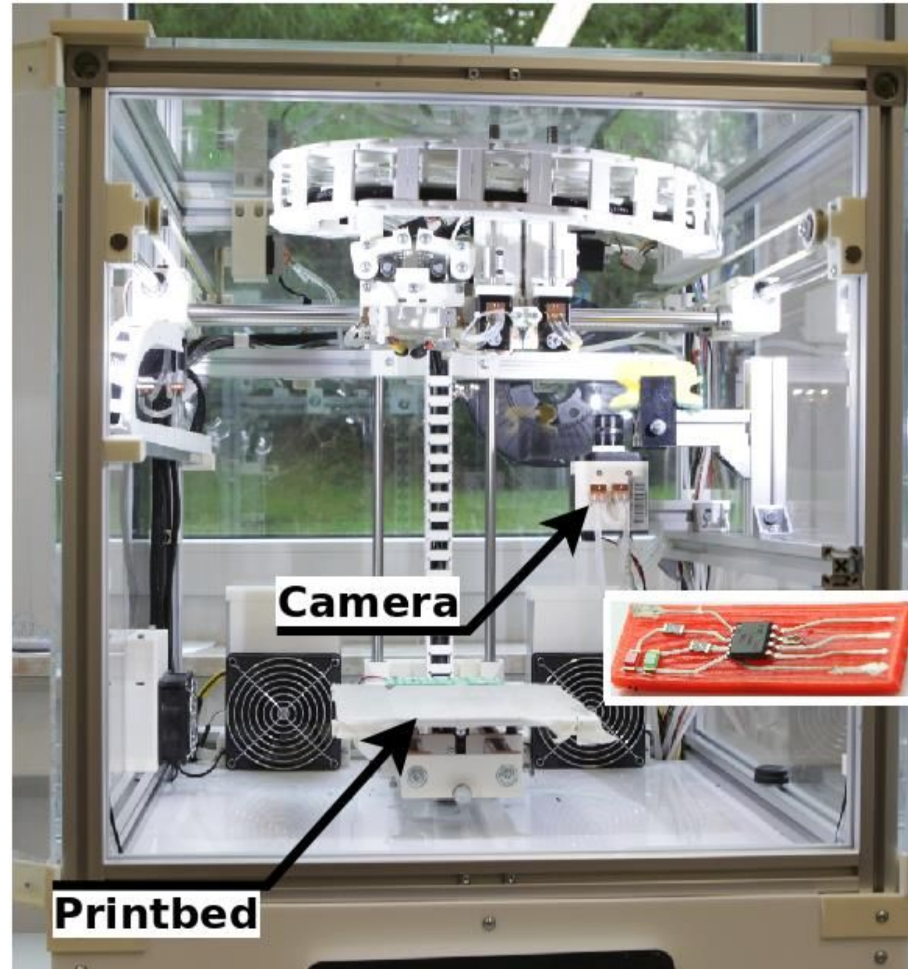
TAMS GPU

- TamsGPU2 NVIDIA RTX 2080*2 12G*2
- TamsGPU3 NVIDIA RTX 2080 Ti *2 12G*2
- TamsGPU4 NVIDIA RTX 3090*2 24G*2
- TamsGPU5 NVIDIA RTA A6000*2 48G*2
- TamsGPU6 NVIDIA H100*4 80G*2 and 94G*2
- TamsGPU7 NVIDIA RTX 4090*1 24G
- TamsGPU8 NVIDIA A6000*1(48G) RTX 4090*1 (24G)
- TamsGPU9 NVIDIA RTX 4090*1 24G

You should notice me or other group member before using it.

3D Printing related projects





Integration of Electronics into FDM Printed Objects


Bit-Bots
2023

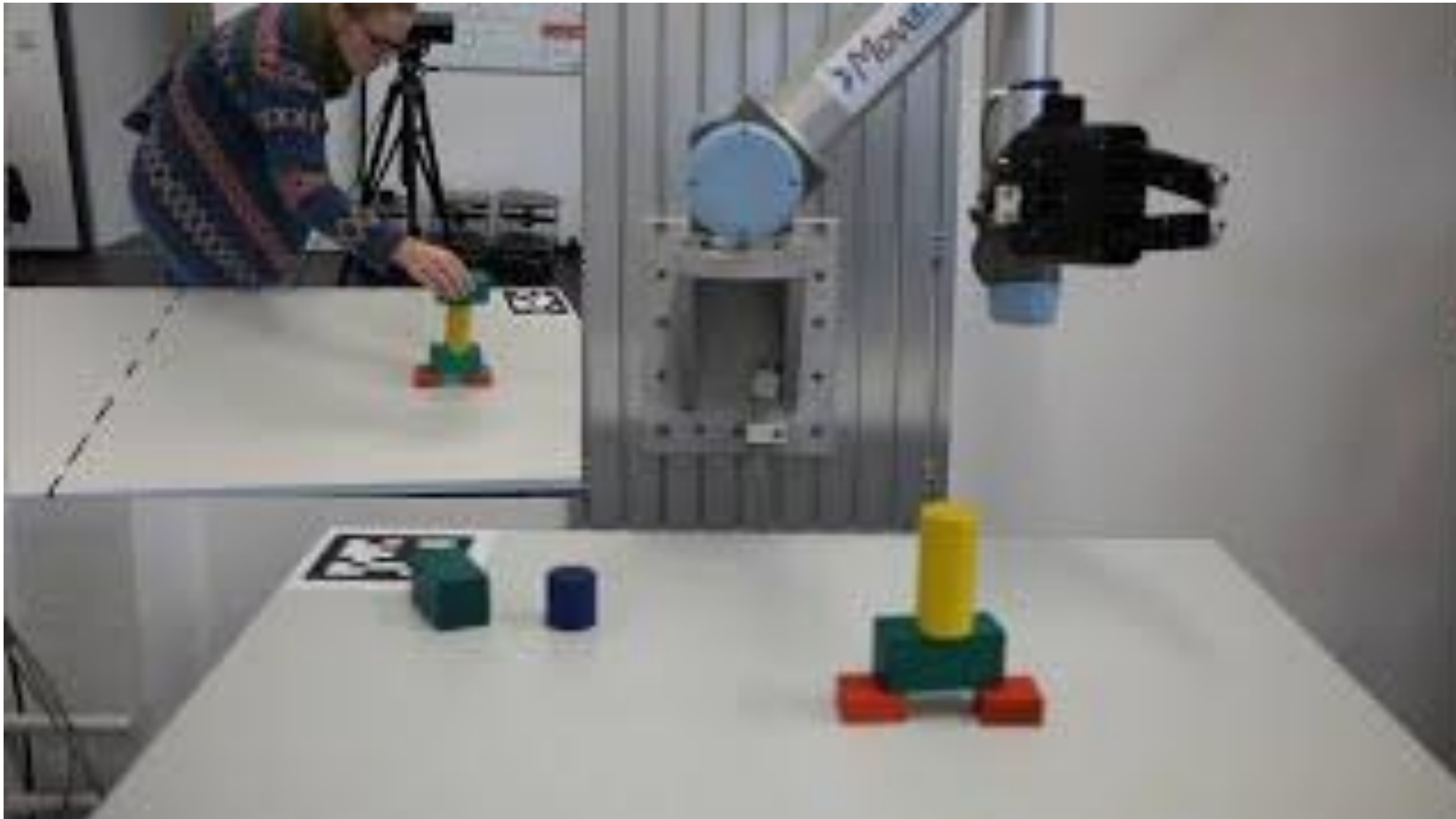
TAMS Application & Research


PR2

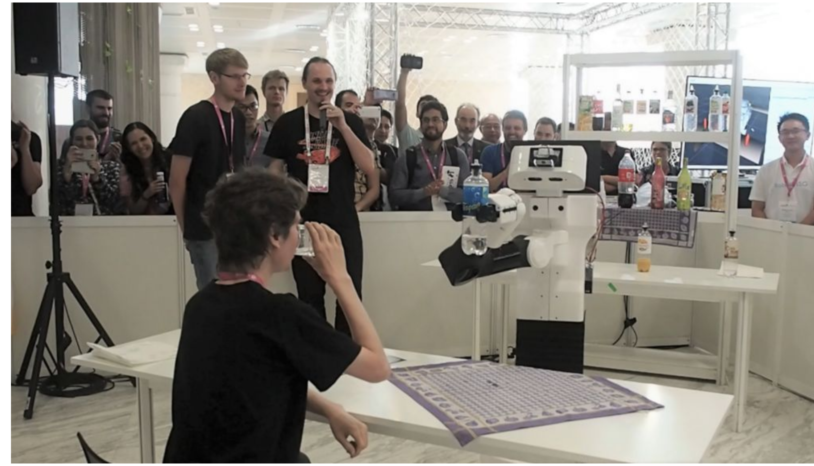
Luis The Bartender - Master Project 2016/17

PR2 Bartender - Master Project 2018/19

Tower Building - Master Project 2017/18



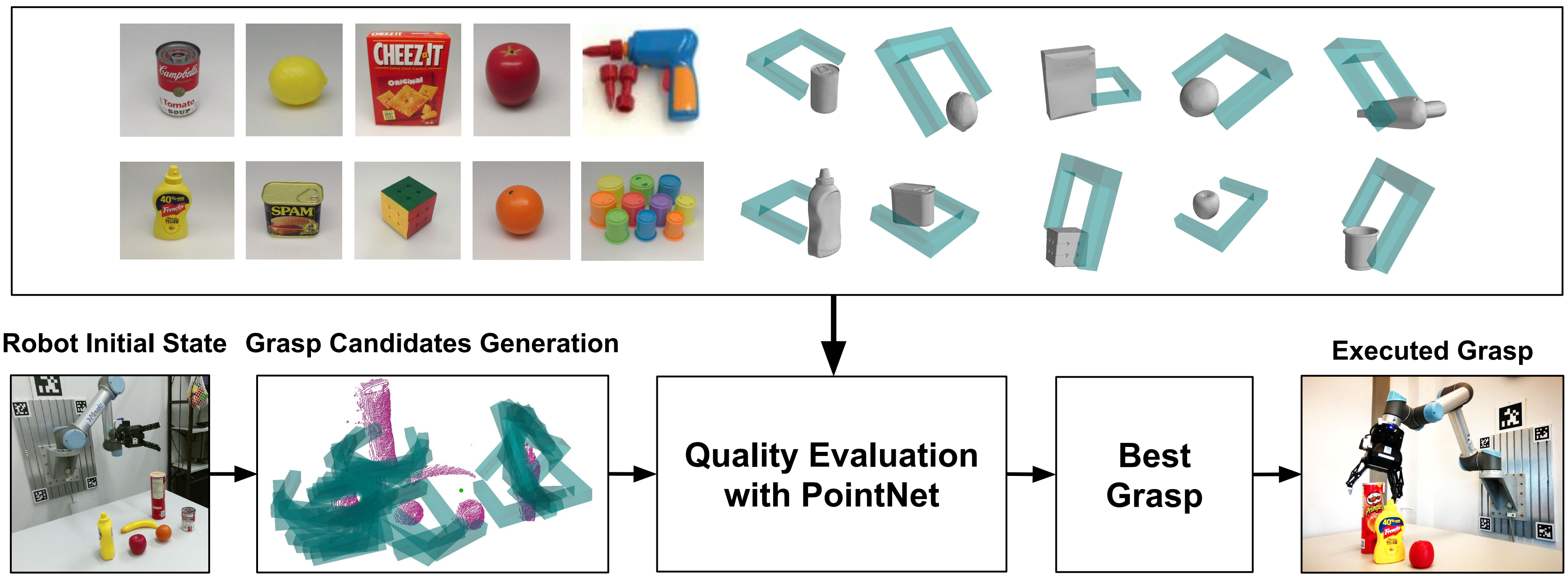
Mobile Manipulation Hackathon



Blackjack Dealer - Master Project 2019/20

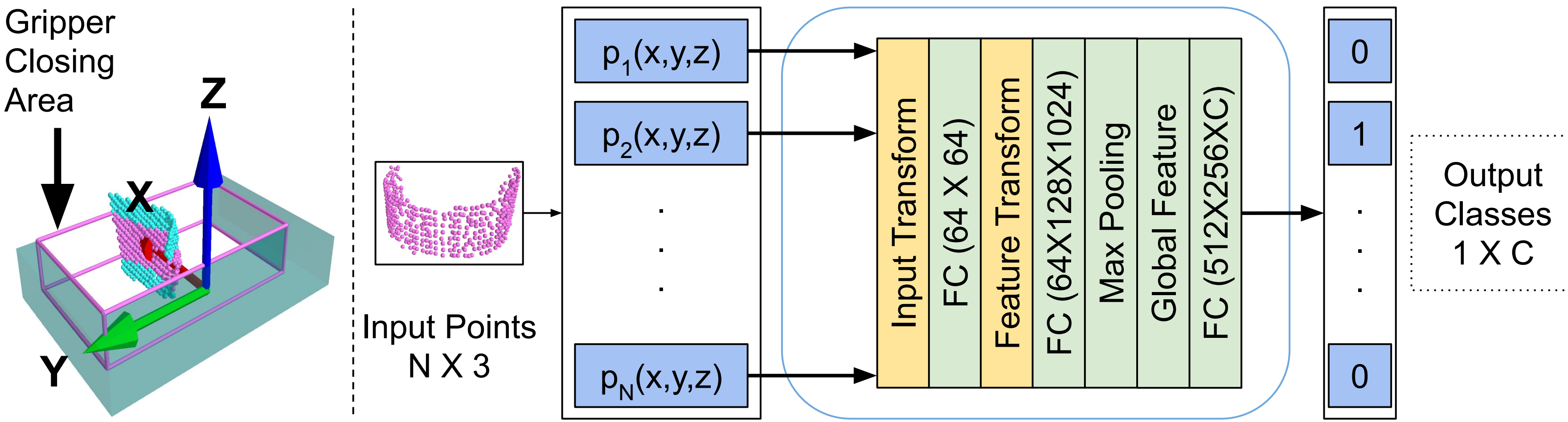
PointNetGPD

- How to improve grasp success rate?
  - Better dataset?
  - Better input modality? Better network?
  - More diverse grasps candidates?



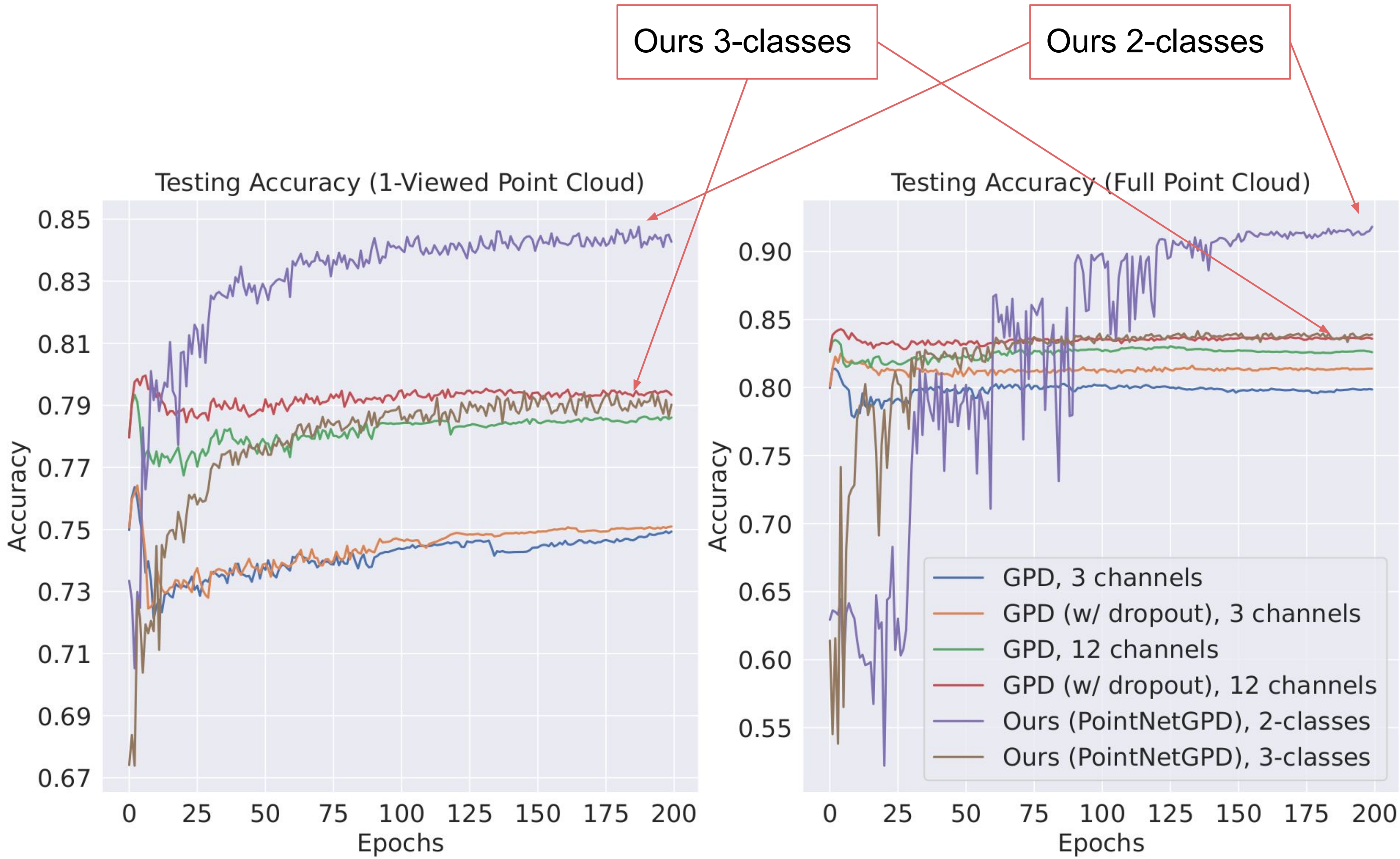
Liang et al., “PointNetGPD: Detecting Grasp Configurations from Point Sets”, ICRA 2019.

PointNetGPD

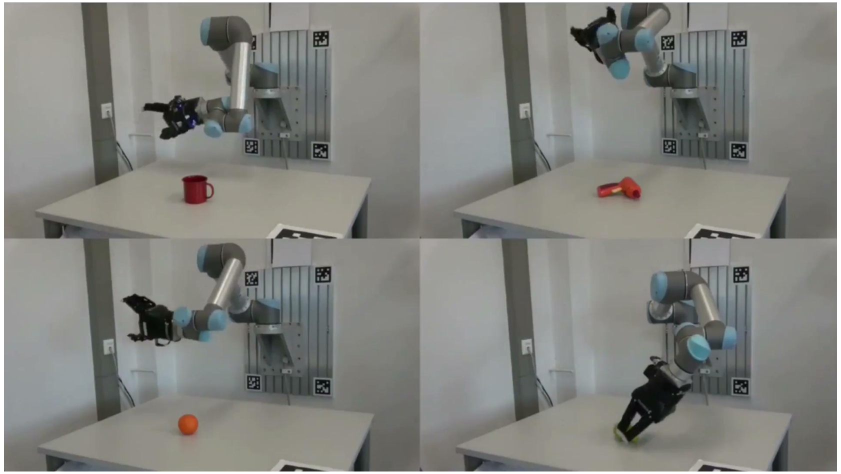


- Input: points within the closing area of the gripper
- Output: quality level of the input grasp
- The first work that uses PointNet for grasp evaluation

Network evaluation



Robot experiments

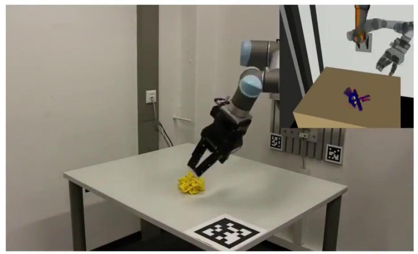


Method	Avg.	cleanser bottle	mug	meat can	tomatoo soup can	banana	toy power drill	chain	mustard bottle	wood block	screw driver
GPD	49%	**100%**	30%	60%	90%	20%	**80%**	0%	90%	**90%**	20%
Ours 2-classes	81%	**100%**	50%	**80%**	**100%**	**90%**	70%	**60%**	**100%**	**90%**	70%
Ours 3-classes	**82%**	90%	**70%**	70%	**100%**	**90%**	**80%**	**60%**	90%	**90%**	**80%**

Robot experiments

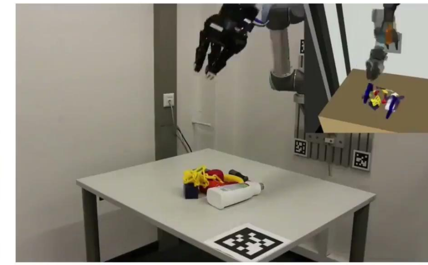
Comparative experiments on object set 2

PointNetGPD ×4.8
7/7 Succeed/Trail



GPD ×7.0
7/11 Succeed/Trail

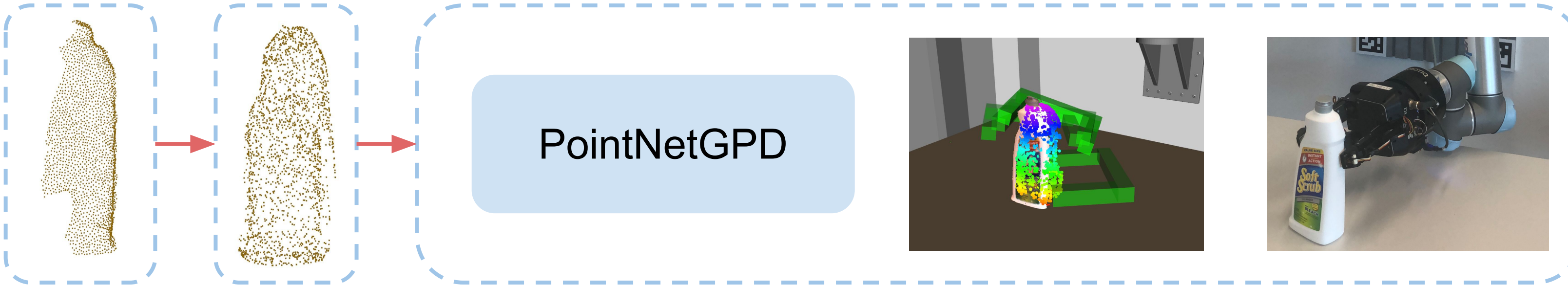
PointNetGPD ×4.8
0/0 Succeed/Trial



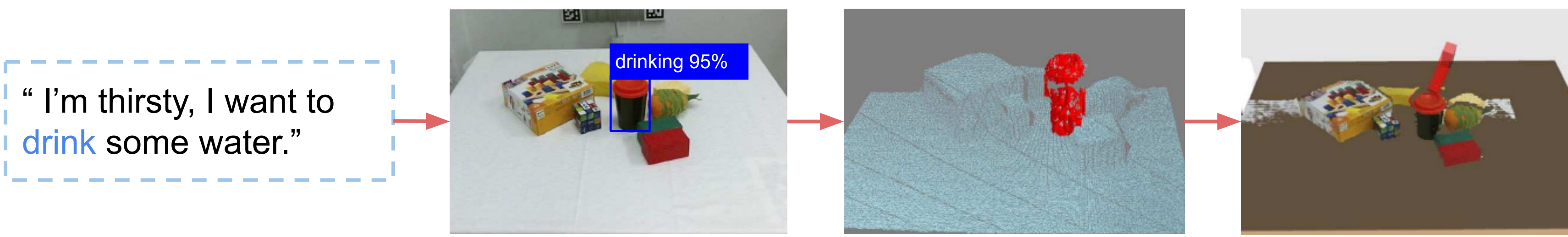
GPD ×7.0
0/0 Succeed/Trial

	GPD		Ours 2-classes		Ours 3-classes	
	Success rate	Completion rate	Success rate	Completion rate	Success rate	Completion rate
Set 1	84.83%	95%	86.54%	94.08%	**89.33%**	**100%**
Set 2	61.13%	81.50%	61.07%	84.38%	**66.20%**	**95%**

Further applications



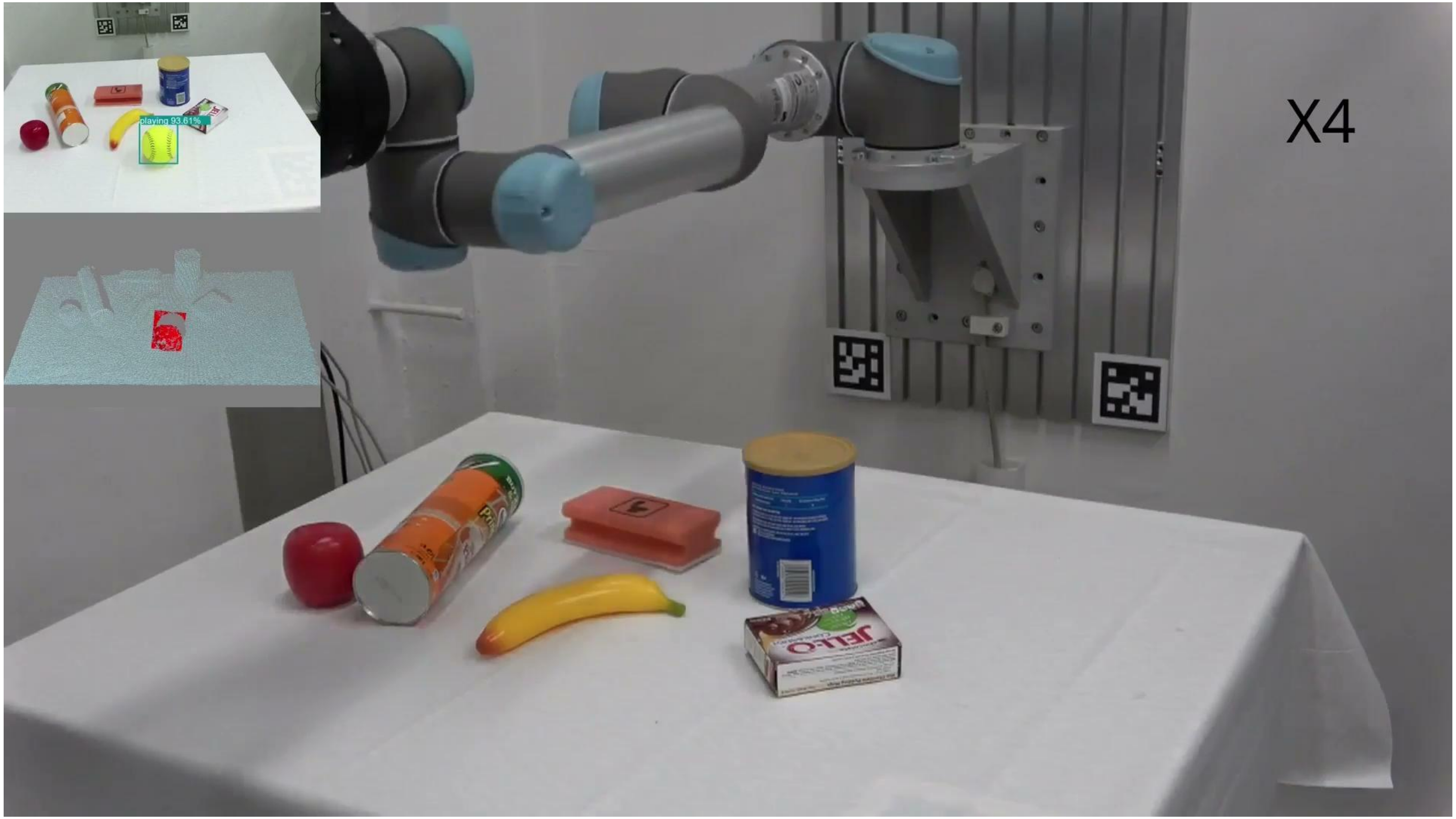
[2] Chen *et al.*, **Improving Object Grasp Performance via Transformer-based Sparse Shape Completion**. *Journal of Intelligent & Robotic Systems*, 2022.



Mi et al., “Intention-Related Natural Language Grounding via Object Affordance Detection and Intention Semantic extraction”, Frontiers in Neurorobotics, 2020.

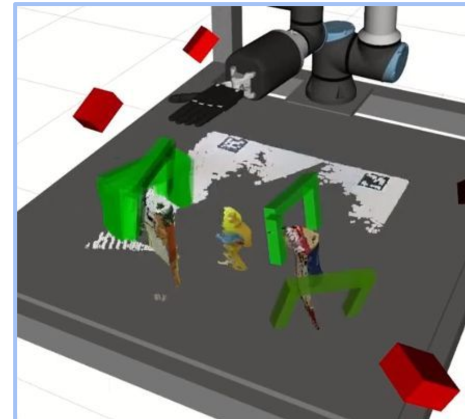

X4

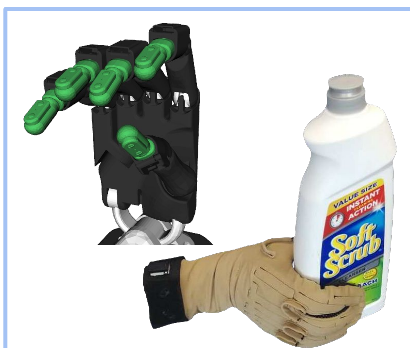
Intention-Related Natural Language Grounding via Object Affordance Detection and Intention Semantic extraction



Multifingered Grasping

- Why use dexterous hand
    - For dexterous manipulation
    - Human-like motion (human friendly)
- Why train in simulation
    - Training in simulation is fast
    - More efficient than the real world
    - Safer than the real world





Synergy dataset

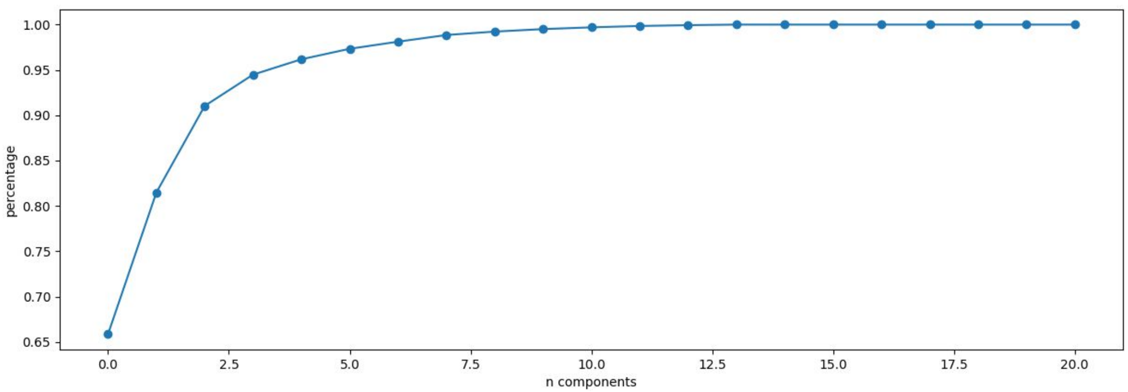
Simulation environment

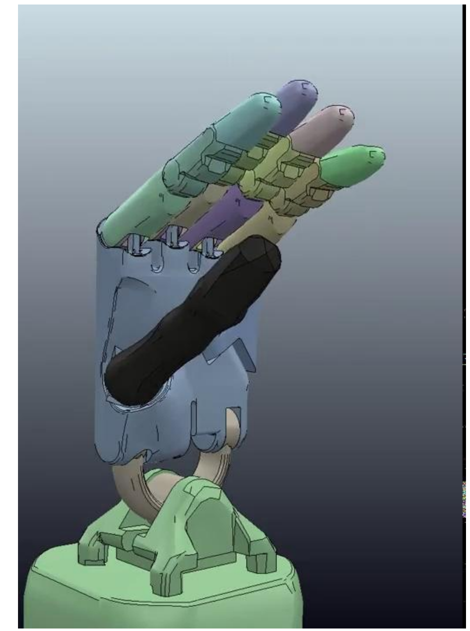
Real robot environment

Liang et al., “Multifingered Grasping Based on Multimodal Reinforcement Learning”, IEEE Robotics and Automation Letters (RA-L), 2022.

PCA based grasp synergy

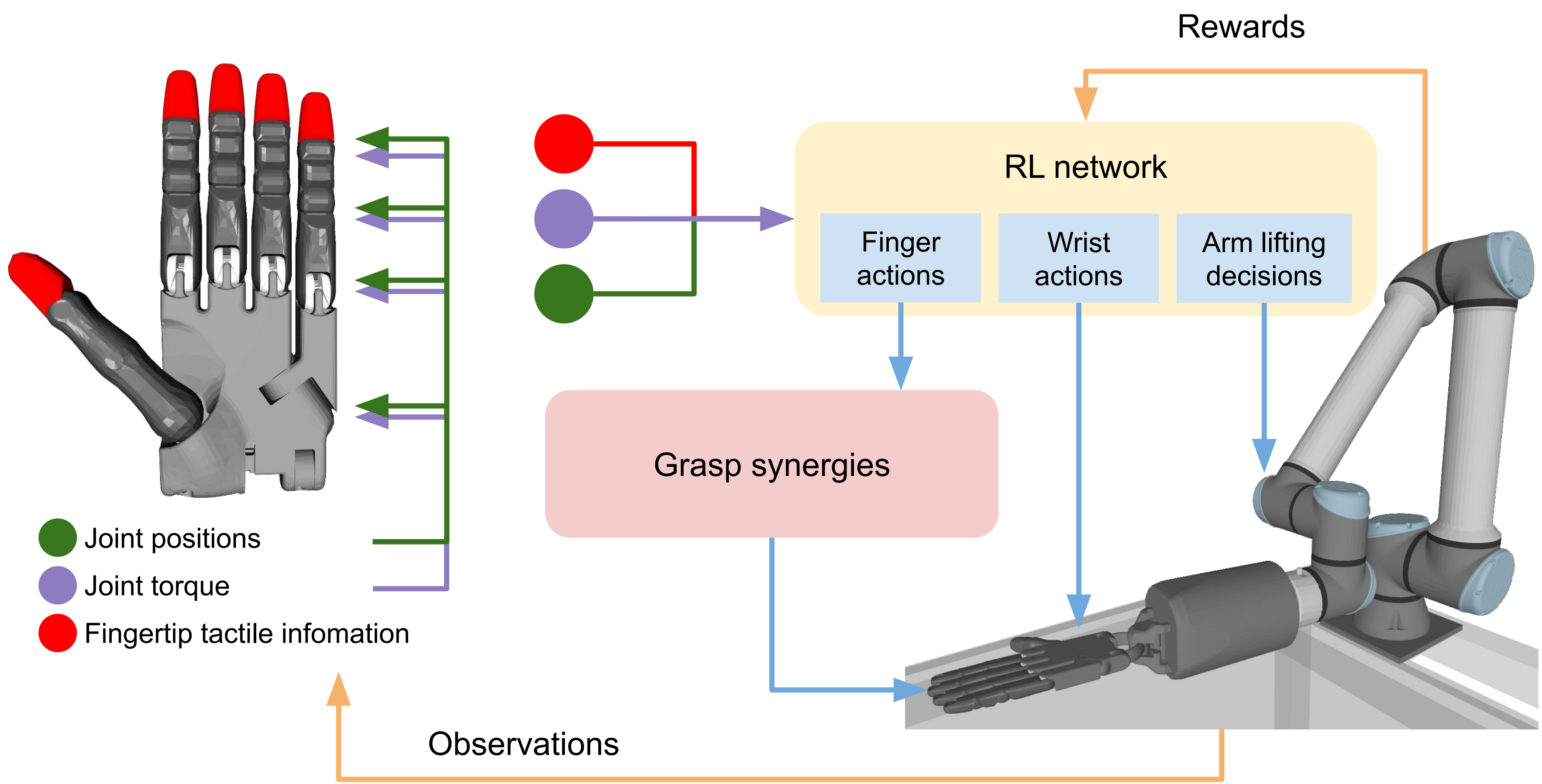
- Cyberglove teleopration for data collection
- Calculate PCA for the 22 joints (without wrists)
- Plot the graph to see the reconstructed precision
- When n components is 5, can reconstruct 95%





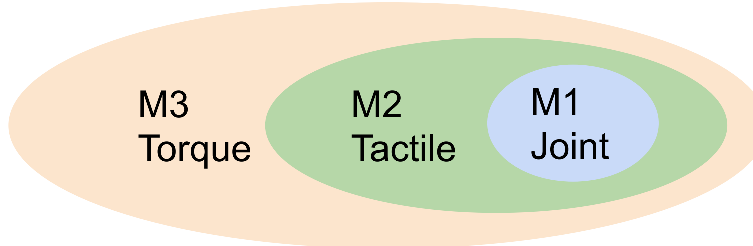
First three components demo

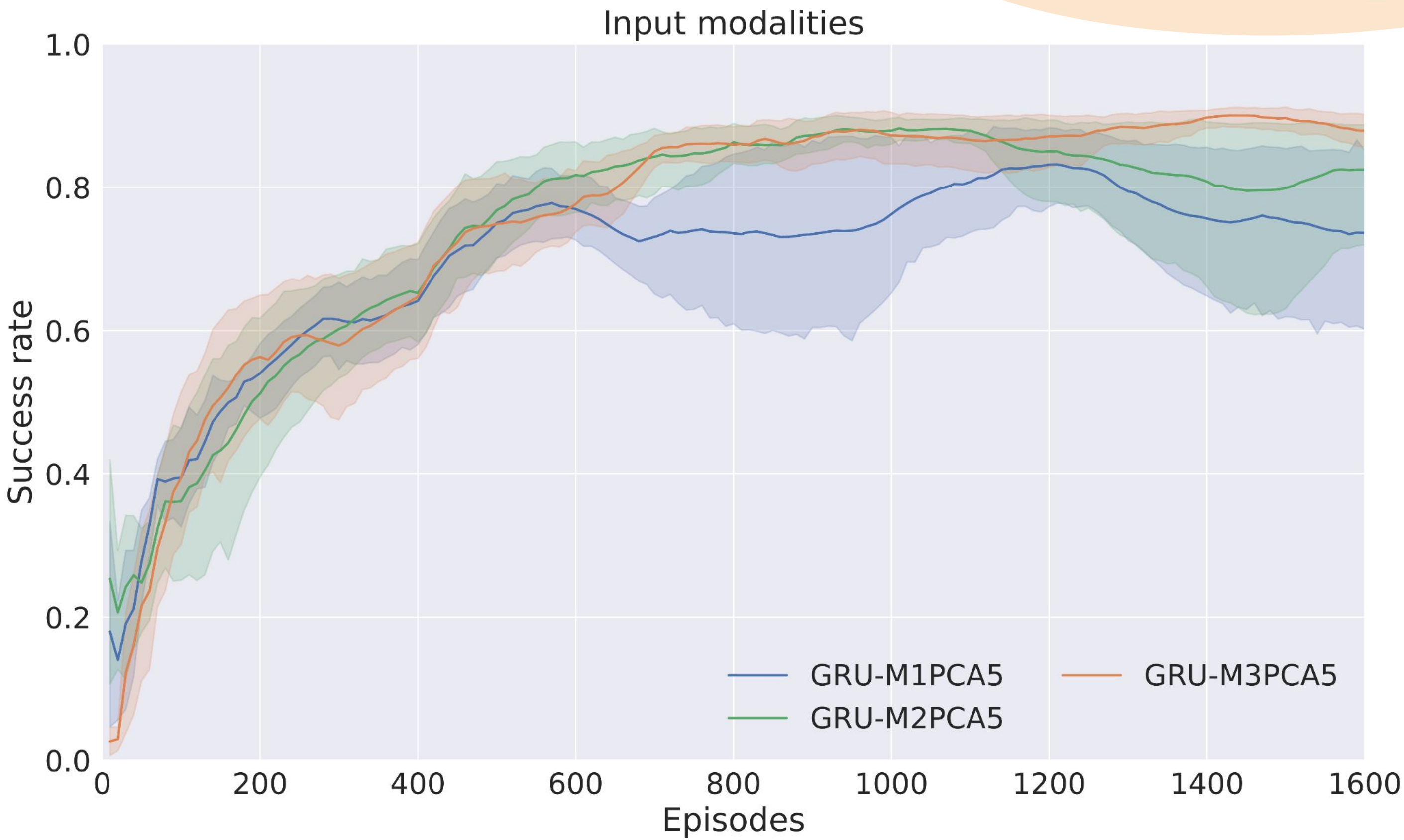
RL framework



Joint, Tactile, Torque

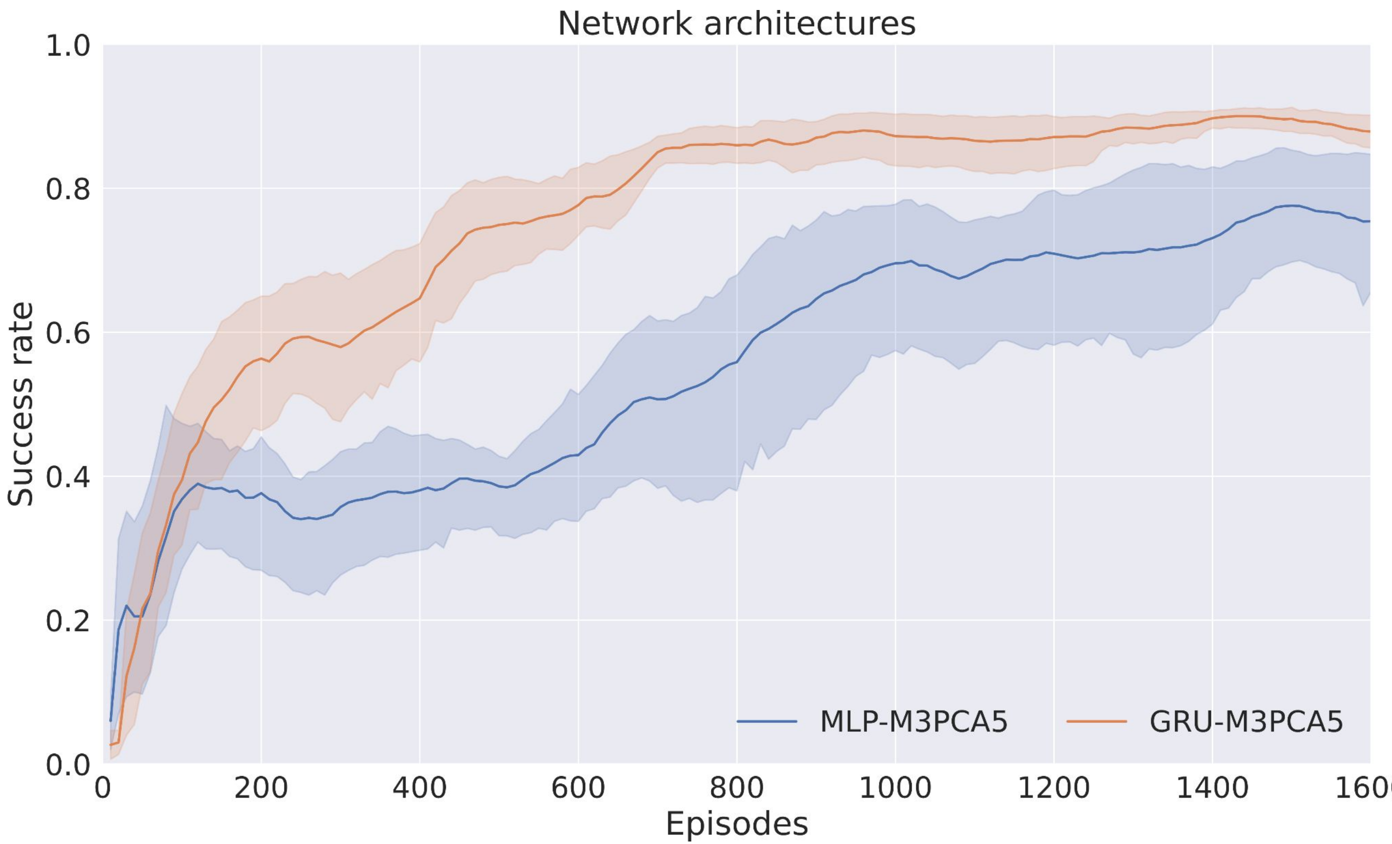
Different input modalities number M1, M2, M3





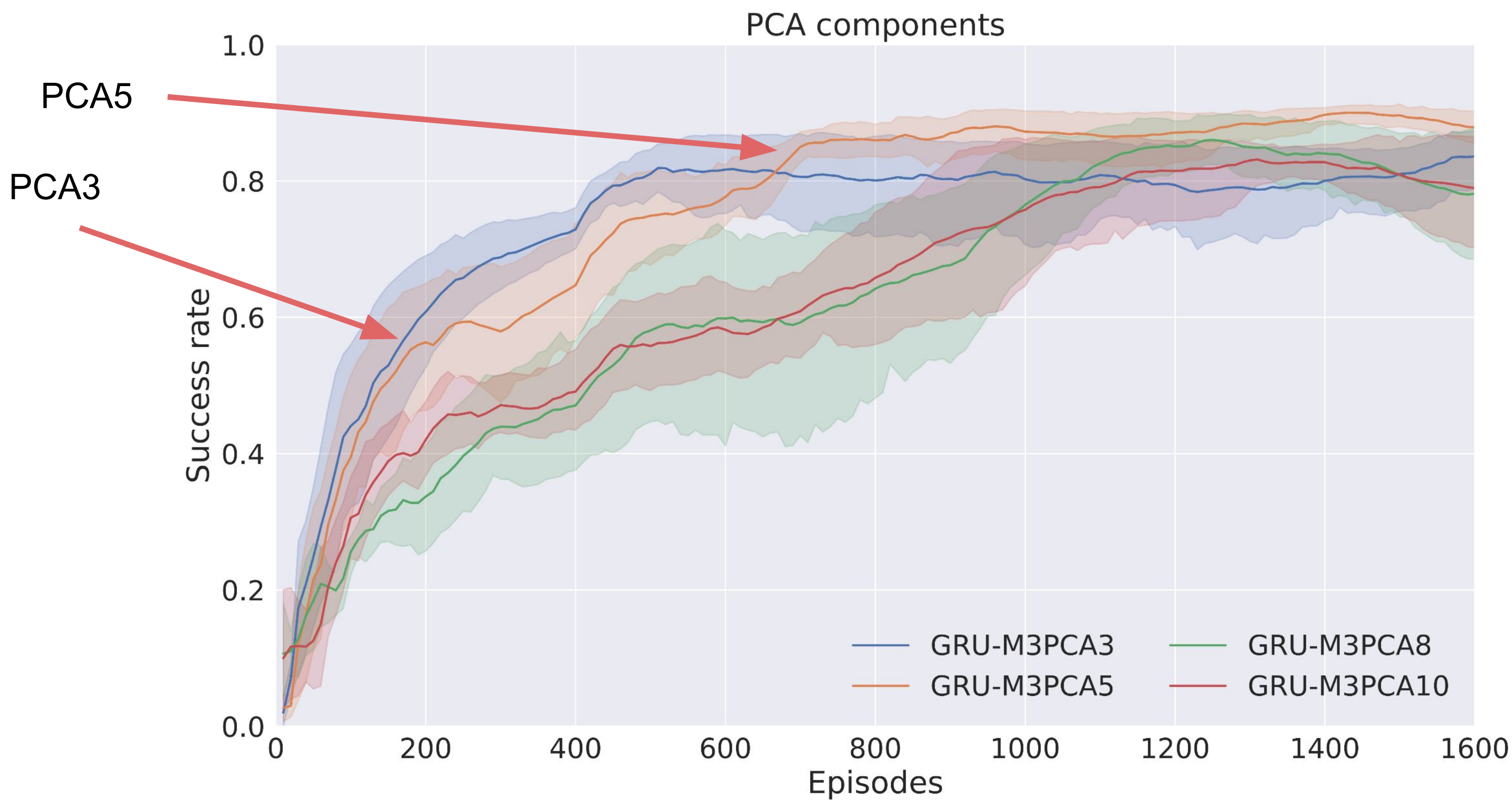
MLP and GRU

Different network types: MLP and RNN (GRU)



PCA 3, 5, 8, 10

Different PCA components: 3, 5, 8, 10



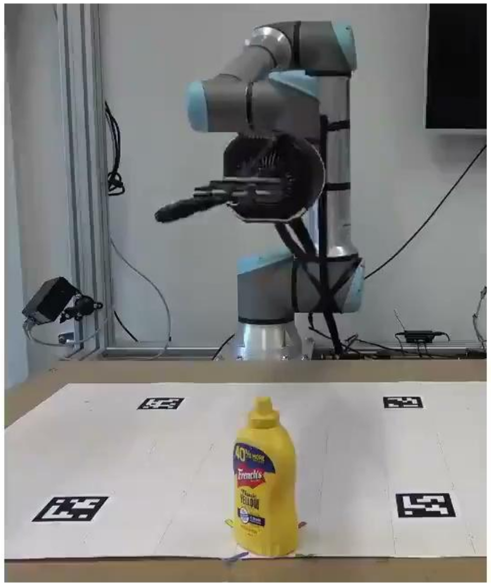
Real robot experiments

Model used in the experiment: GRU-M3PCA5

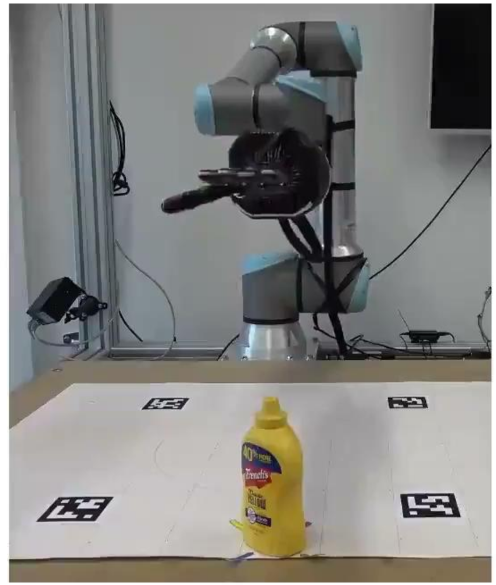
Real robot experiments

Grasp comparison with different models

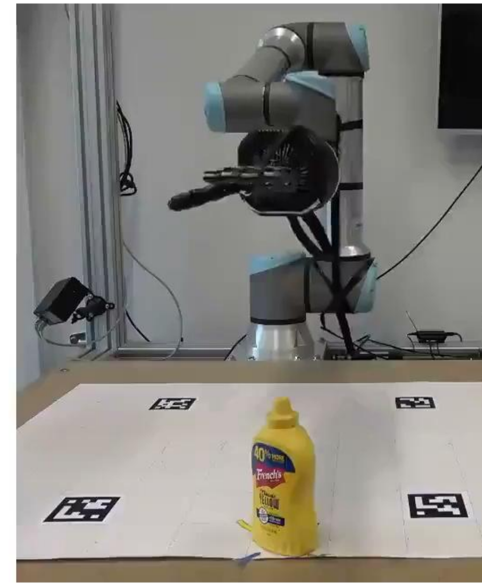
X8



GRU-M3PCA5 RL



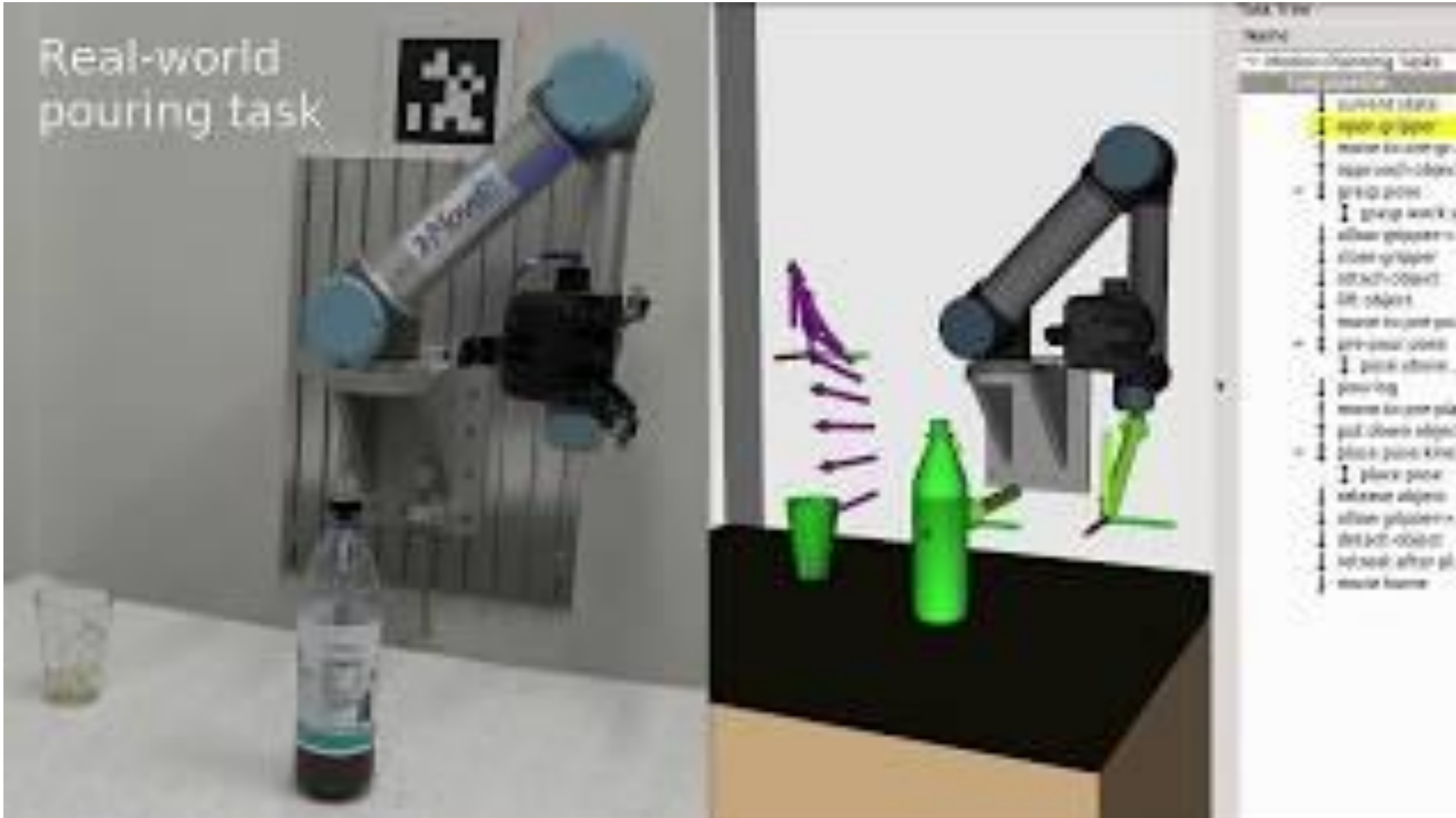
GRU-M2PCA5 RL



No RL

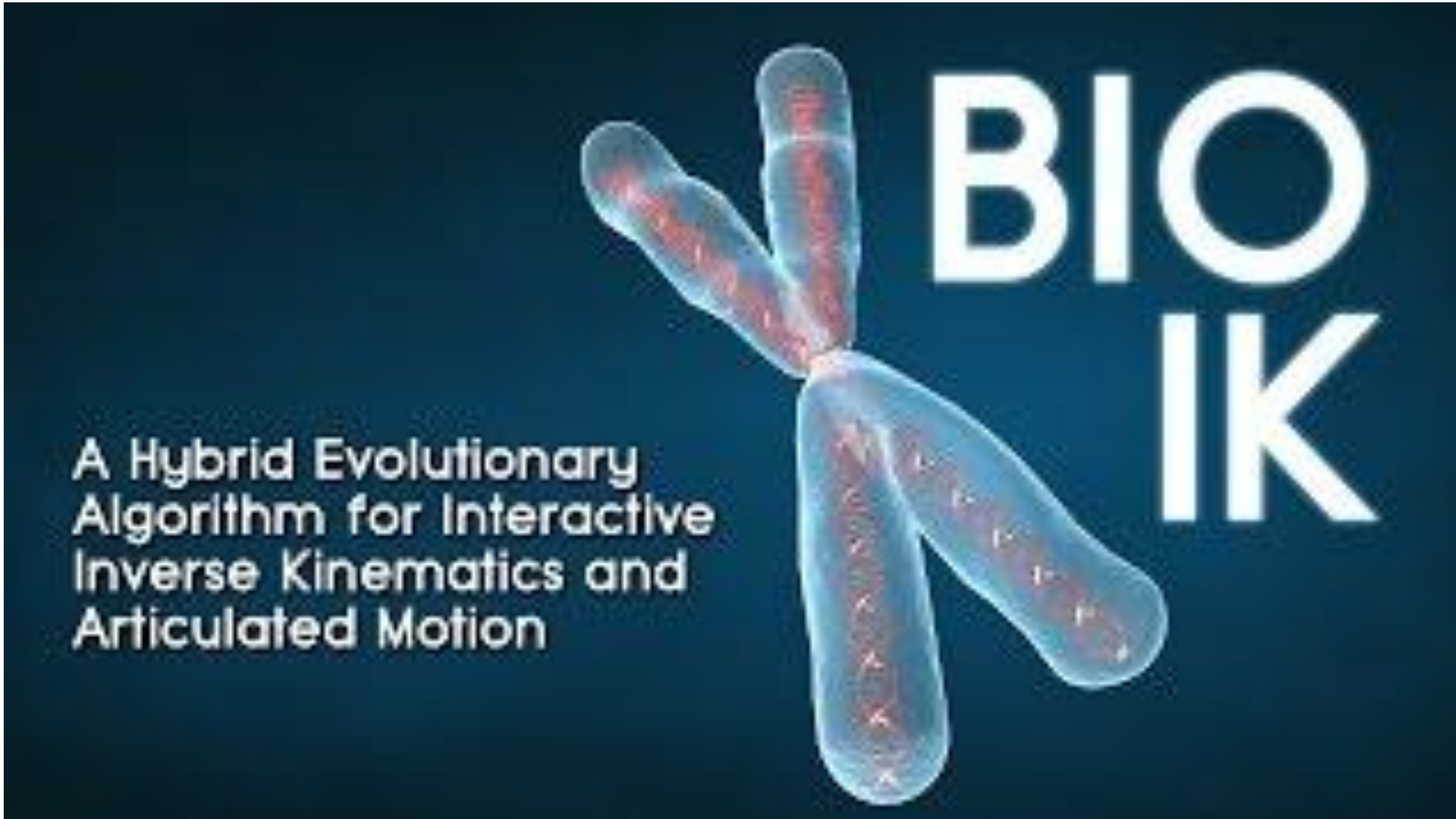
Note: M3 means three modalities input, M2 means two modalities input. PCA5 means the Shadow hand joint space is reduced to 5 using PCA.

MoveIt! Task Constructor



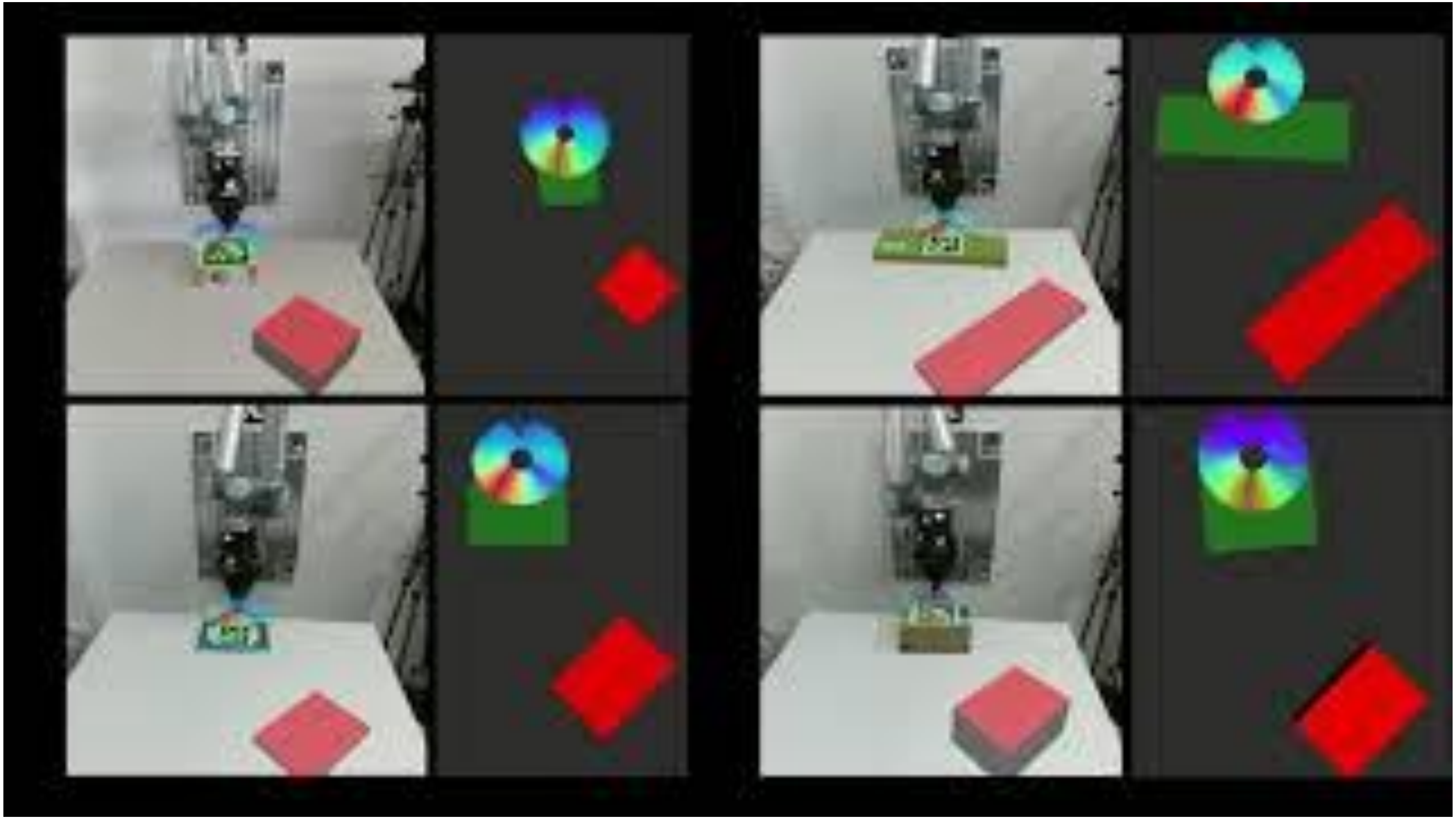
Görner et. al., MoveIt! Task Constructor for task-level motion planning, ICRA-2019.

Bioik - IK for genetic kinematic trees



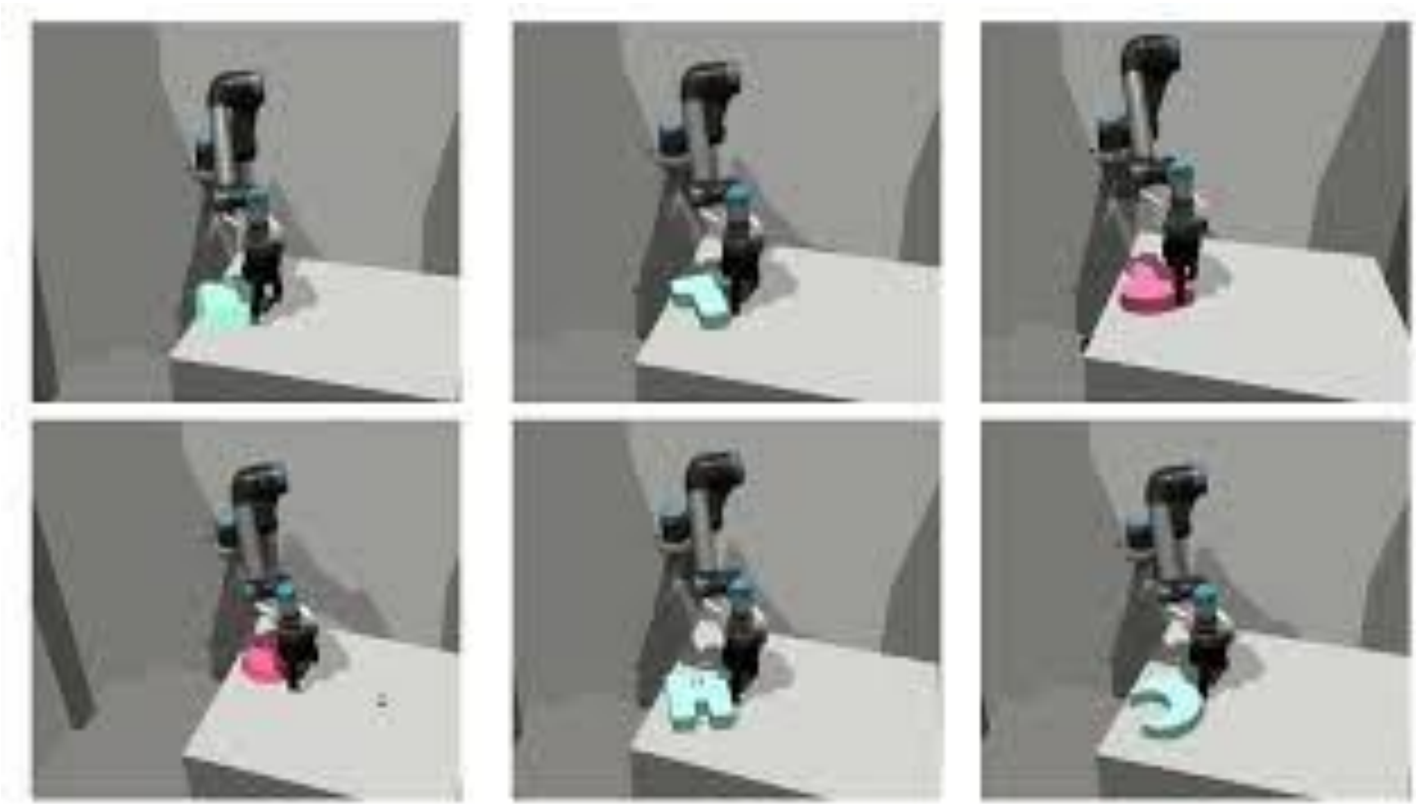
Ruppel, et. al. Cost Functions to Specify Full-Body Motion and Multi-Goal Manipulation Tasks, ICRA-2018.
Starke, et. al., Memetic Evolution for Generic Full-Body Inverse Kinematics in Robotics and Animation, Transactions on Evolutionary Computation, 2018.

Pushing



Cong et. al. Self-Adapting Recurrent Models for Object Pushing from Learning in Simulation. IROS-2020

Pushing and Grasping



Zhang, H., "Reinforcement Learning Based Pushing and Grasping Objects from Ungraspable Poses", International Conference on Robotics and Automation (ICRA) 2023.

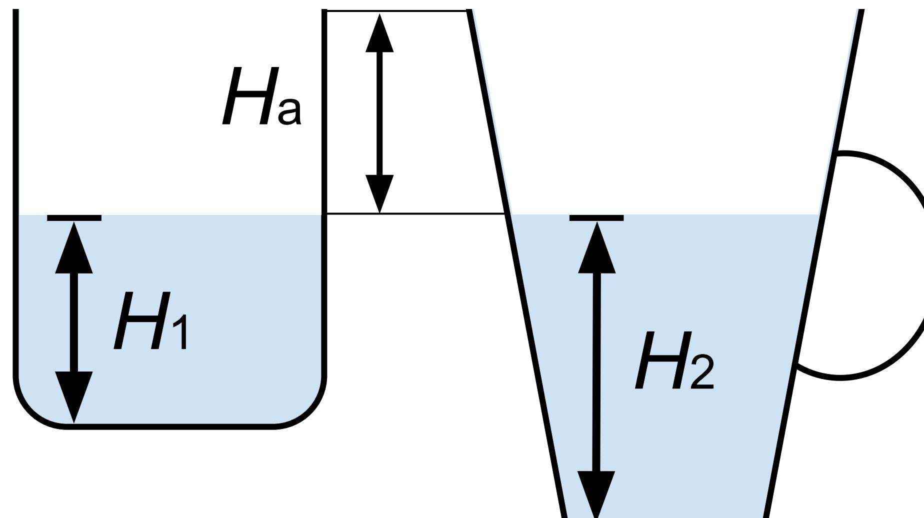
Robotic pouring

Why use multimodal:

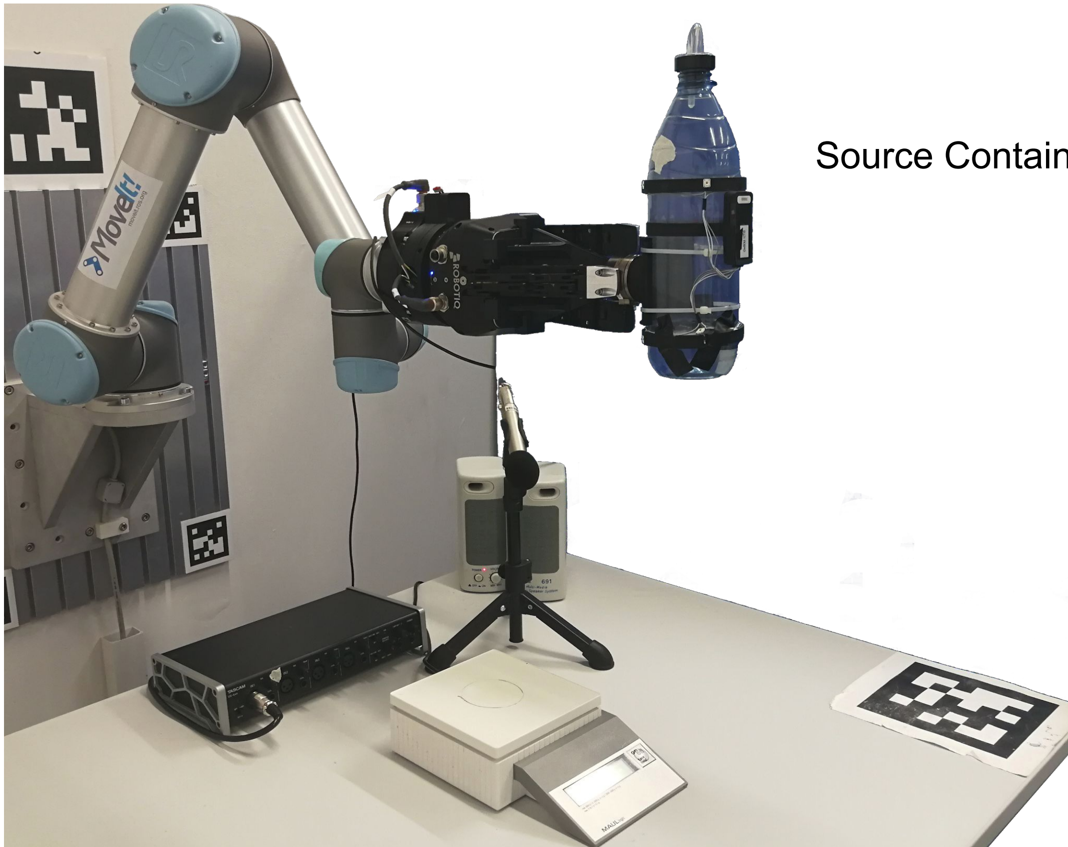
- Visual sensing methods cannot generalize to occluded situations
- Audio sensing requires a quiet environment

Why use ***H_a*** :

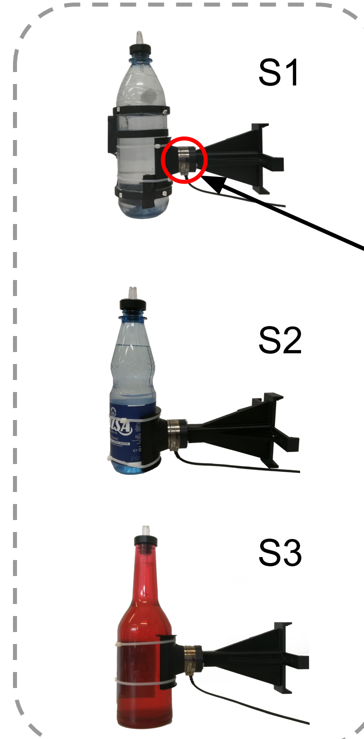
- Length of air column ***H_a*** has a direct relationship with **resonance frequency** of the air.



Hardware setup



Source Containers



ATI Force / Torque Sensor

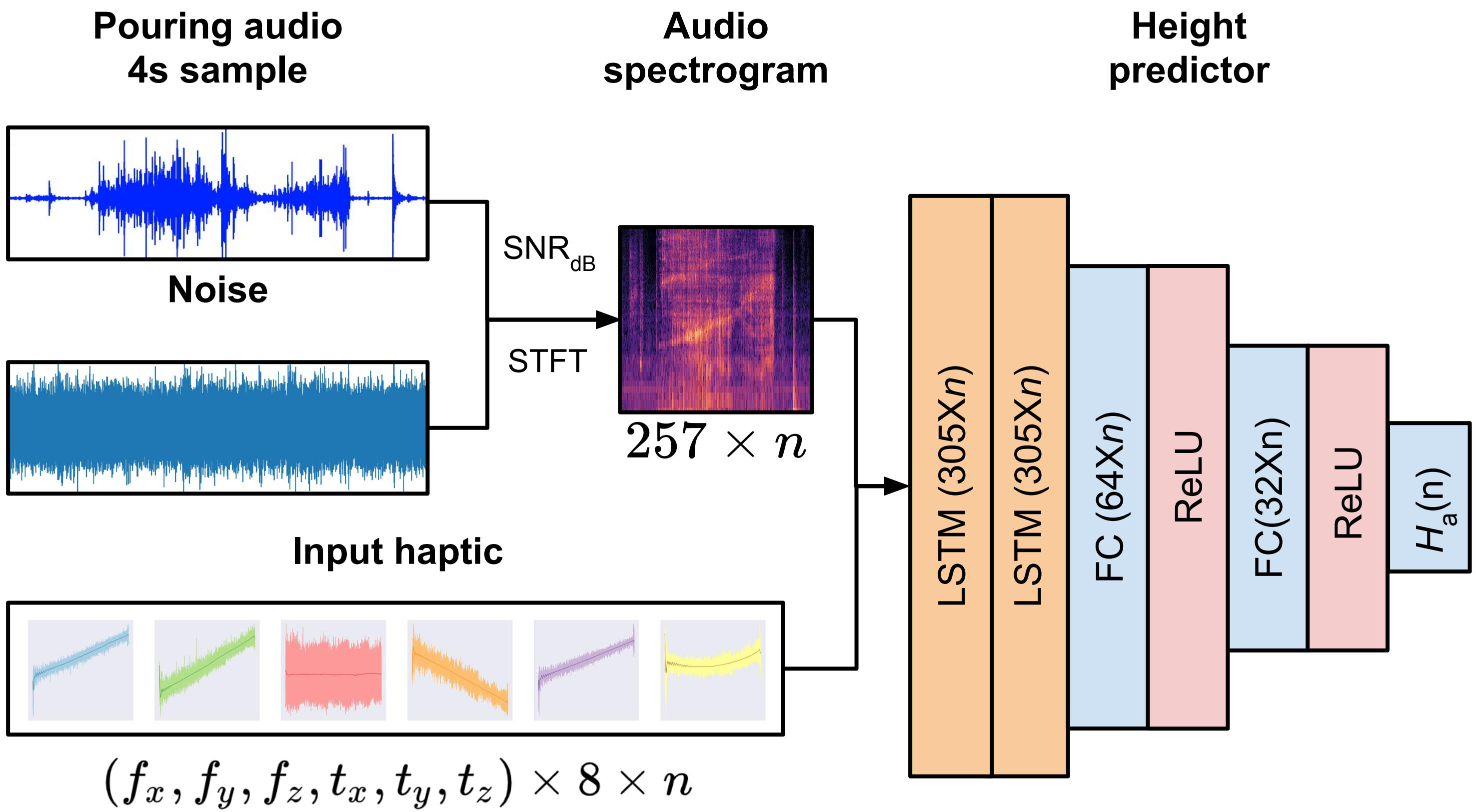
Target Containers

Pouring network

Input modalities:

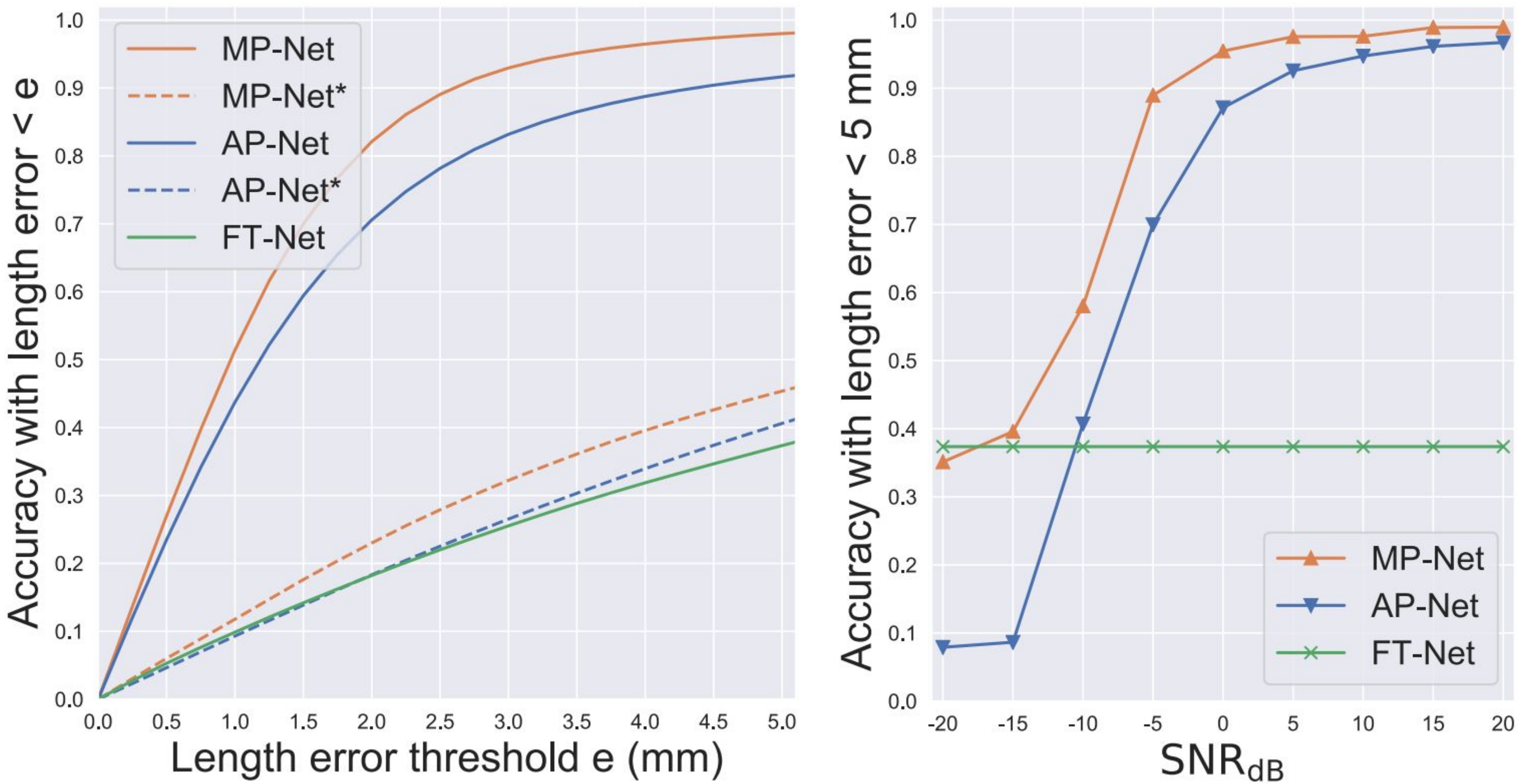
Input	audio	mixed noise	haptic
AP-Net*	✓		
AP-Net	✓	✓	
MP-Net*	✓		✓
MP-Net	✓	✓	✓
FT-Net			✓

MP-Net



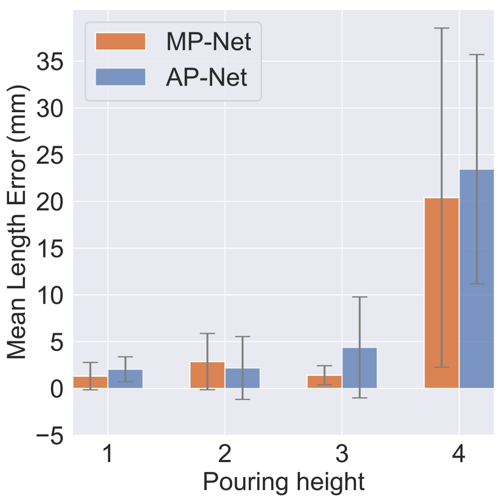
Liang et al., "Making sense of audio vibration for liquid height estimation in robotic pouring". IROS 2019.
Liang et al., "Robust robotic pouring using audition and haptics". IROS 2020.

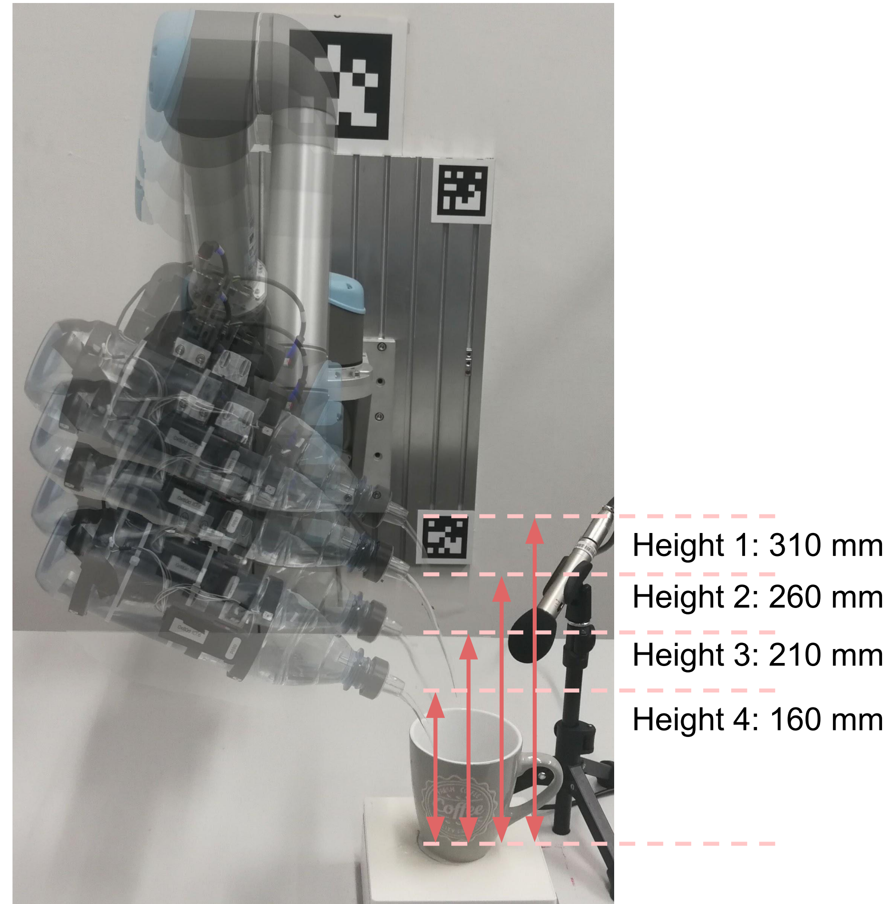
Network evaluation



Robot experiments

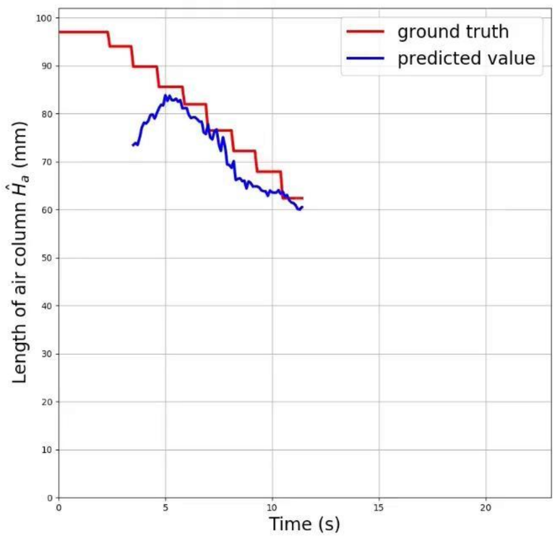
1. Different target containers
2. **Different pouring heights**
3. Different source containers
4. Varying noise conditions
5. Varying positions of noise source
6. Varying initial liquid height in target containers
7. Different types of liquid
8. Different types of noise sources

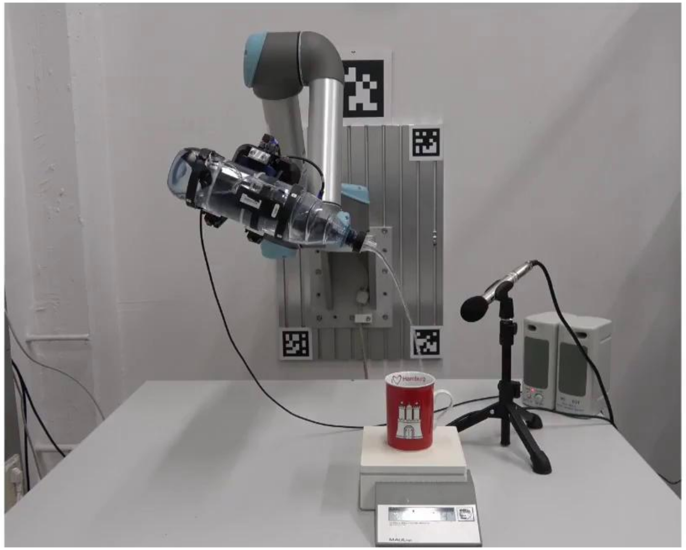




Robot experiment

Pouring into unknown container

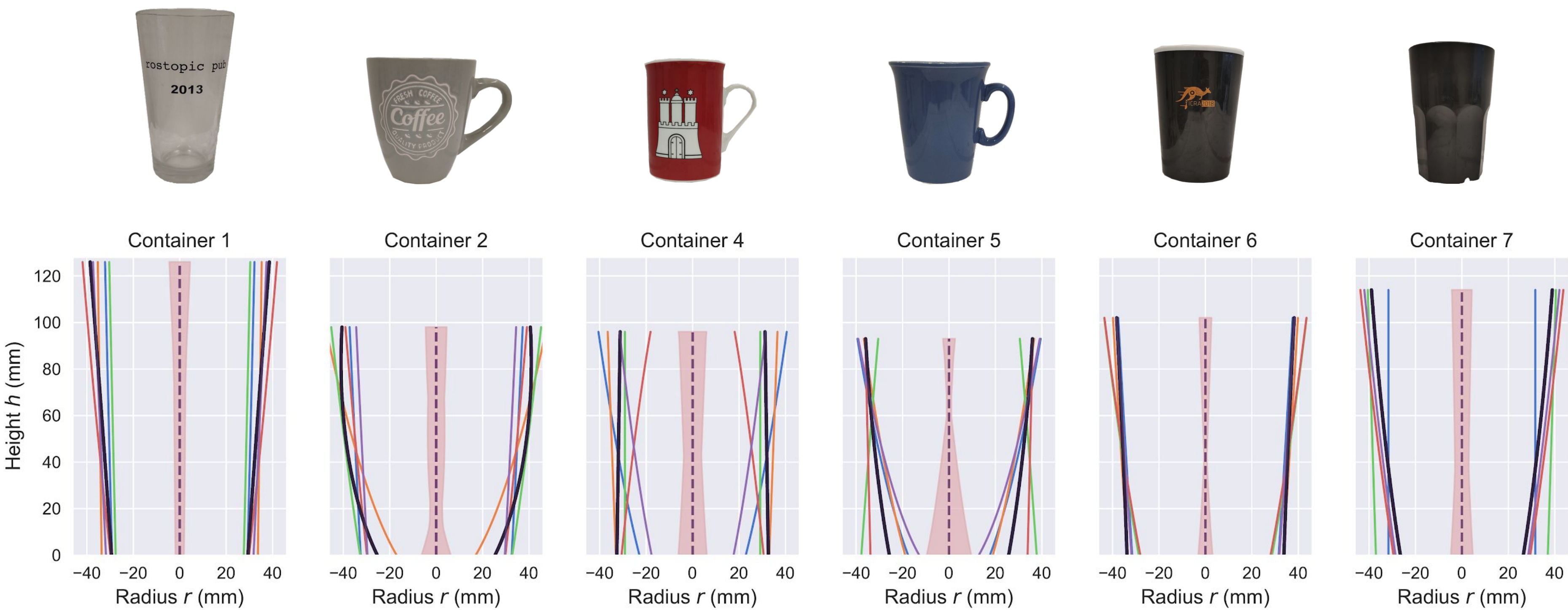




$SNR_{dB} = 5$ $H_a = 40$ mm

Shape prediction

- Height + Weight + Liquid density → Inner shape of symmetry containers



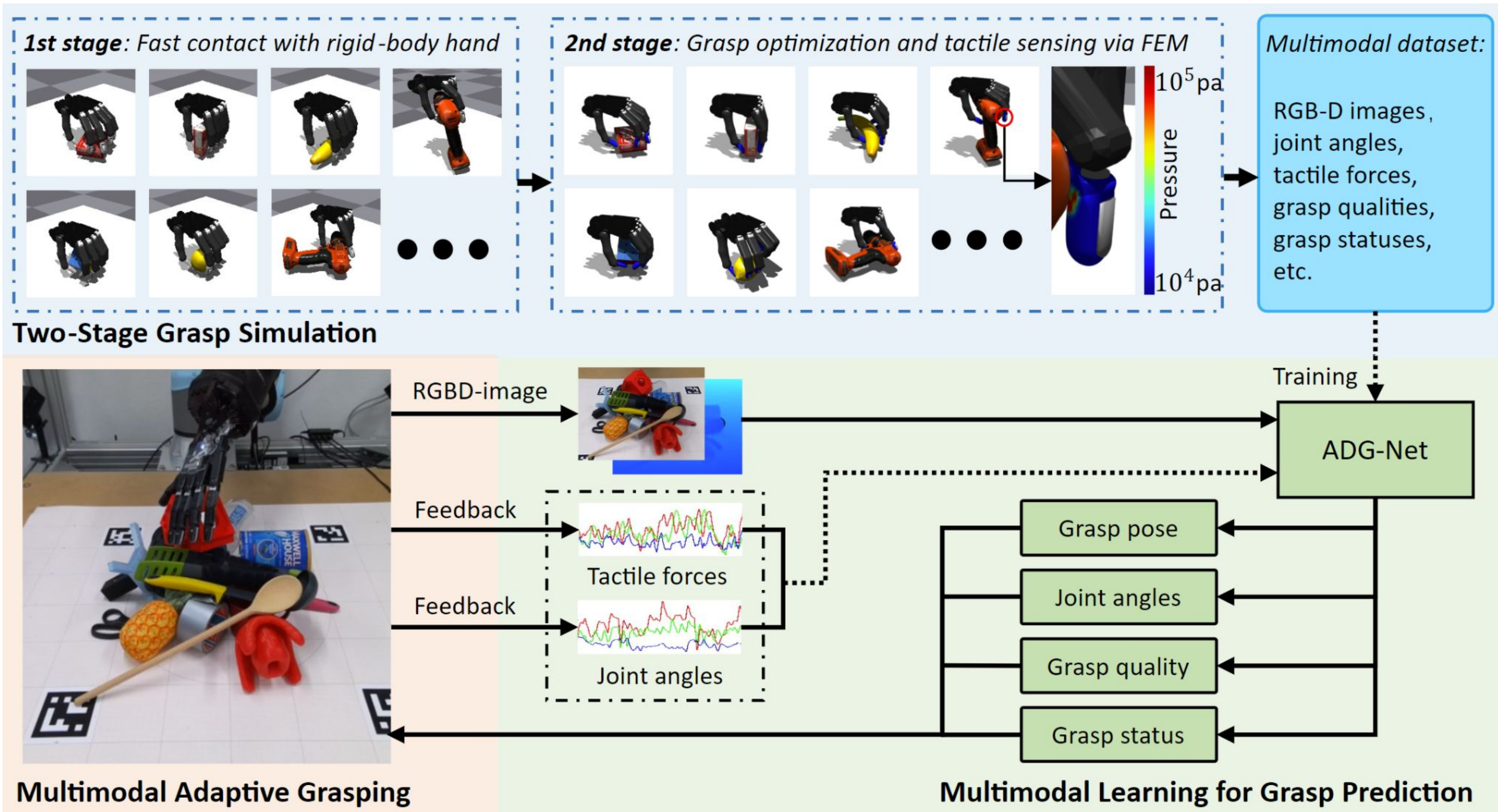
Dexterous Hand-Arm Teleoperation

Li, S., et al., "A Dexterous Hand-Arm Teleoperation System based on Hand Pose Estimation and Active Vision", *IEEE Transactions on Cybernetics*.

AR-based teleoperation

Dennis Krupke, et.al, "Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction", IROS 2018

Multimodal Sim2Real Dexterous Grasping



Zhang, H. et.al, "ADG-Net: A Sim-to-Real Multimodal Learning Framework for Adaptive Dexterous Grasping", IEEE Trans. Cybernetics, under review.

Multimodal Sim2Real Dexterous Grasping

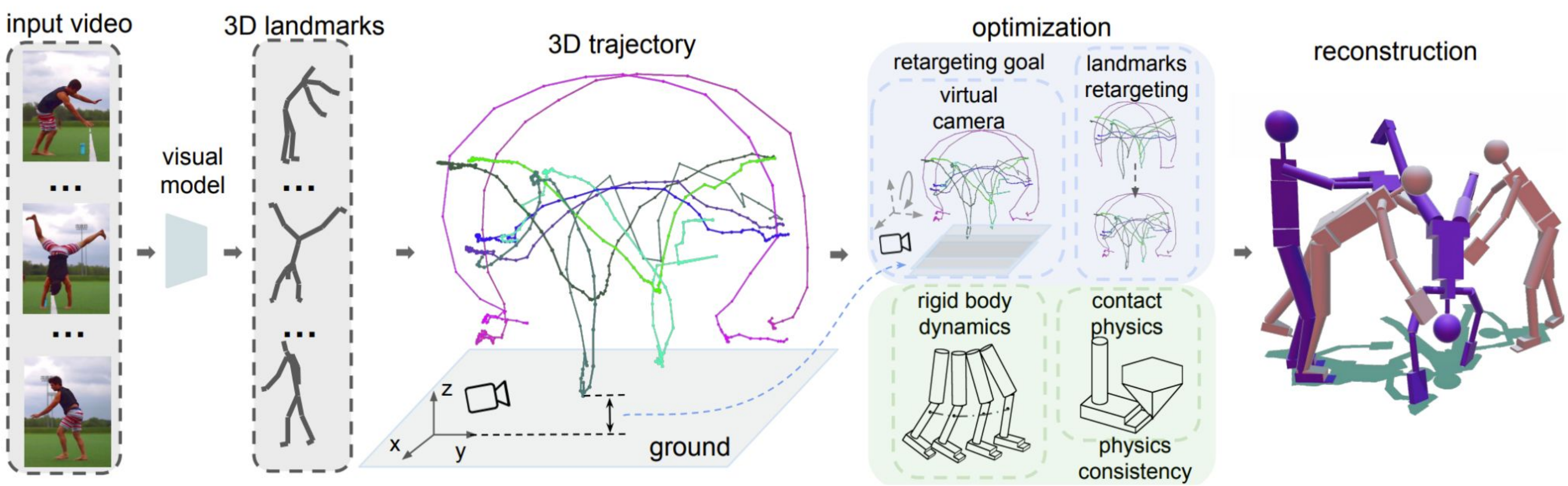
ADG-Net: A Sim-to-Real Multimodal Learning Framework for Adaptive Dexterous Grasping

Hui Zhang, Jianzhi Lyu, Hongzhuo Liang, Chuangchuang Zhou, Fuchun Sun, and Jianwei Zhang

TAMS group, Department of Informatics, Universität Hamburg.

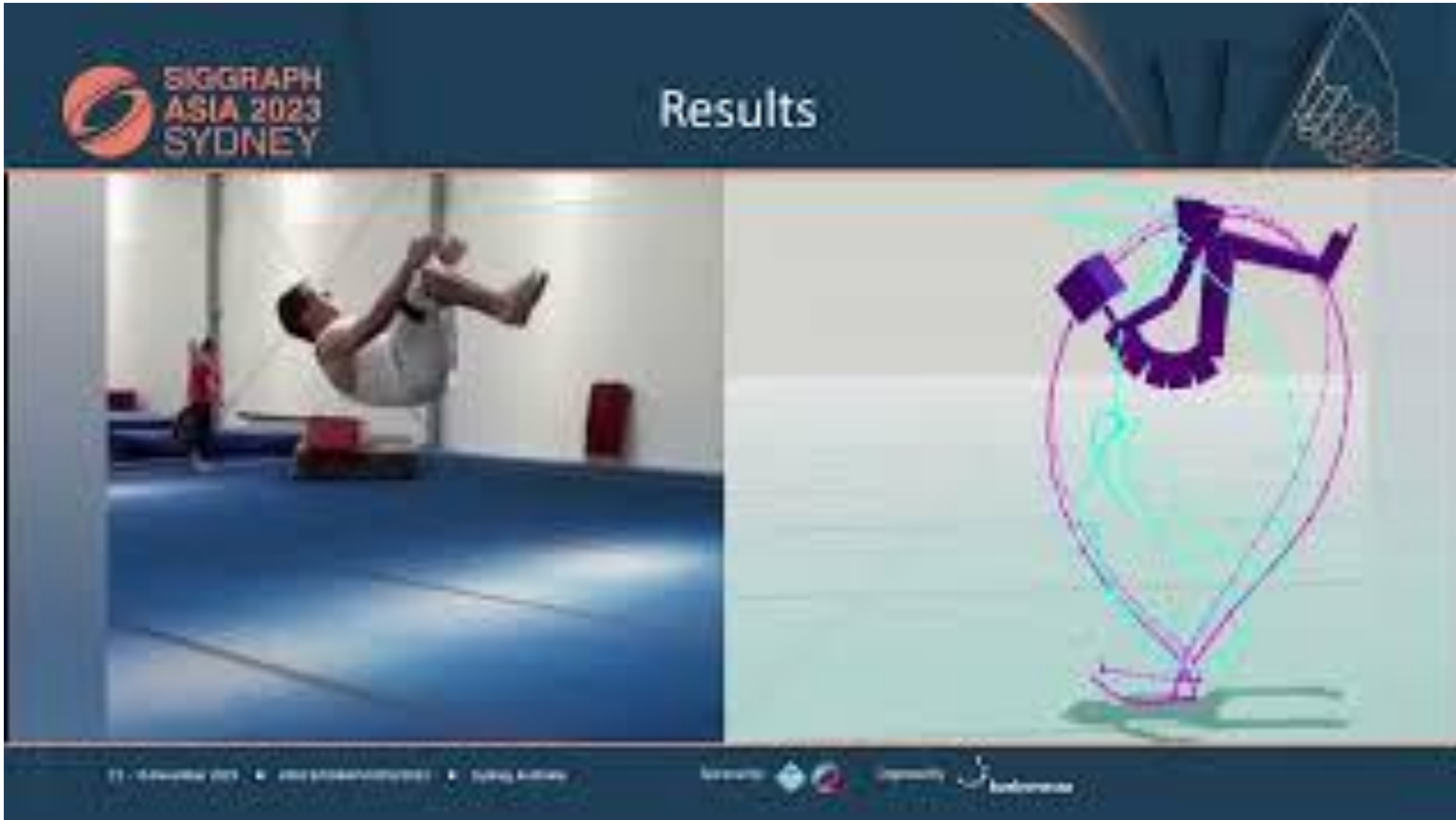
Zhang, H. et.al, "ADG-Net: A Sim-to-Real Multimodal Learning Framework for Adaptive Dexterous Grasping", IEEE Trans. Cybernetics, under review.

Human Motion Reconstruction



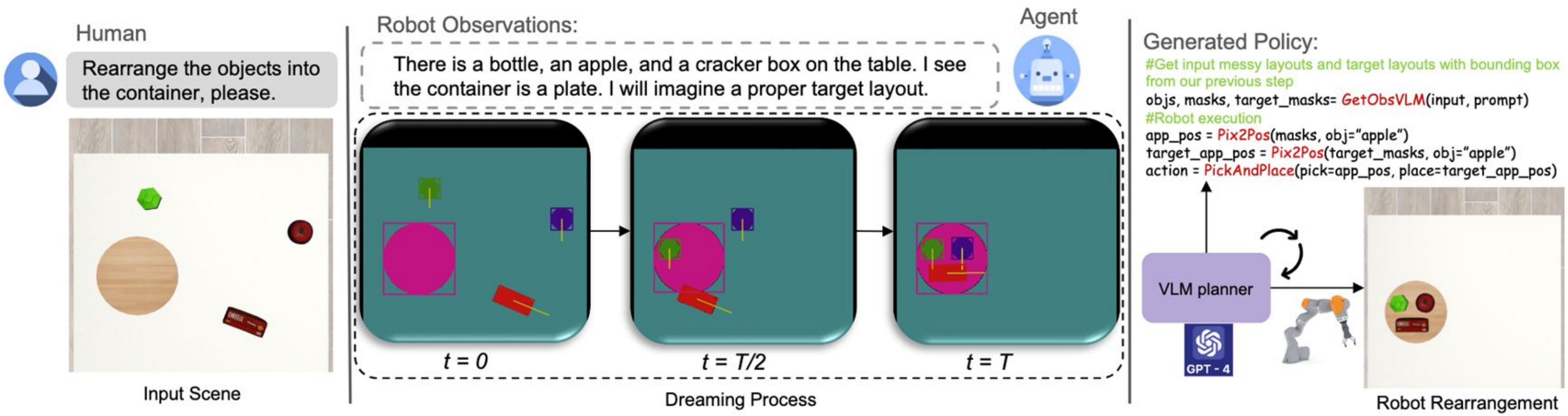
Lin Cong, et al., "Efficient Human Motion Reconstruction from Monocular Videos with Physical Consistency Loss", Siggraph-Asia 2023

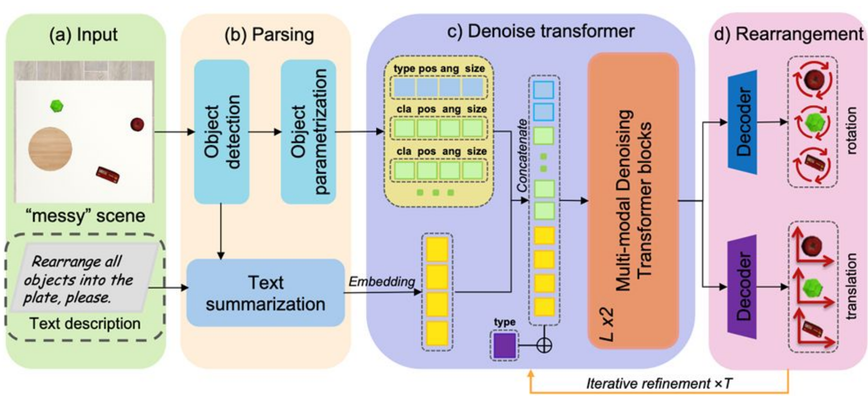
Human Motion Reconstruction



Lin Cong, et al., "Efficient Human Motion Reconstruction from Monocular Videos with Physical Consistency Loss", Siggraph-Asia 2023

VLM-Based Robotic Rearrangement





Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

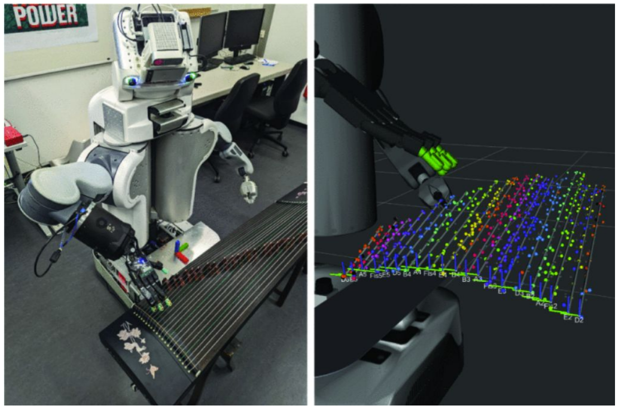
DreamArrangement: Learning Language-conditioned Robotic Rearrangement of Objects via Denoising Diffusion and VLM Planner

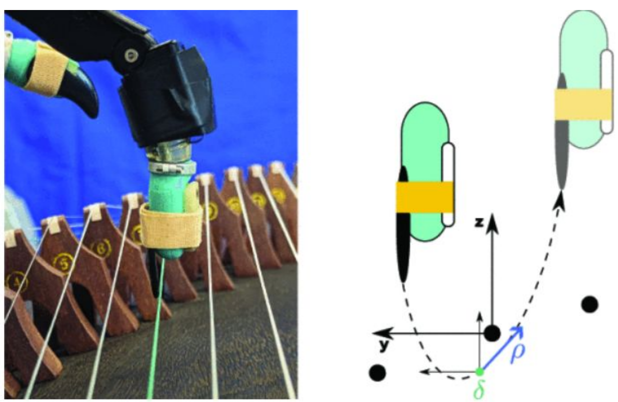
Wenkai Chen[1], Changming Xiao[2], Ge Gao[3], Fuchun Sun[2], Changshui Zhang[2], Jianwei Zhang[1]

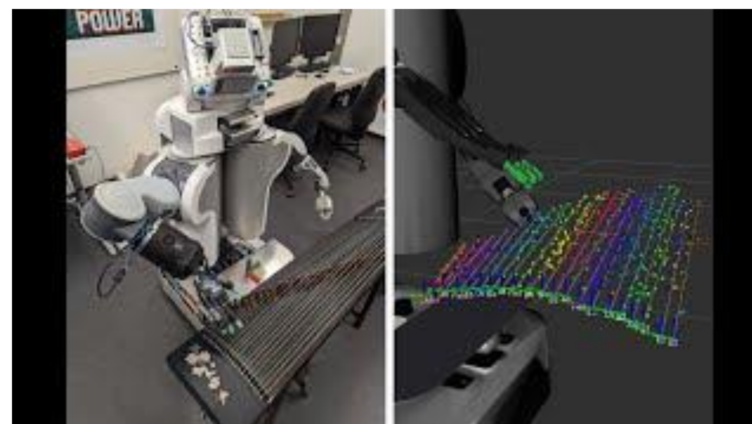
[1]Universität Hamburg, [2]Tsinghua University, [3]Mech-Mind Robotics

Chen, W., et al., 2024. "DreamArrangement: Learning Language-conditioned Robotic Rearrangement of Objects via Denoising Diffusion and VLM Planner". TechRxiv, minor revisions.

Pluck and Play Guzheng

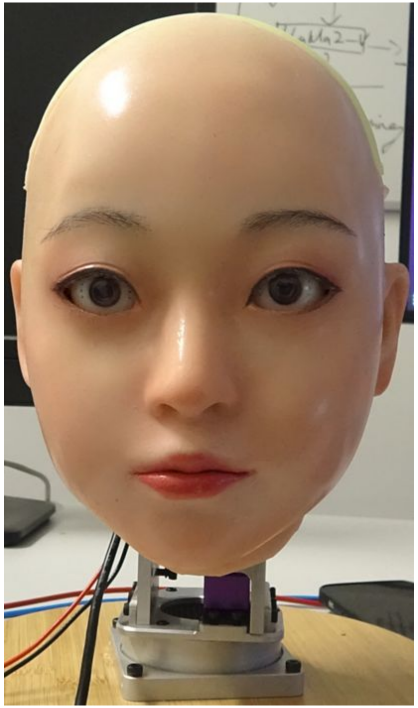






Görner, M., et al., 2024. “Pluck and Play: Self-supervised Exploration of Chordophones for Robotic Playing”, IEEE ICRA 2024.

Human-Robot Interaction



Robot Head Athena2

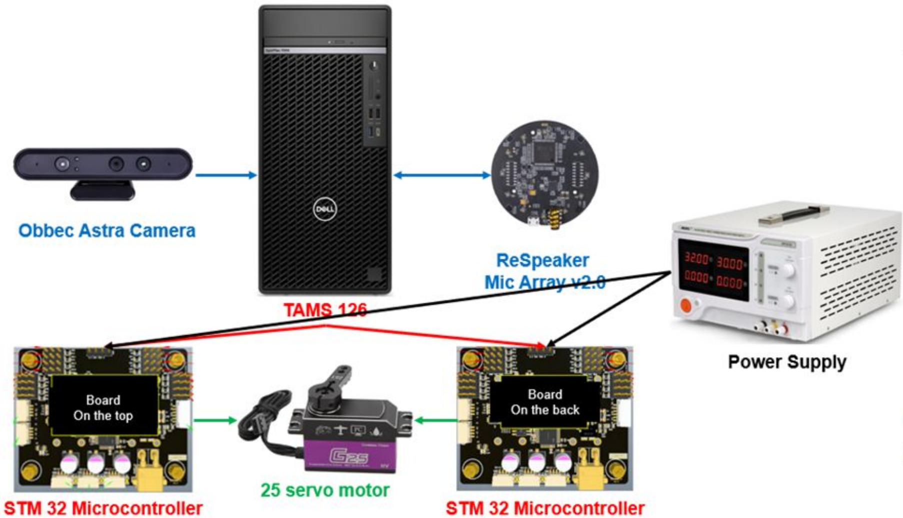
1. Understand human ourself

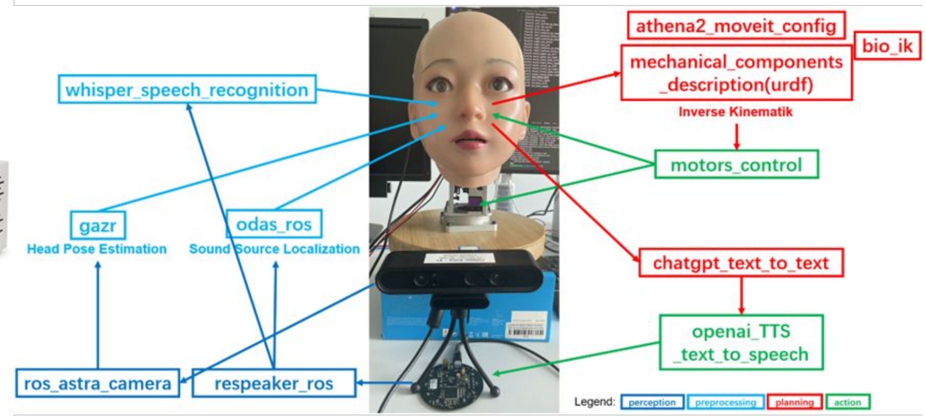
Scientists can use it as a experiment platform to understand human social behavior. Models from Psychology and Human movement science could be implemented, tested and analyzed.

2. Human-Robot Collaboration and Interaction

We want to make robots natural and intuitive to use (Humanoid Robot) and to interact with (Social Assistant, Disease Treatment)

Human-Robot Interaction





Human-Robot Interaction