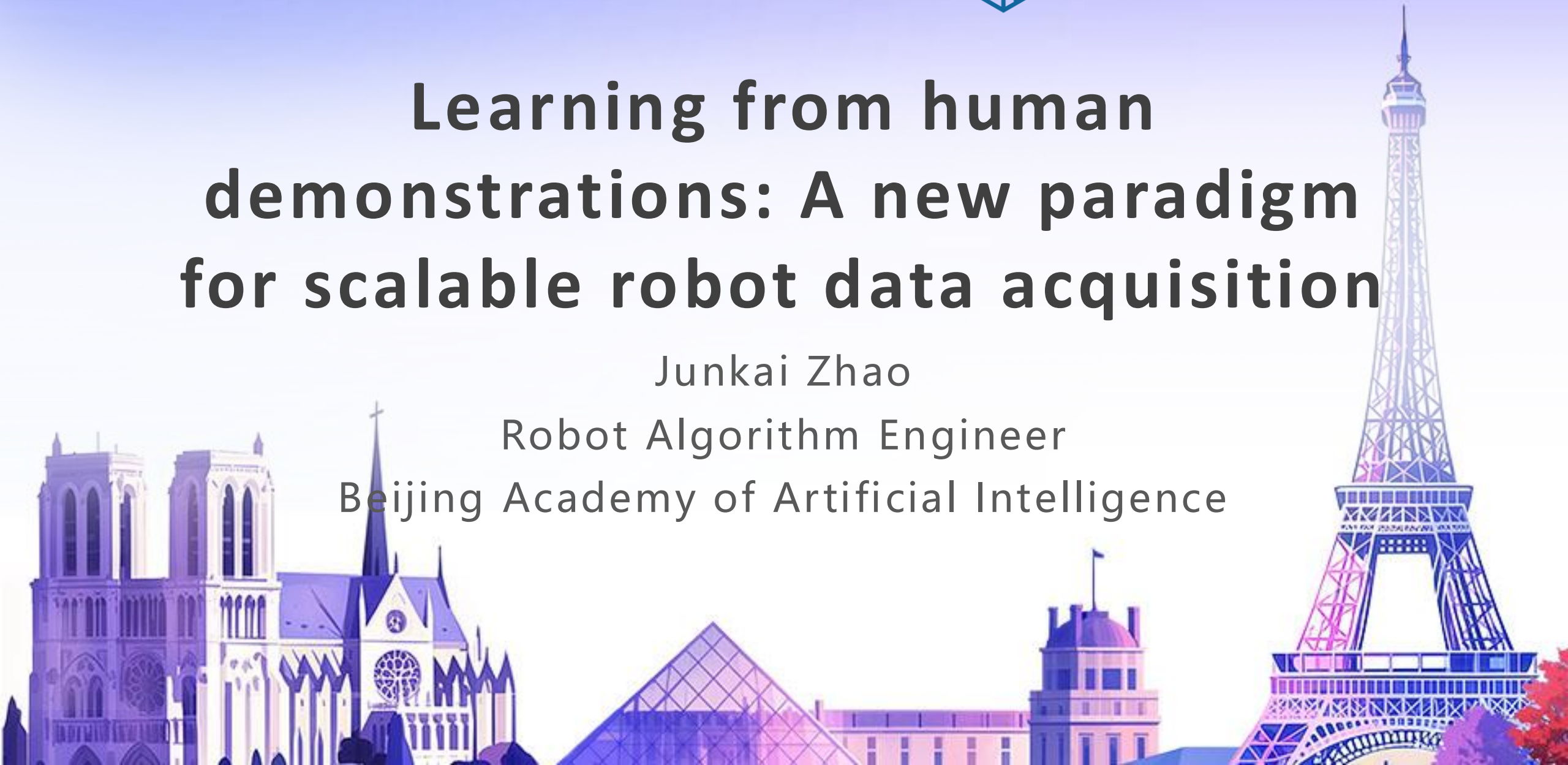
GOSIM

BAAI 智源研究院 GOSIM AI Paris 2025

Learning from human demonstrations: A new paradigm for scalable robot data acquisition

Junkai Zhao
Robot Algorithm Engineer
Beijing Academy of Artificial Intelligence



GOSIM

BAAI 智源研究院

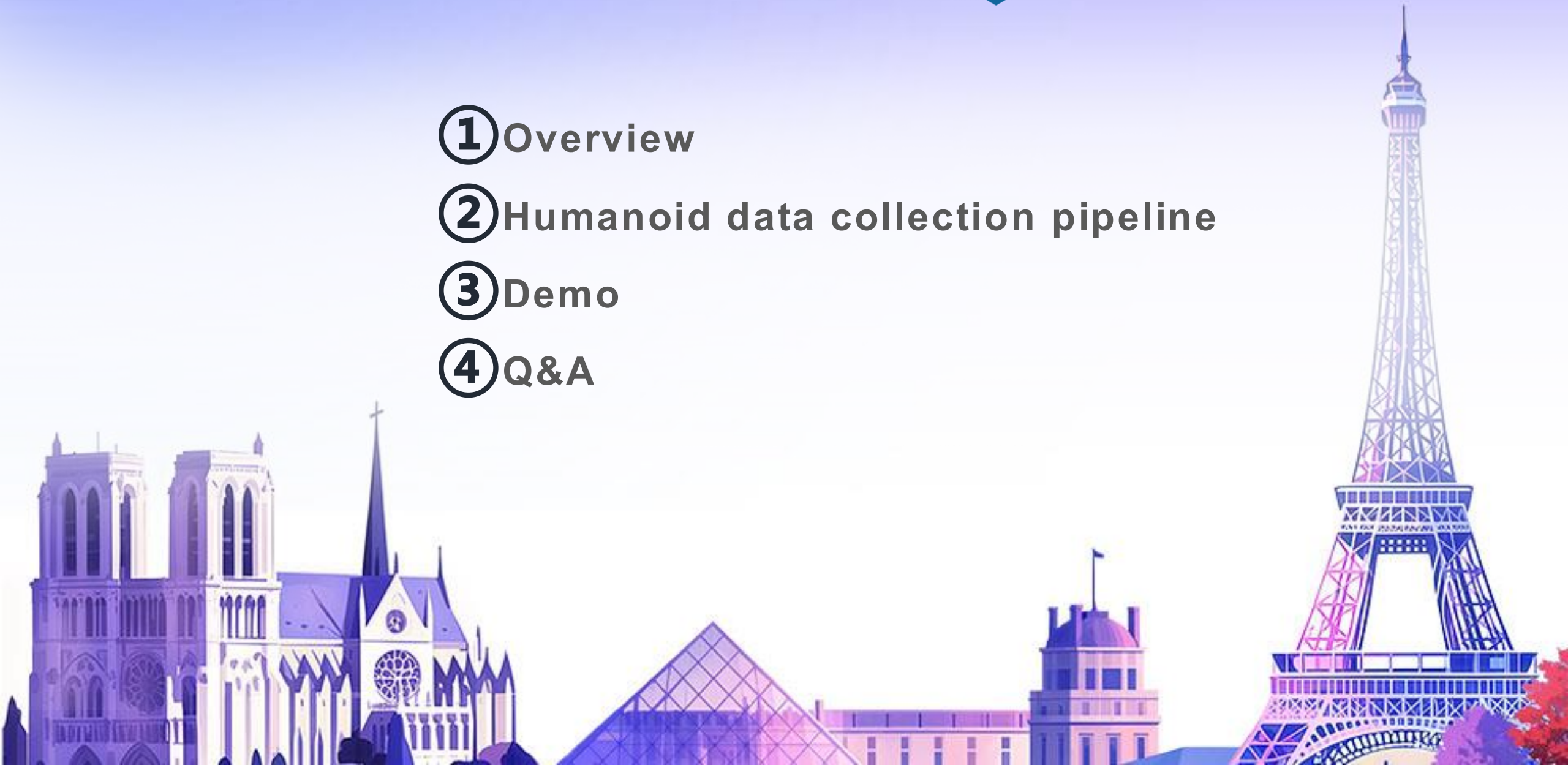
GOSIM AI Paris 2025

① Overview

② Humanoid data collection pipeline

③ Demo

④ Q&A

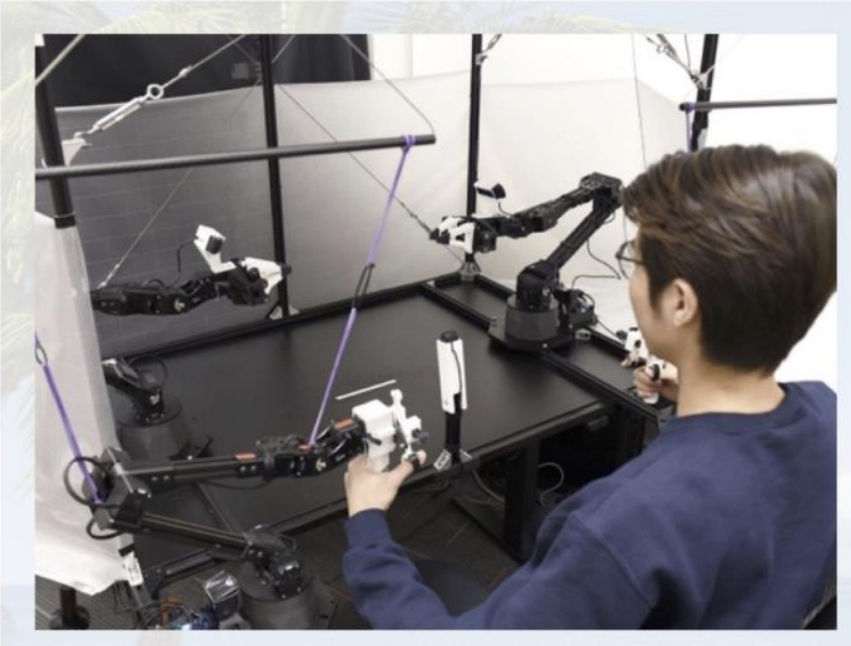


Overview

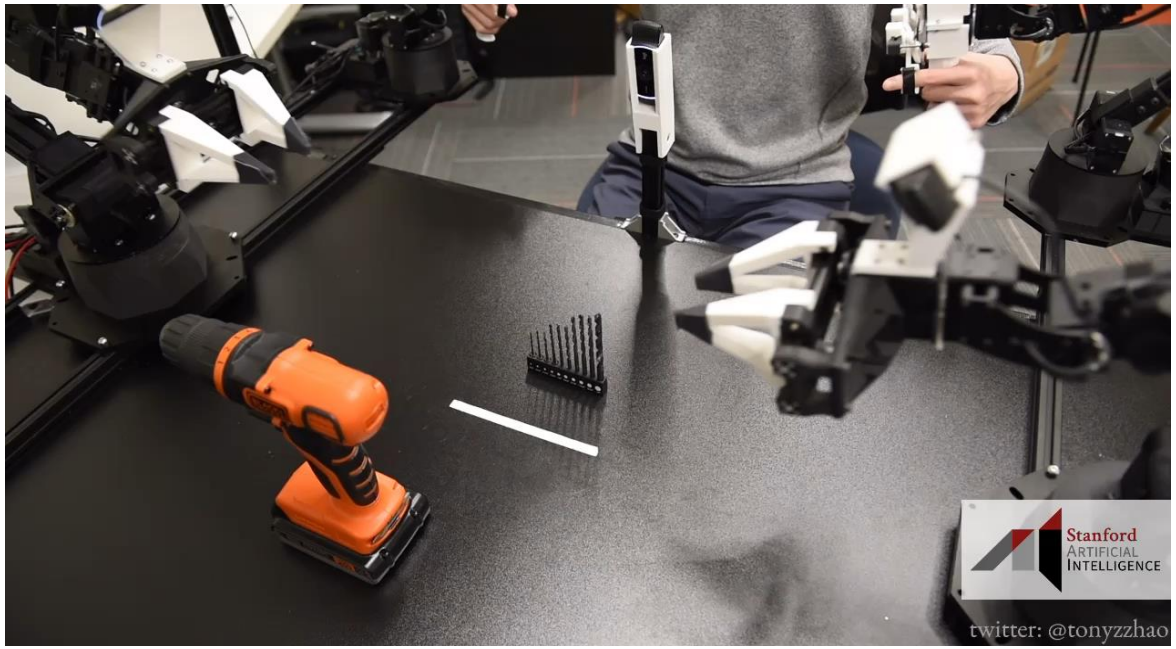
Learning from demonstration is an effective approach for robot manipulation, but how can we get large-scale robot data?

Overview

Leader-follower arms



RSS 2023: ALOHA

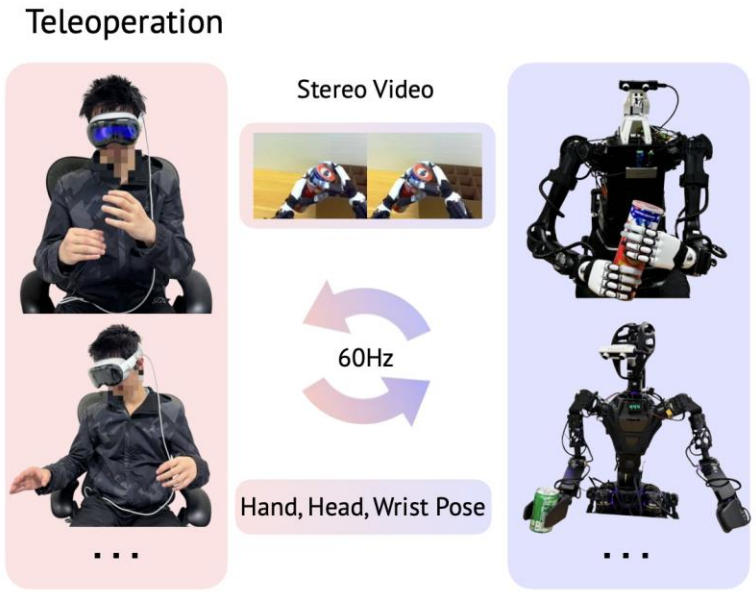


ALOHA data collection

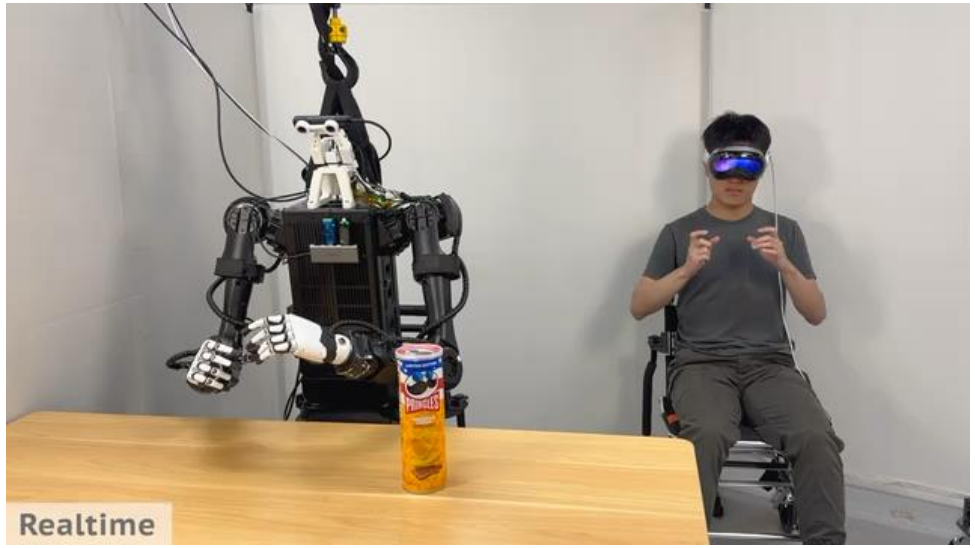
- Precise joint recording
- Mainly focus on parallel-jaw grippers
- Restricted to specific robot platforms

Overview

Vision based VR



CoRL 2024: Open-Television

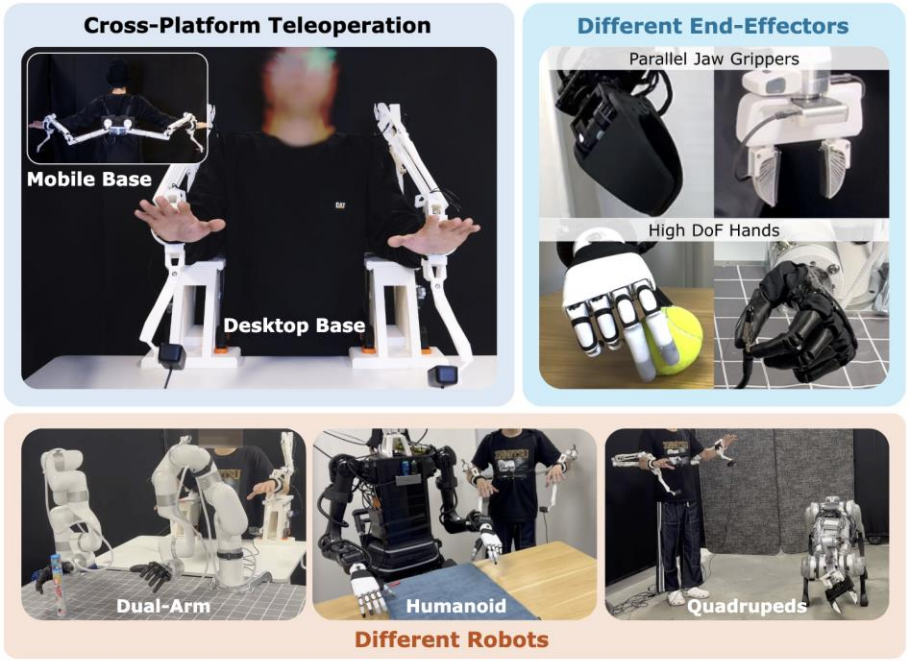


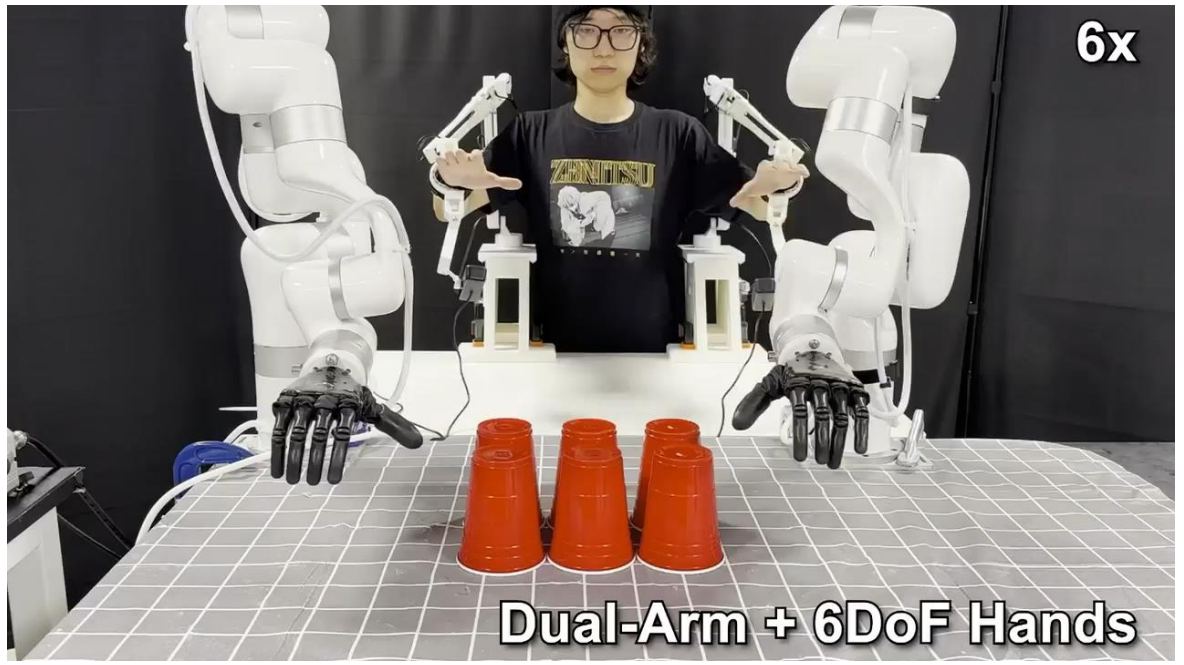
Open-Television data collection

- Wrist and fingertip tracking
- 3 dof wrist tracking
- Tiring for human teleoperators
- Inaccurate finger joints tracking under occlusion

Overview

Vision Exoskeletons teleoperation system





ACE data collection

CoRL 2024: ACE

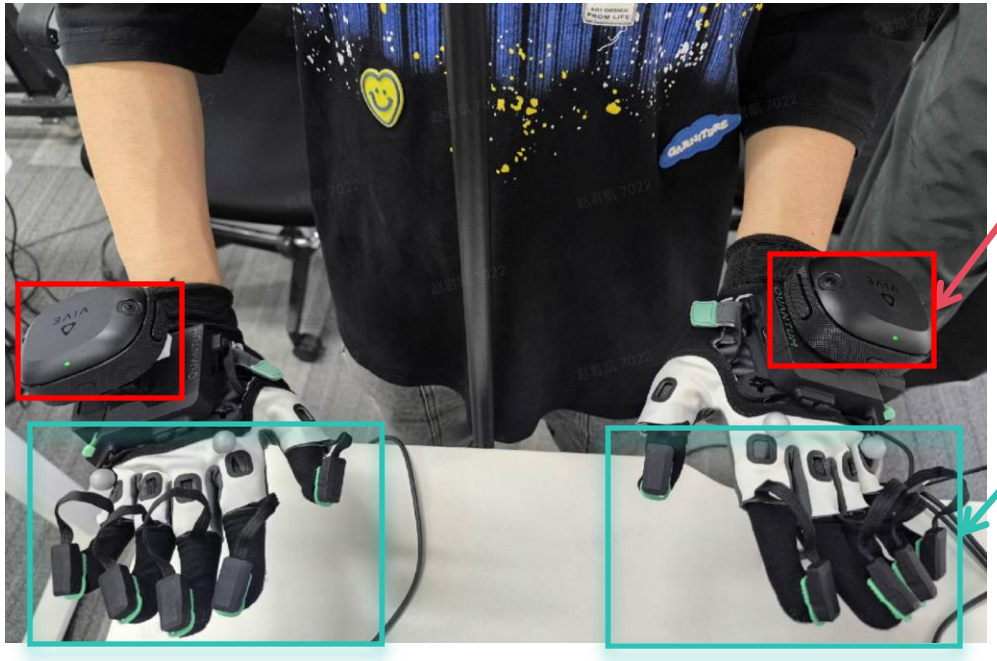
- Accurate joint(wrist) tracking
- No head tracking
- Unstable fingertip motion tracking
- Additional hardware configuration for different robot platforms

Humanoid teleoperation pipeline

How can we setup an efficient, precise, comfortable and cross embodiment robot teleoperation pipeline

Humanoid teleoperation pipeline

Head



Vive tracker(wrist pose motion tracking)

Manus Metagloves(fingertip motion tracking)

- No external tracking devices and skeleton needed, portable
- Precise hand and fingertip tracking under occlusion
- Comfortable for operators
- Efficient
- Cross platform

Humanoid teleoperation pipeline

Human operator

Manus SDK

Human hand joints

Fingertip retargeting

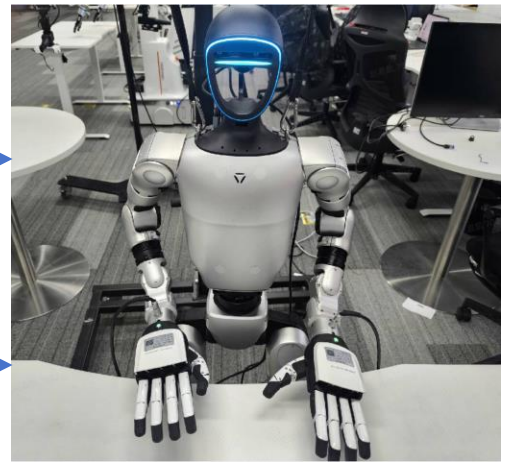
Joint control

Vive tracker SDK

Human wrist pose

Wrist pose retargeting

IK



Unitree G1 with inspire hands

Robot wrist pose

Dex-hand joints

Head camera stream

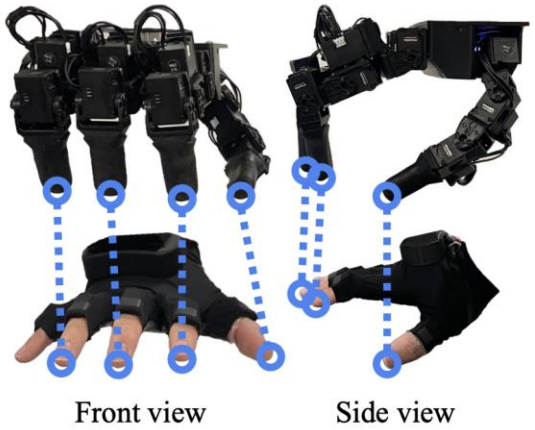
Data collection

Policy training&deployment

Humanoid teleoperation pipeline

Fingertip retargeting

- Size difference between human hand and robot hand.
- It's hard to directly transfer human hand motions to robot hand motions.
- Ensure the same motion between human hand fingertips and robot hand fingertips.



RSS 2024:Dexcap

$$\min_{q_t} \sum_{i=0}^{N} ||\alpha v_t^i - f_i(q_t)||^2 + \beta ||q_t - q_{t-1}||^2$$

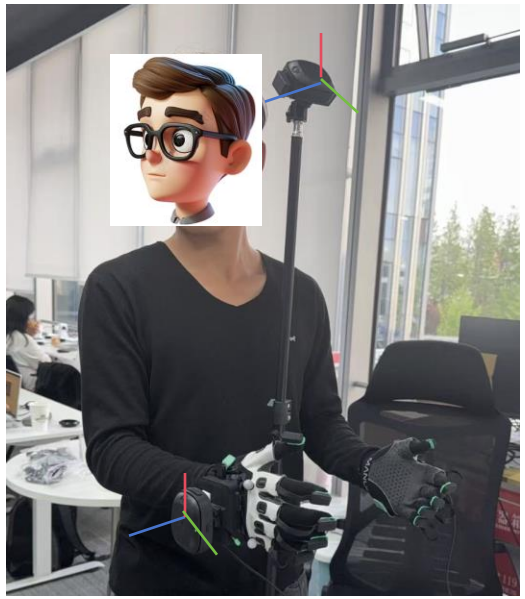
$$\text{s.t.} \quad q_l \leq q_t \leq q_u,$$

q_t: Dex-hand joint positions at timestep t
$f_i(q_t)$: Dex-hand forward kinematics
v_t^i: the i-th keypoint vector from the detected finger keypoints
q_l, q_u: lower and upper dex-hand joint limits

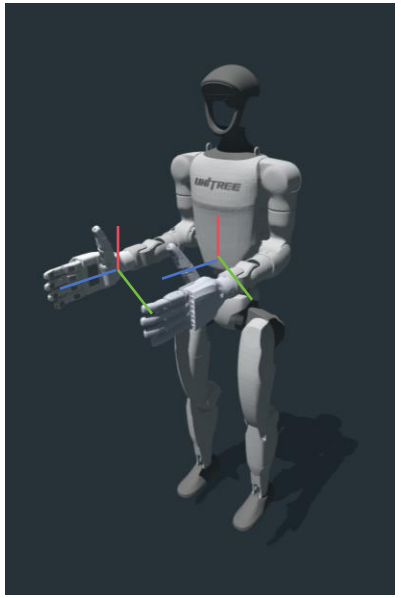
Humanoid teleoperation pipeline

Wrist pose retargeting

- Same wrist motion between human and robot
- After retargeting, use inverse kinematics(IK) to get arm joints
- In-hand tracker ensures smooth and precise wrist pose tracking

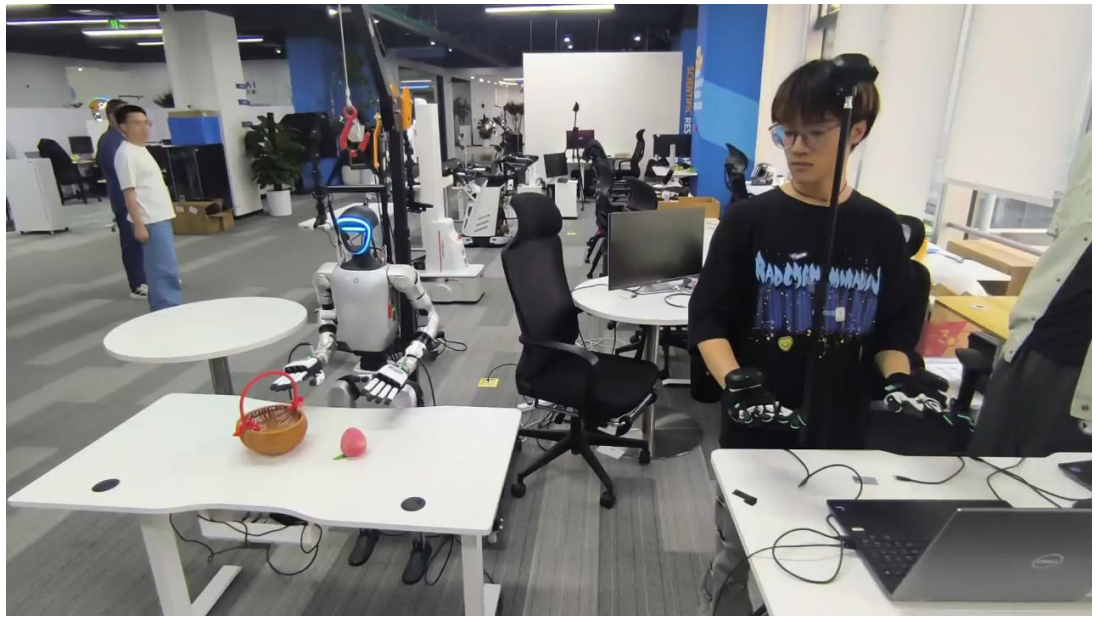


Human hand and head frame coordinate



G1 frame coordinate

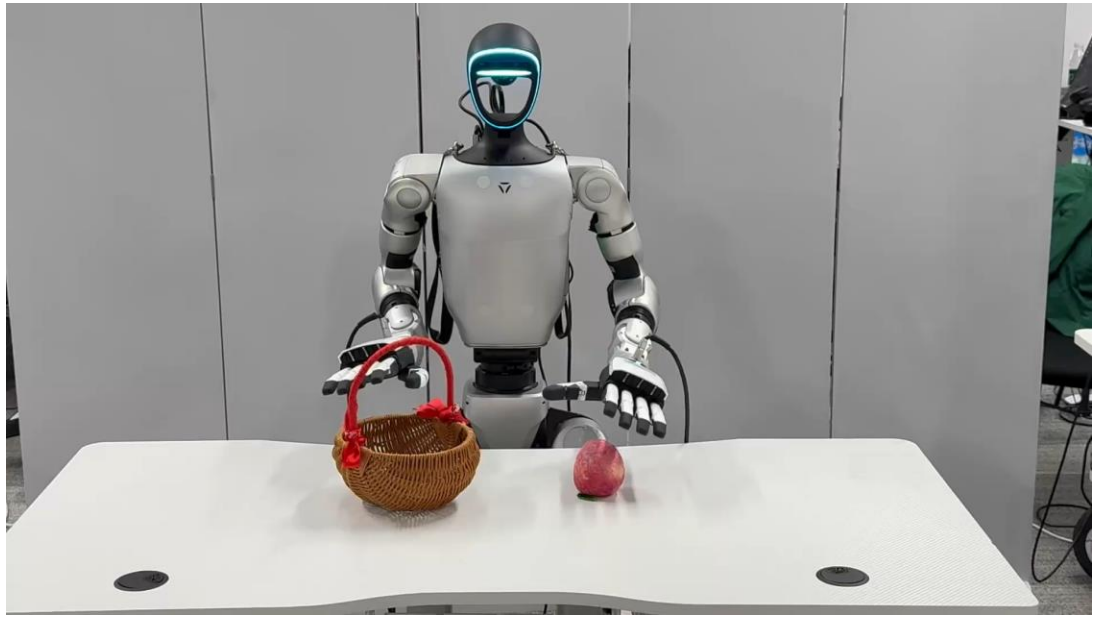
Demo humanoid data collection



Teleoperation

- Precise fingertip and wrist position tracking
- 60 hz of robot arm joints and wrist poses collection
- Rgb head camera image collection

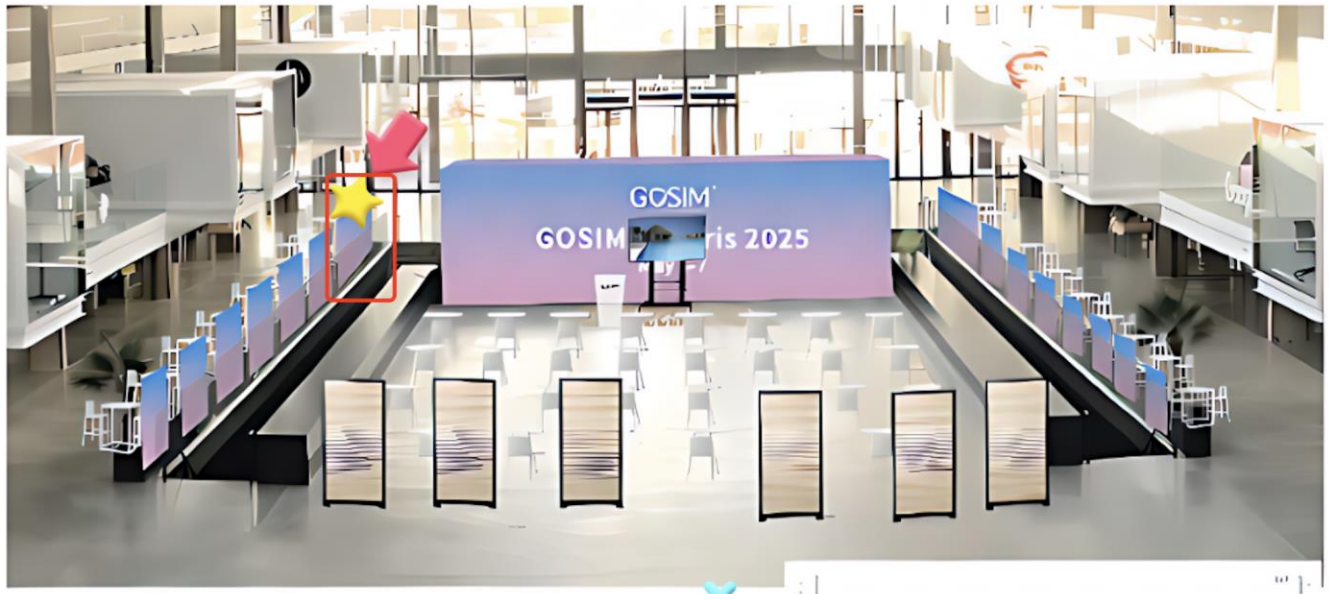
Demo Policy rollout



Policy deployment

- Precise data ensures stable robot policy
- Policy rollout frequency of 30hz
- Integration to VLM to solve long-horizon tasks

Welcome to our Booth!

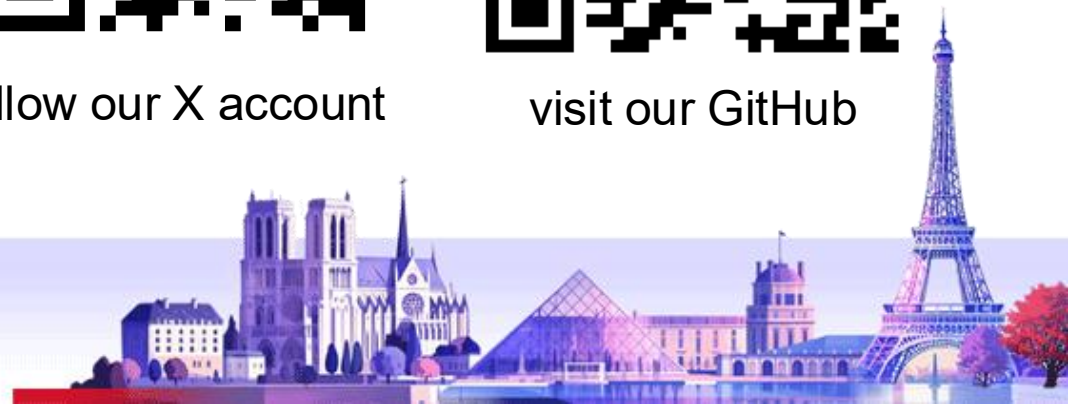


1st Floor, Open Platform Area
Entering the main gate, the first booth on the right side (next to the GOSIM main display board)

Visit us at **BAAI Booth** (with a shining Star in the left picture)

follow our X account

visit our GitHub



GOSIM AI Paris 2025

GOSIM

GOSIM AI Paris 2025

THANK YOU