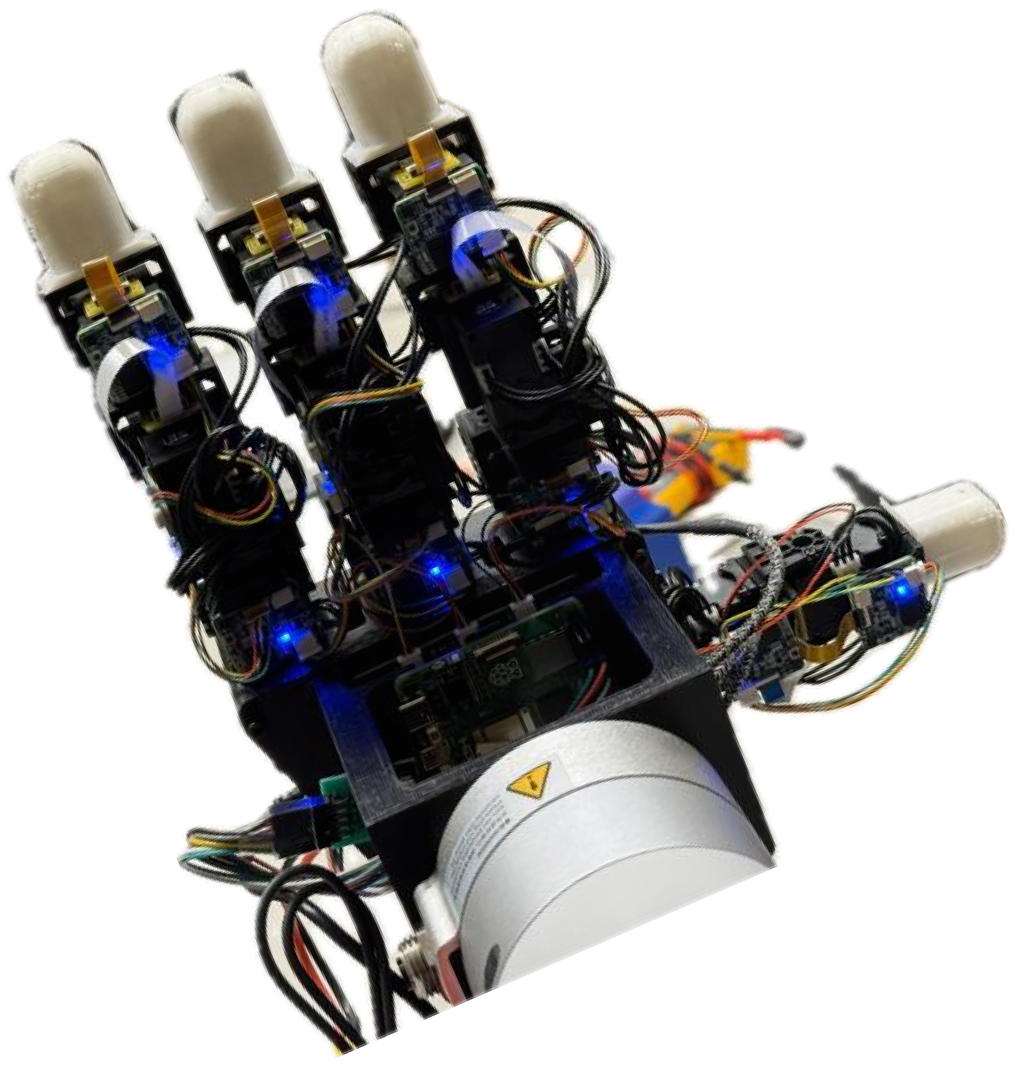
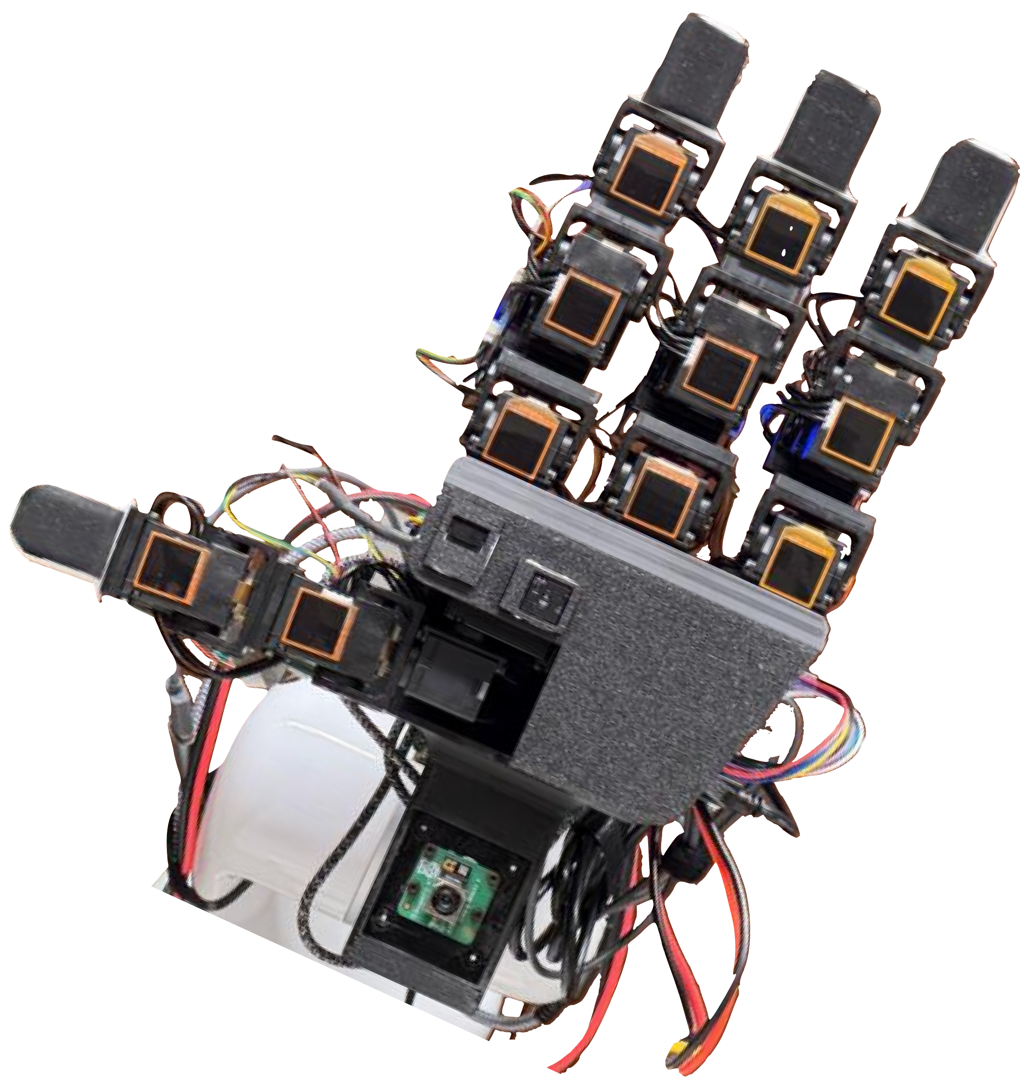
USC Viterbi School of Engineering

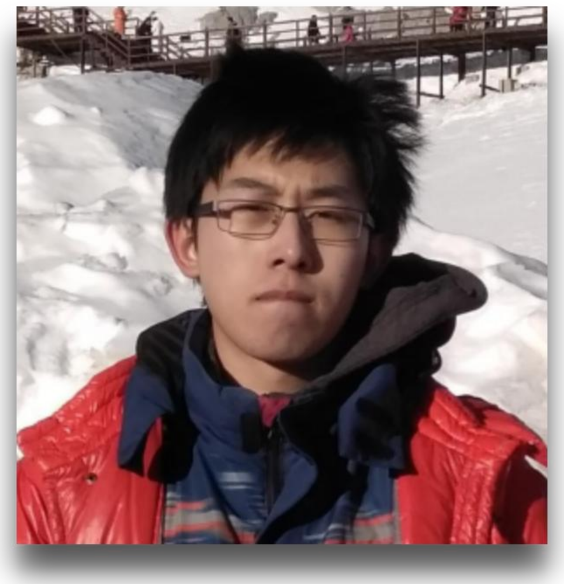
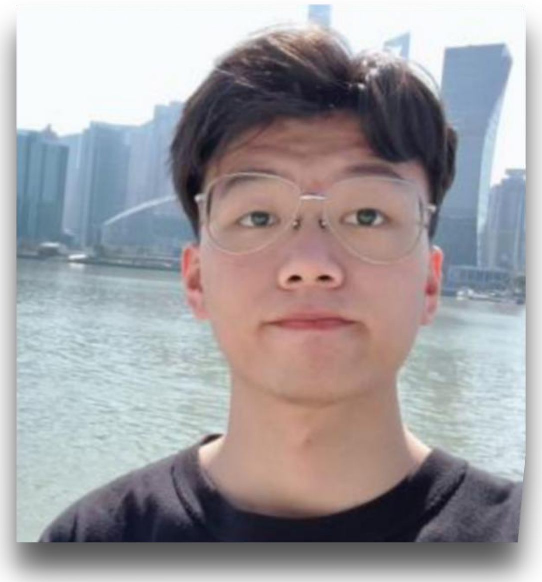
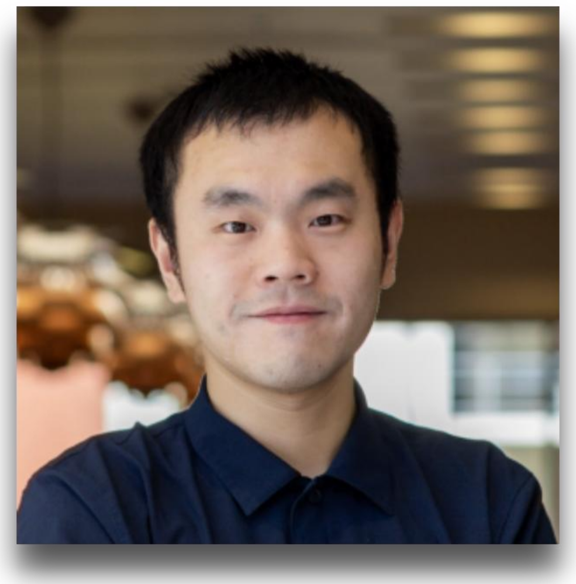
MOTIF Hand





A Robotic Hand for Multimodal Observations with Thermal, Inertial and Force Sensors.



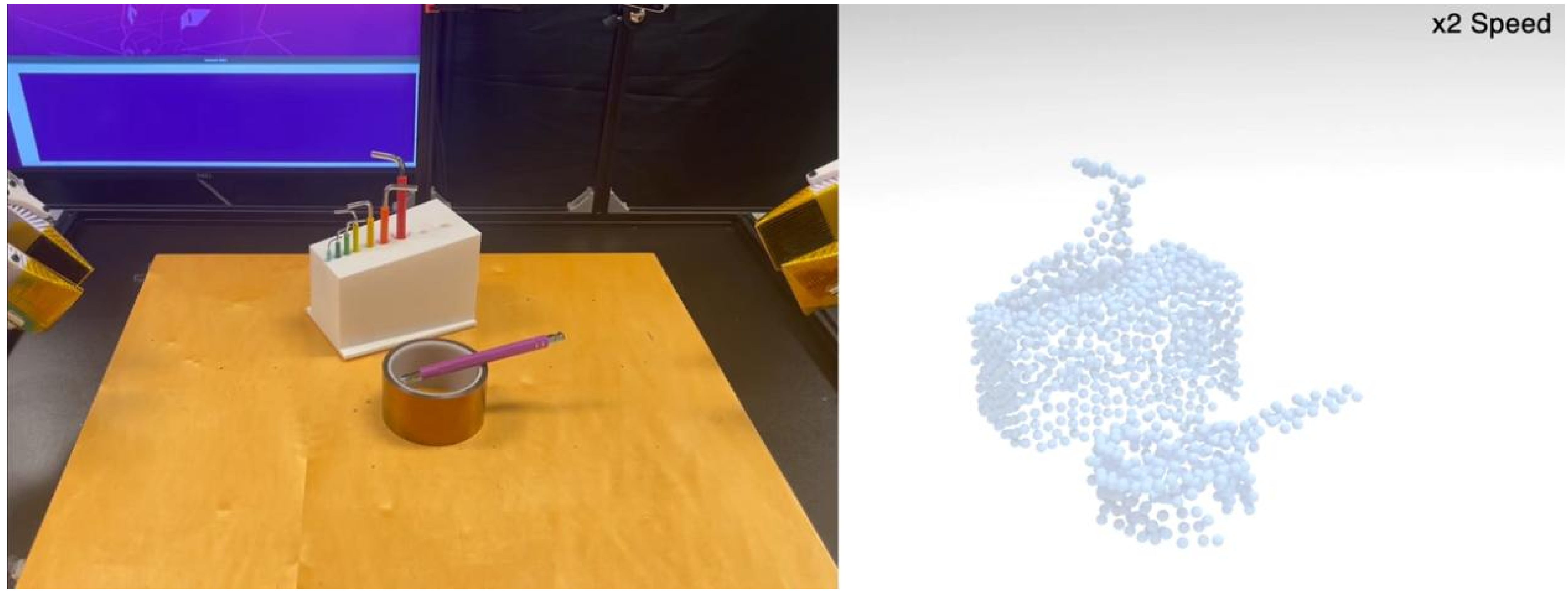

Hanyang Zhou*,[1], Haozhe Lou*, Wenhao Liu*, Enyu Zhao, Yue Wang, and Daniel Seita

*Equal Contribution
[1] Applying to Ph.D

Research Background

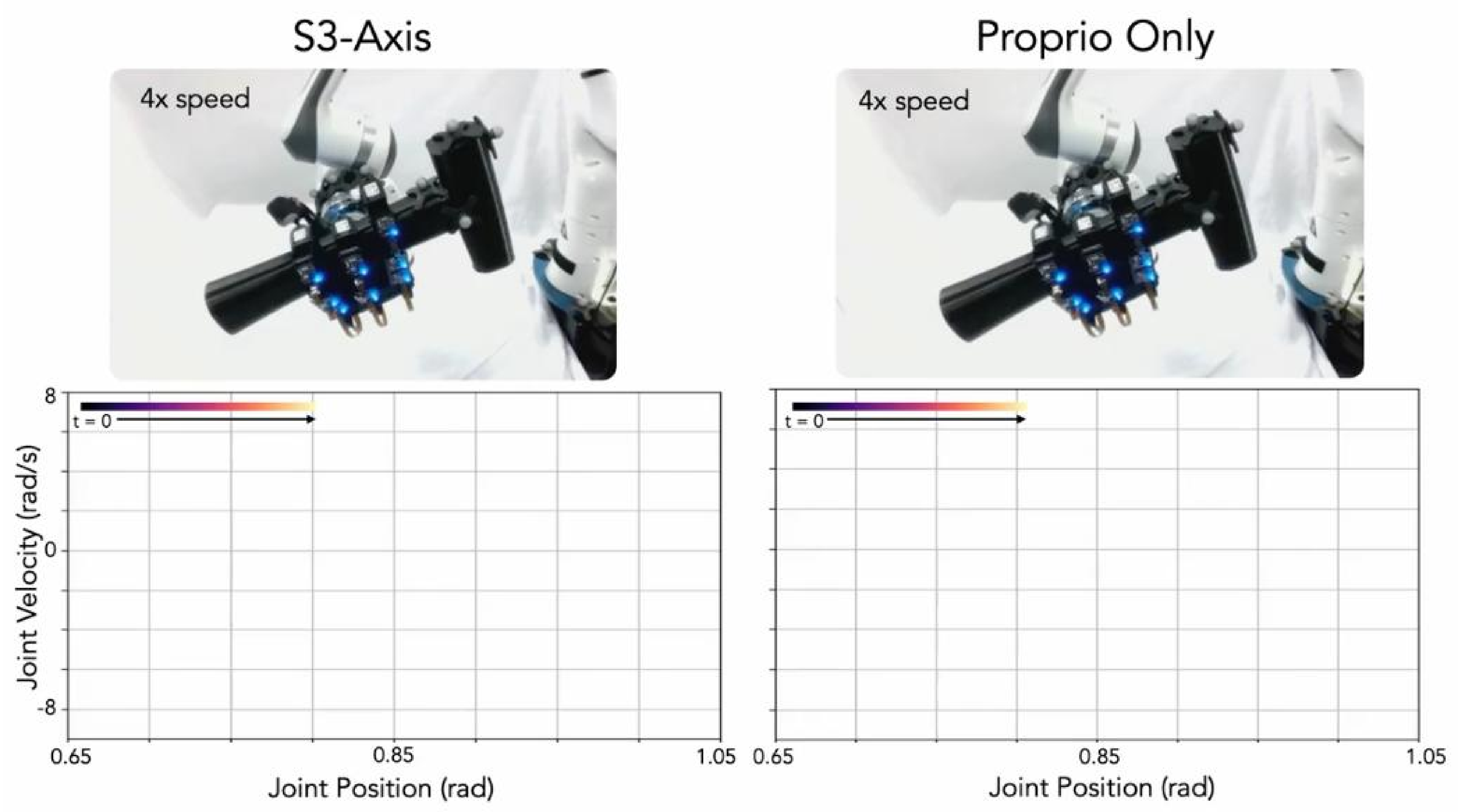
- Recent research has explored vision-based multi-modal sensing for manipulation.
- Additional sensory modalities are essential for robust real-world interaction.
- Our hand is designed around potential senses for manual tasks: **touch**, **force**, and **temperature**.

XELA uSkin Sensors



3D-ViTac

Huang, B. et al., 3D-ViTac: Learning Fine-Grained Manipulation with Visuo-Tactile Sensing. CoRL 2024.

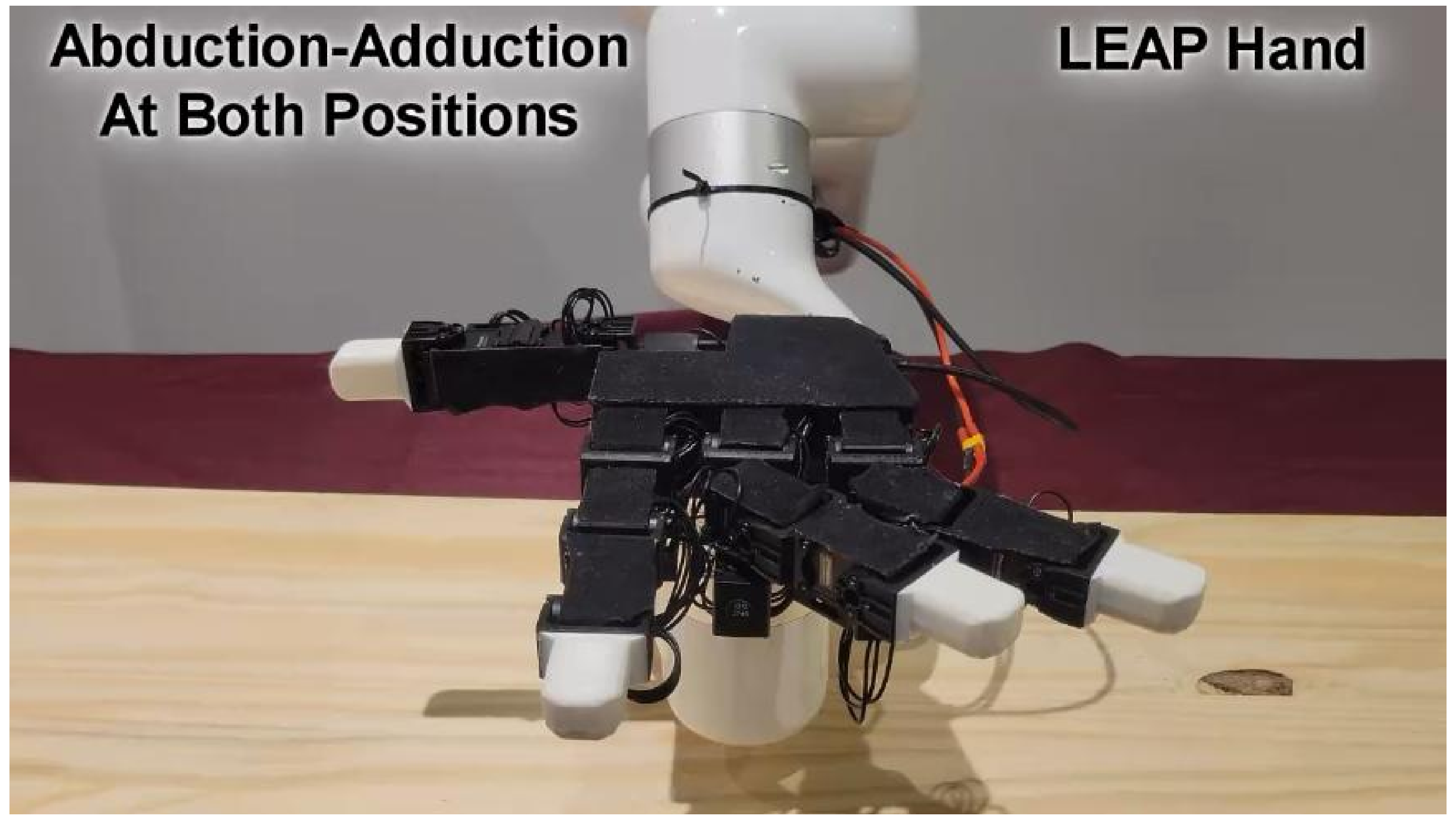


Meta Digit Plexus

Lambeta, M., et al. Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. RA-L (2020).

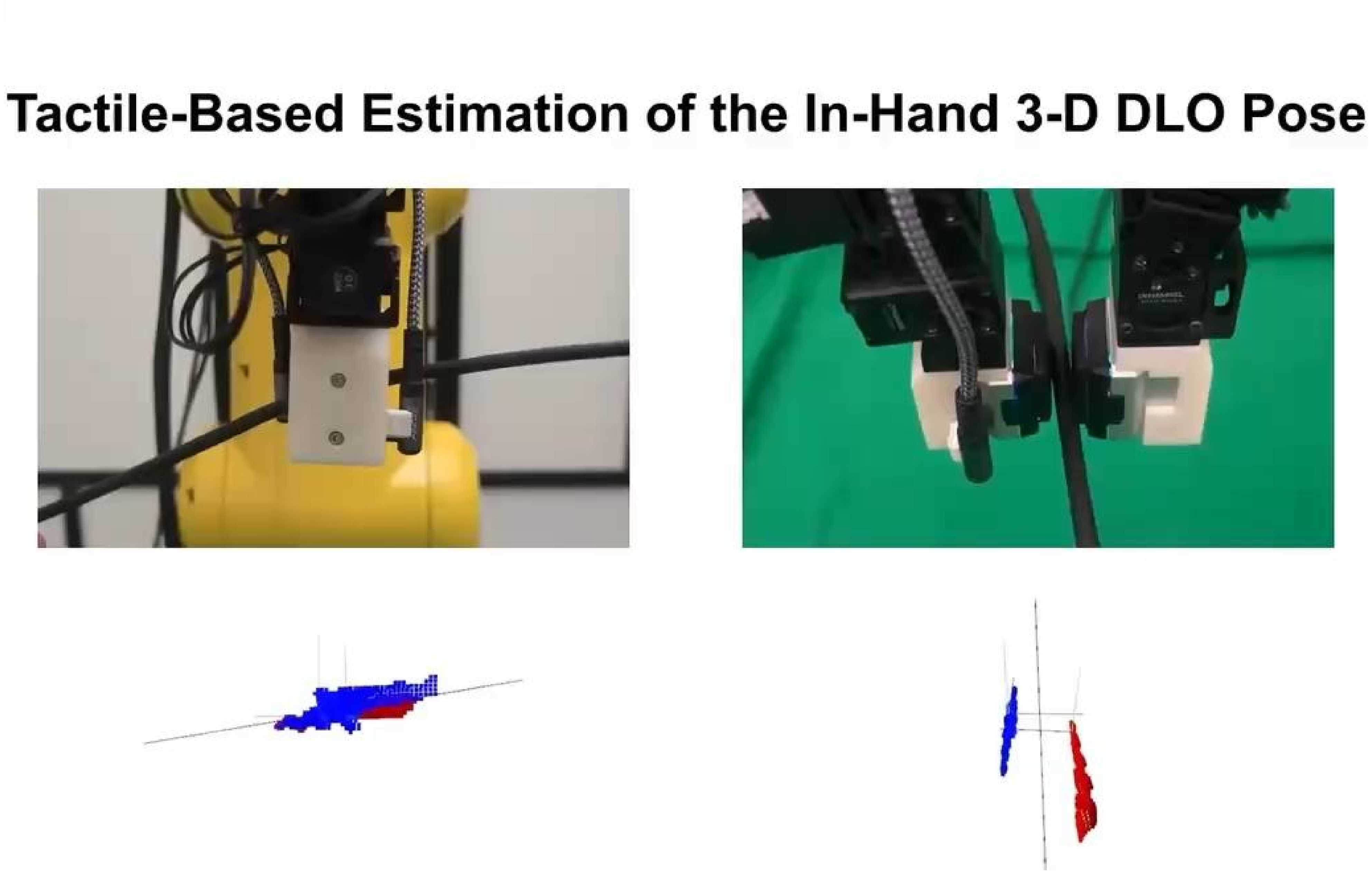
Research Background

LEAP Hand and its variations



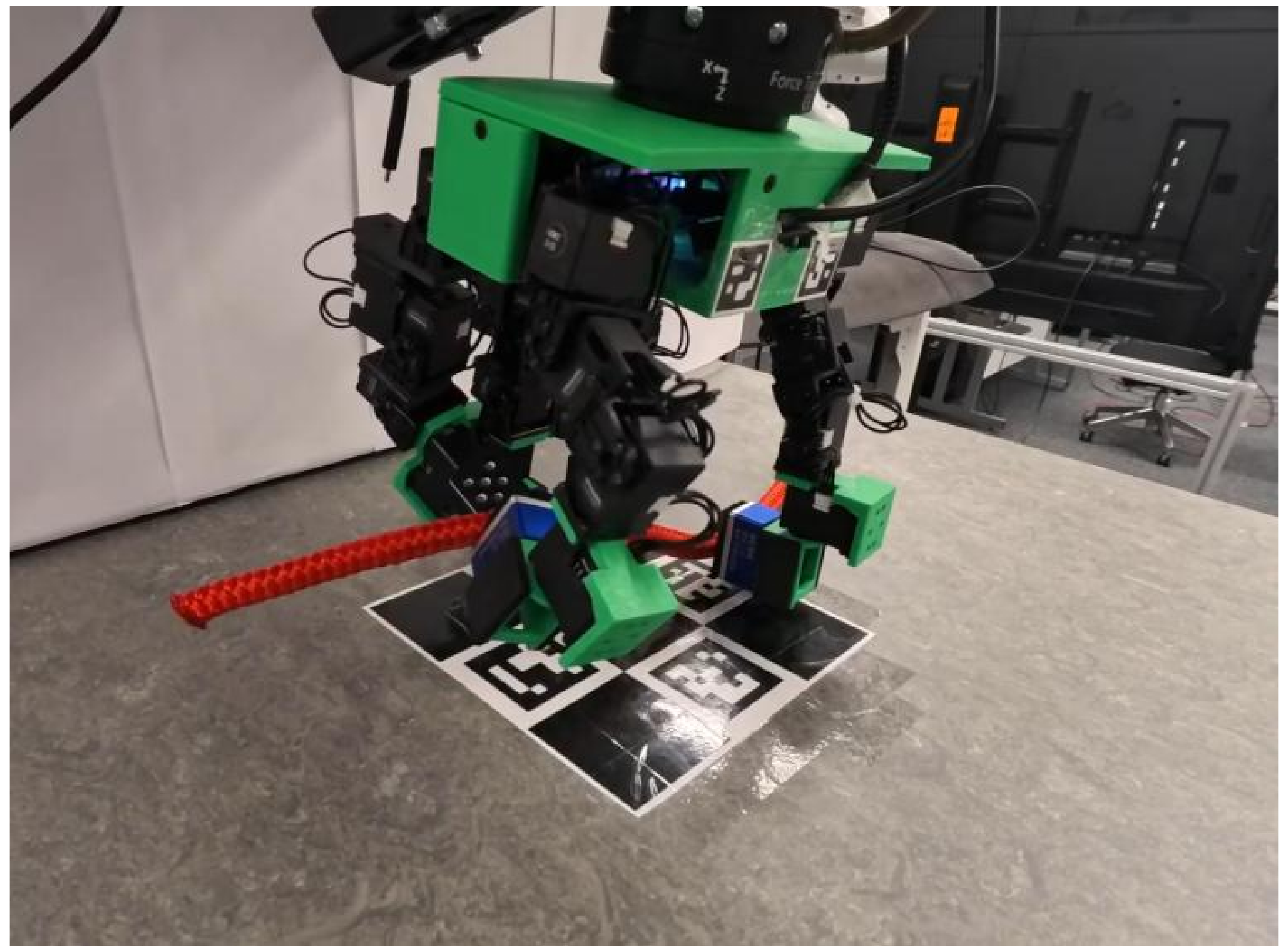
Basic LEAP Hand

Shaw, K., Agarwal, A., & Pathak, D. LEAP Hand: Low-Cost, Efficient, and Anthropomorphic Hand for Robot Learning. RSS 2023.



LEAP Hand with GelSight

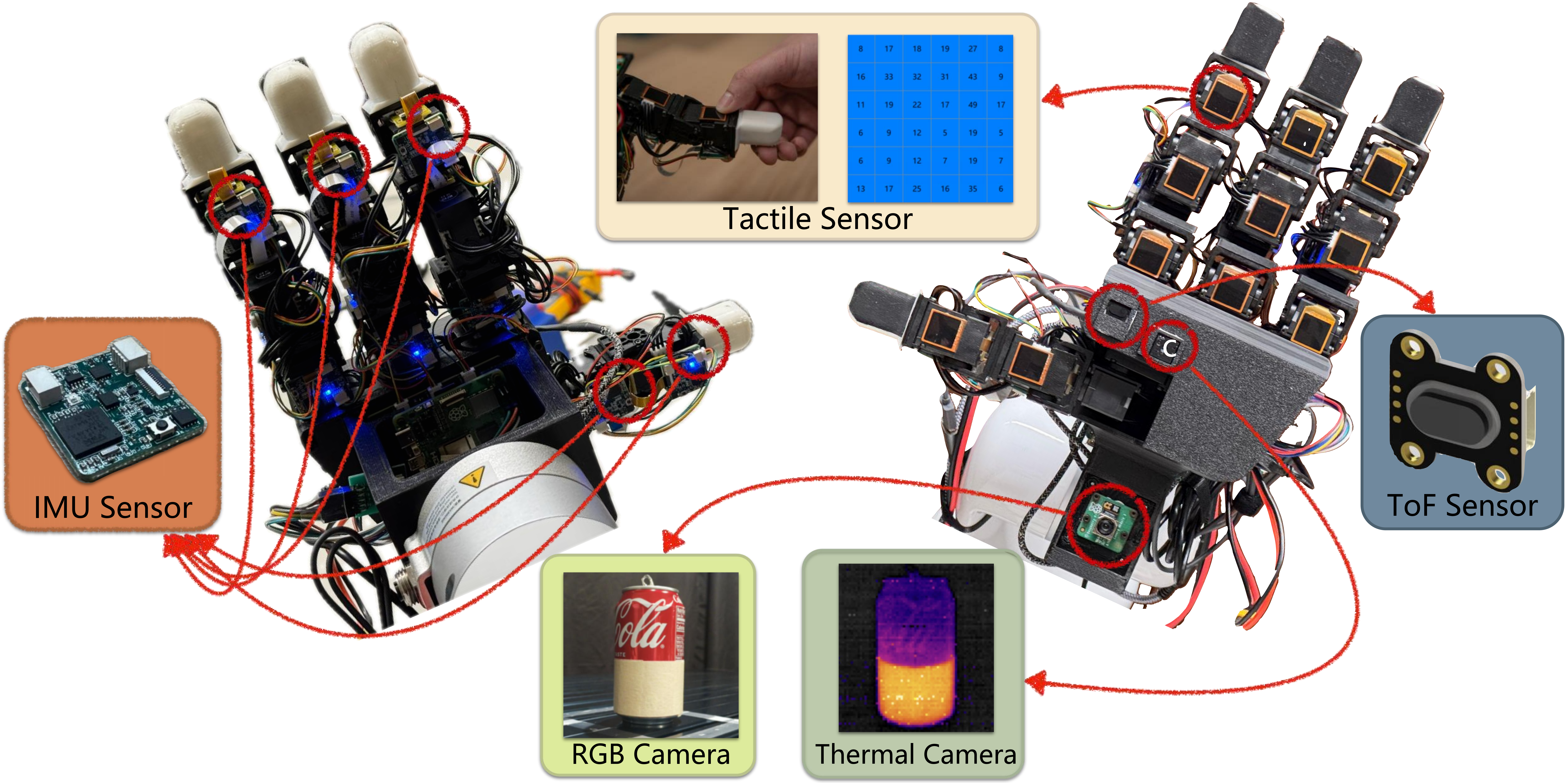
Yu, M.et al., In-hand following of deformable linear objects using dexterous fingers with tactile sensing. IROS 2024.



LEAP Hand with 2-Thumb Design

Zhaole, S., et al., R. B. (2025). Dexterous Cable Manipulation: Taxonomy, Multi-Fingered Hand Design, and Long-Horizon Manipulation. *arXiv preprint arXiv:2502.00396*.

Multi-modality Integrated Design



Design Motivation & Experimental Implementation

Temperature-Aware Manipulation via Thermal Sensing

Why do we introduce thermal sensing?

- Human thermal perception works through skin and hair follicles
  - We **detect heat before physical contact occurs**.

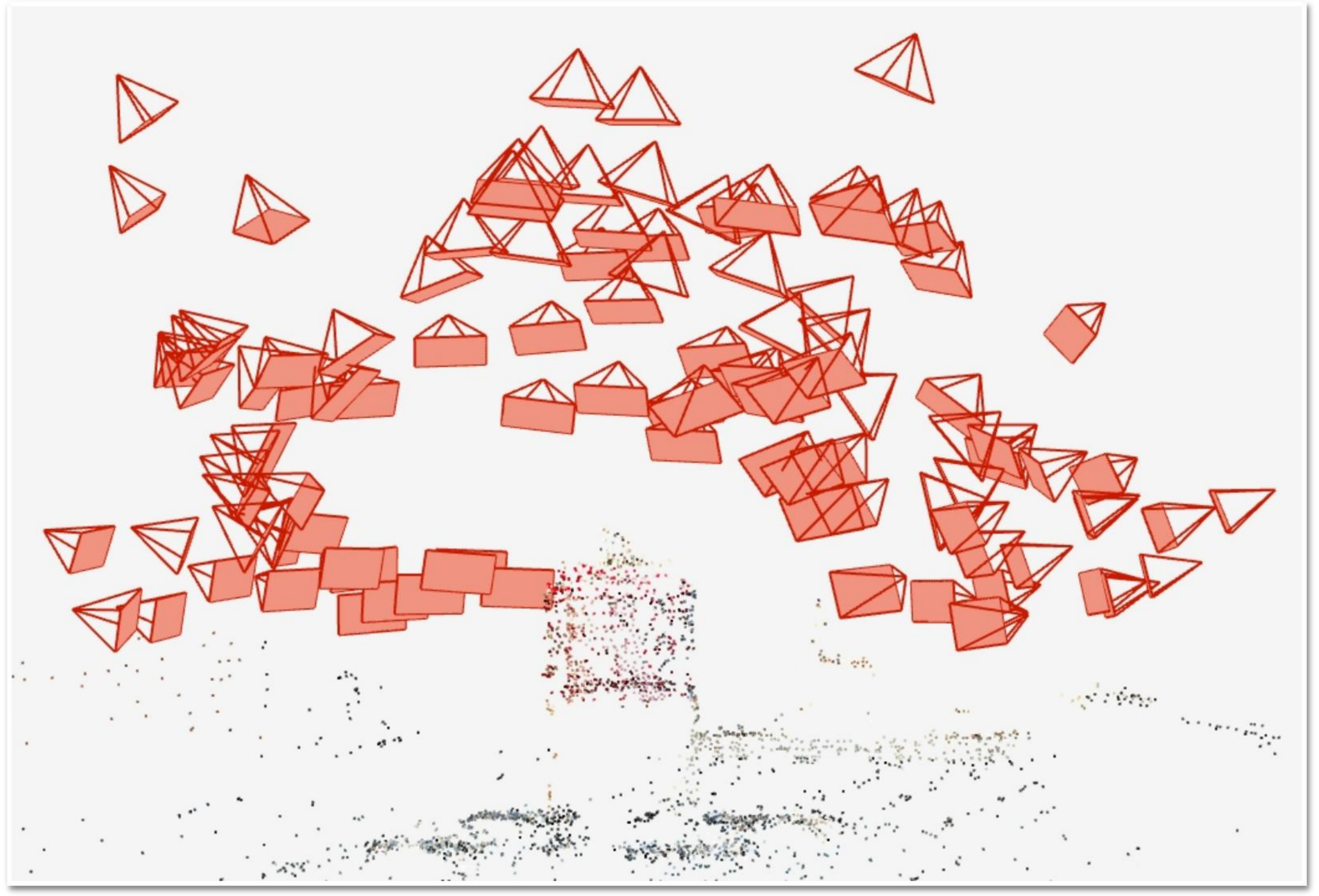
- Infrared sensing replicates this **natural non-contact capability**.
  - Palm-mounted thermal camera enable safe pre-grasp assessment.

Thermal Sensing Pipeline

Object Perception & 3D Gaussian Reconstruction

- Use Structure-from-Motion (SfM) [1], Gaussian Splatting [2] to recover camera poses and spatial representation [3] (Point Cloud, Gaussian Splat, and Mesh) from RGB camera captured images.

110 RGB Images

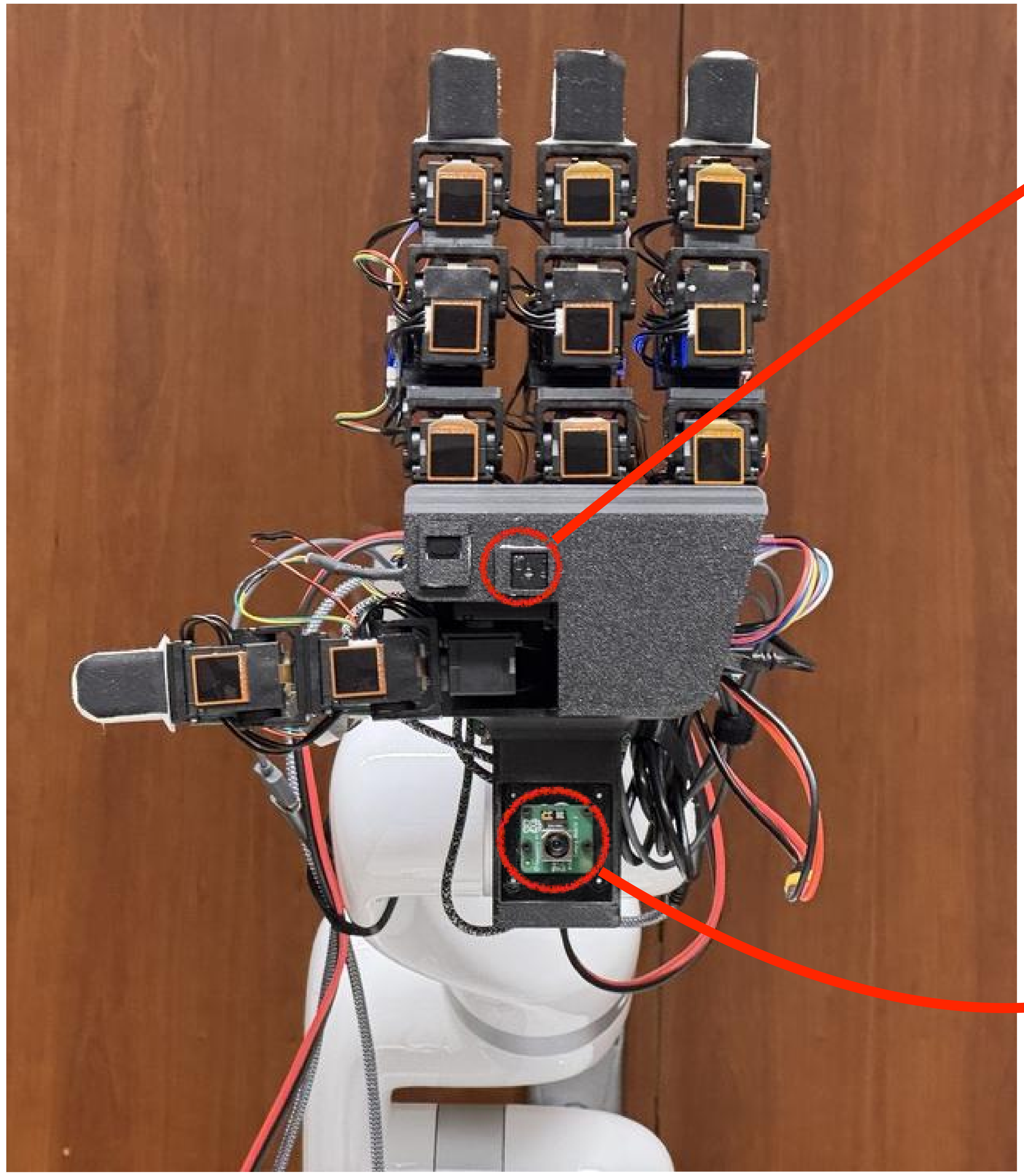


SfM Reconstructed Camera Poses in Space

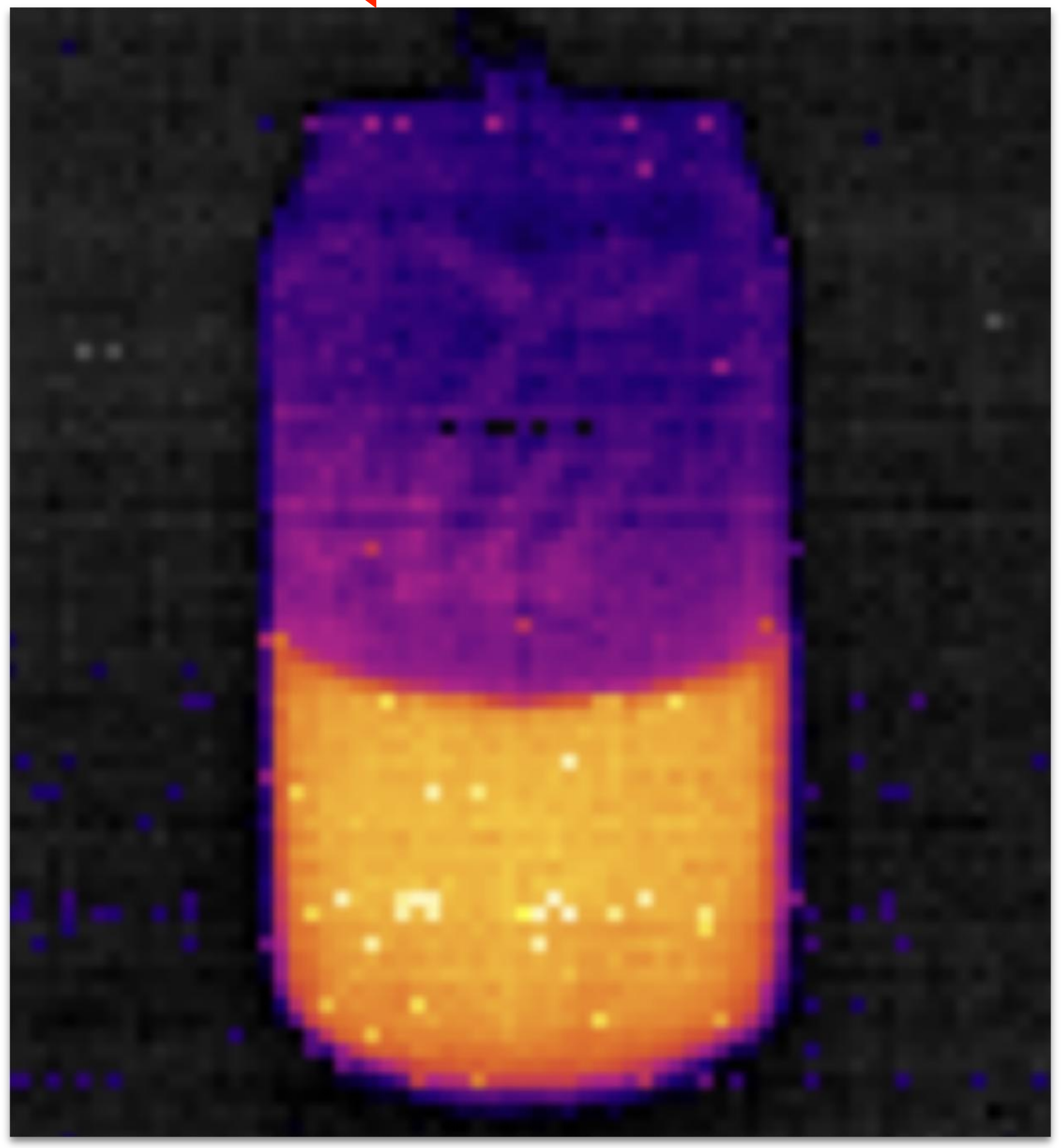
[1] J. L. Schönberger et al., "Structure-from-Motion Revisited," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[2] Kerbl, B., Kopanas et al. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, *42*(4), 139-1.
[3] Lou, H.,et. al.(2024). Robo-gs: A physics consistent spatial-temporal model for robotic arm with hybrid representation. *arXiv preprint arXiv:2408.14873*.

Thermal Sensing Pipeline

RGB & Thermal Image Capture



RGB Image



Thermal Image

- There is a discrepancy in pose and size between the thermal and RGB images.

Thermal Sensing Pipeline

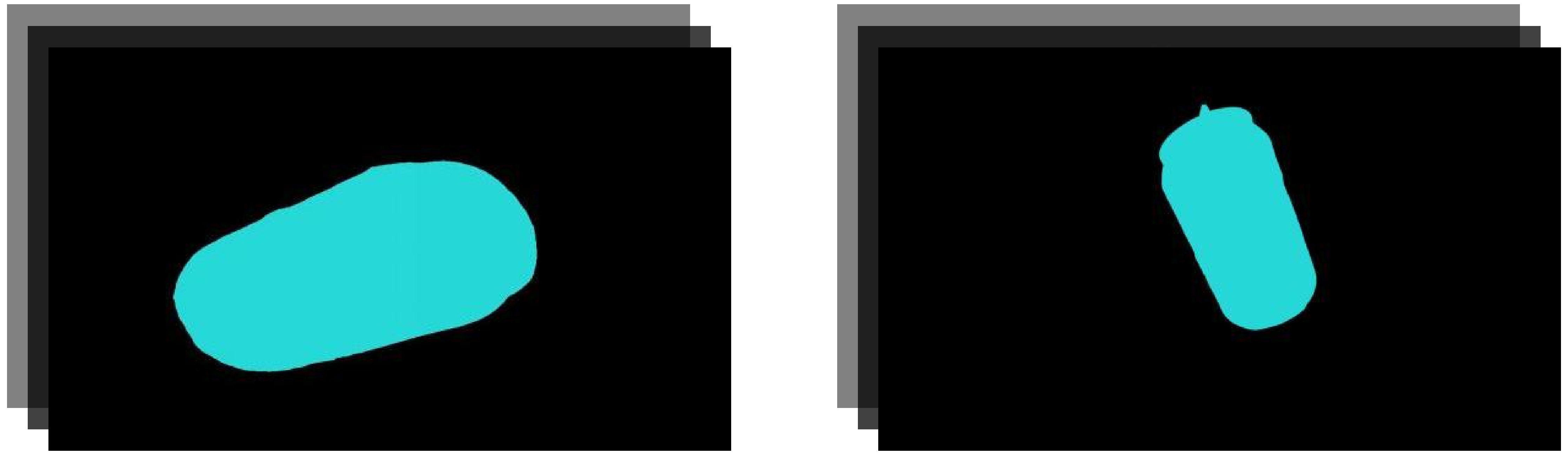
Thermal & RGB Image Alignment

- Use SAM2 to automatically extract objects from both thermal and RGB cameras.
- Apply the alignment transformation matrix between the RGB and thermal images.
- Use the Pixel-Gaussian-Mesh binding to reproject the 2D thermal information to the 3D point cloud.

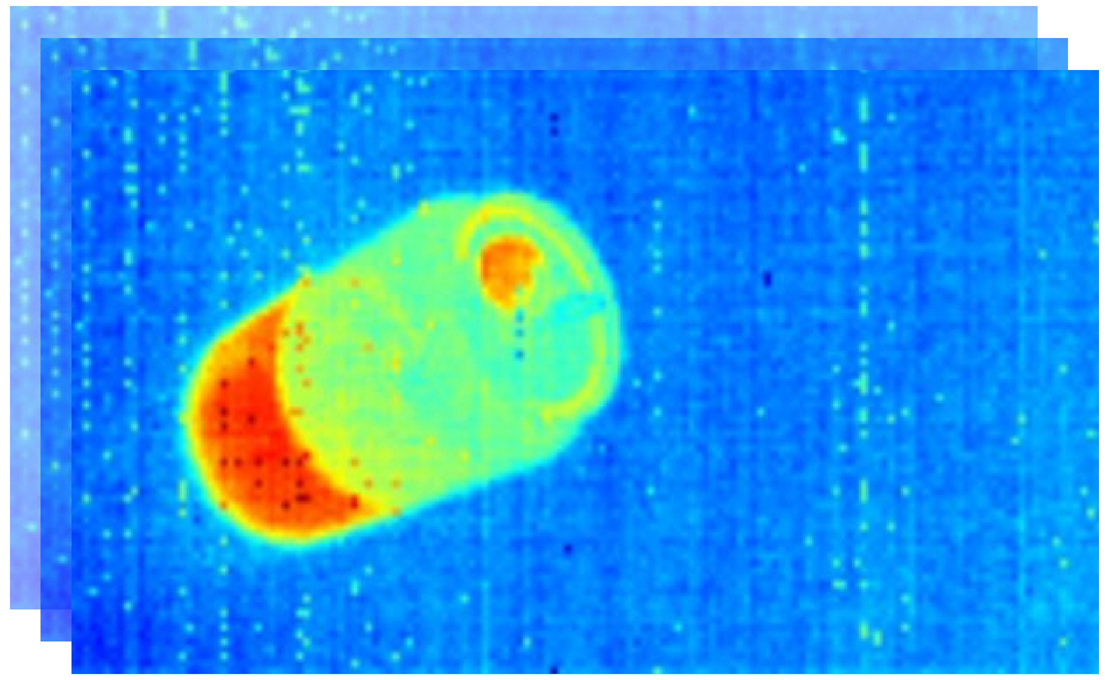
$$\mathbf{x}_{\mathbf{3D}} = {}^{w}T_c^{-1}\left(d \cdot \mathbf{K}^{-1}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}\right)$$

${}^{w}T_c \in SE(3)$: The camera pose in the world coordinate system.
d : The images' associated depth.
$\mathbf{K} \in \mathbb{R}^{3\times3}$: The camera intrinsic matrix.
$\mathbf{x}_{\mathbf{2D}} = [x, y]^T$: The pixel coordinate in images.

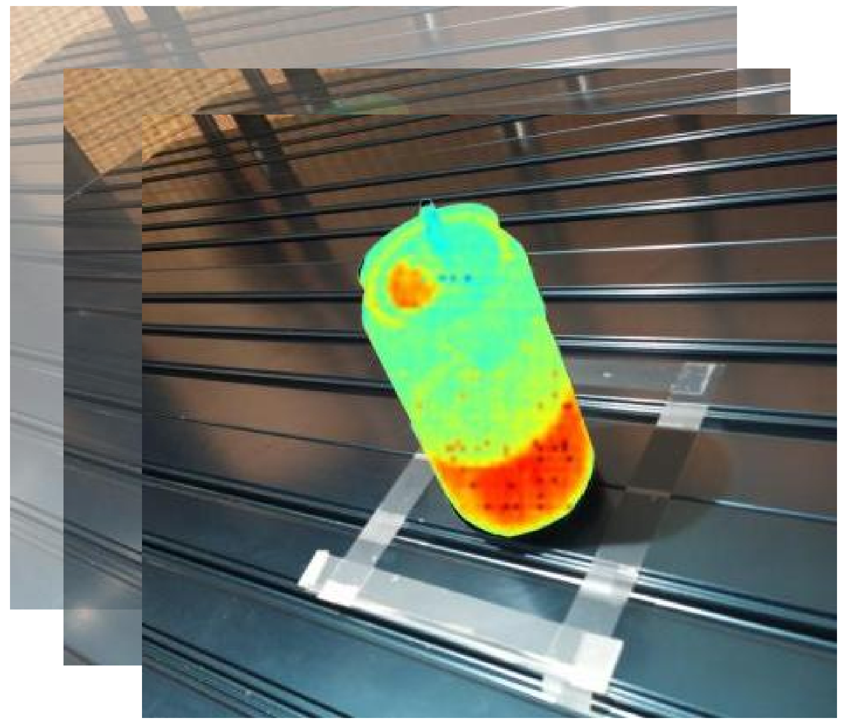
Segmentation Masks



Thermal Images



RGB + Thermal Image Registration

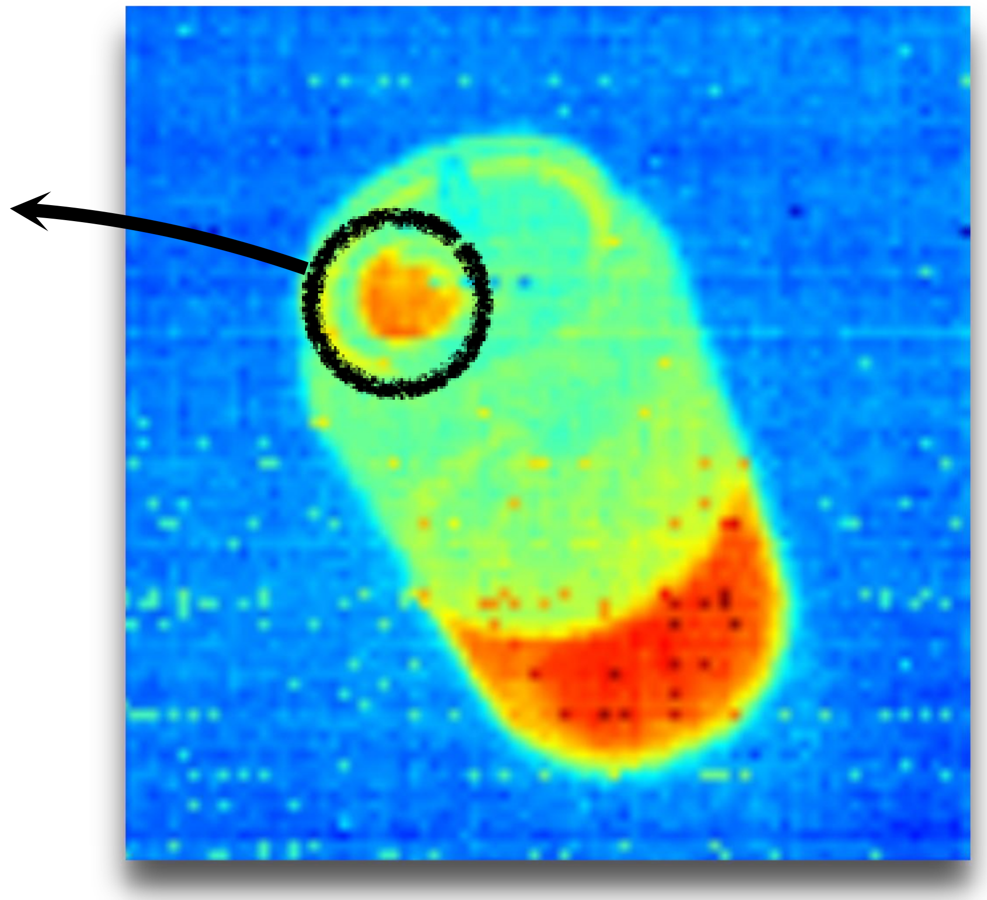


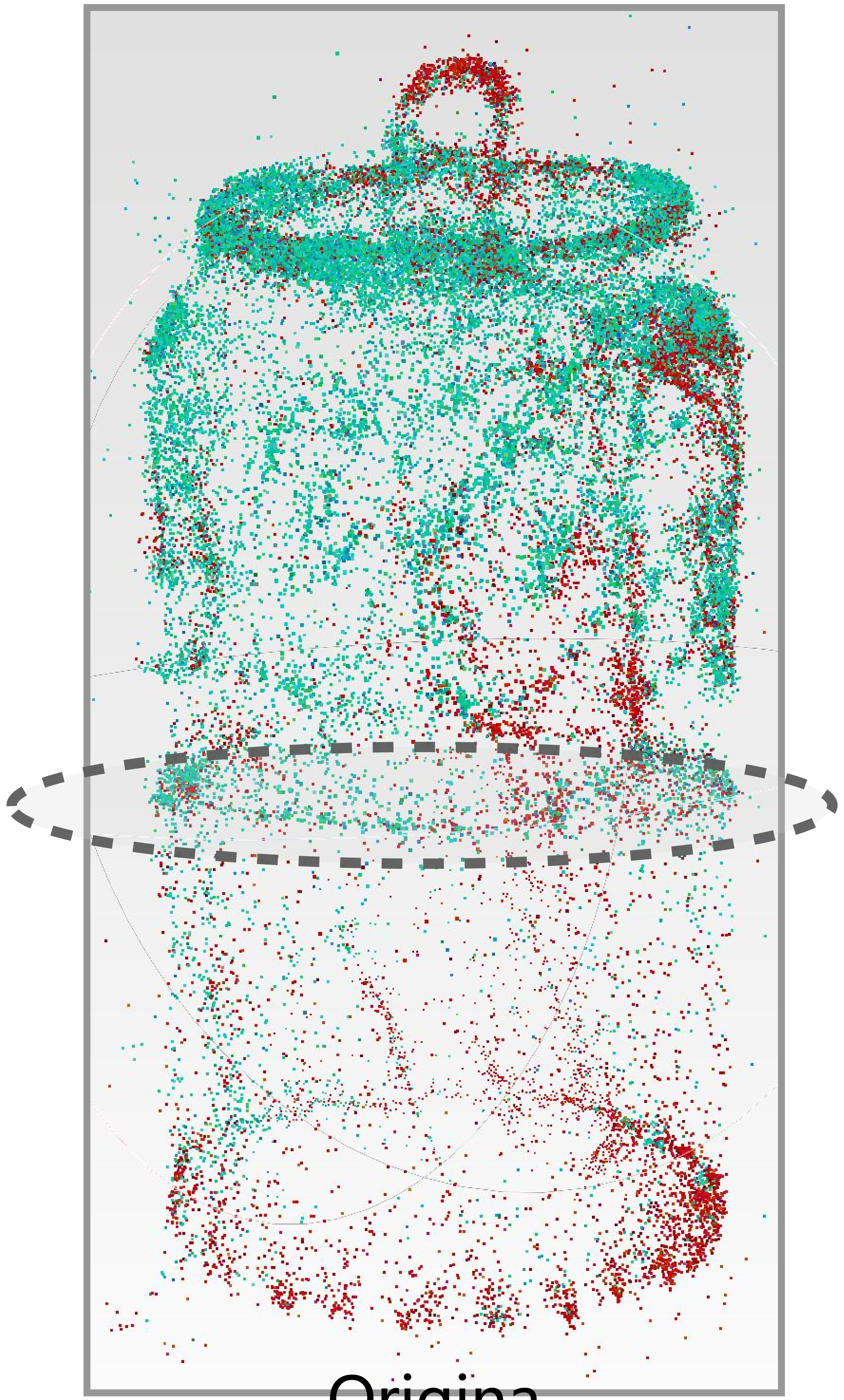
Thermal Sensing Pipeline

Thermal Point Cloud Denoising

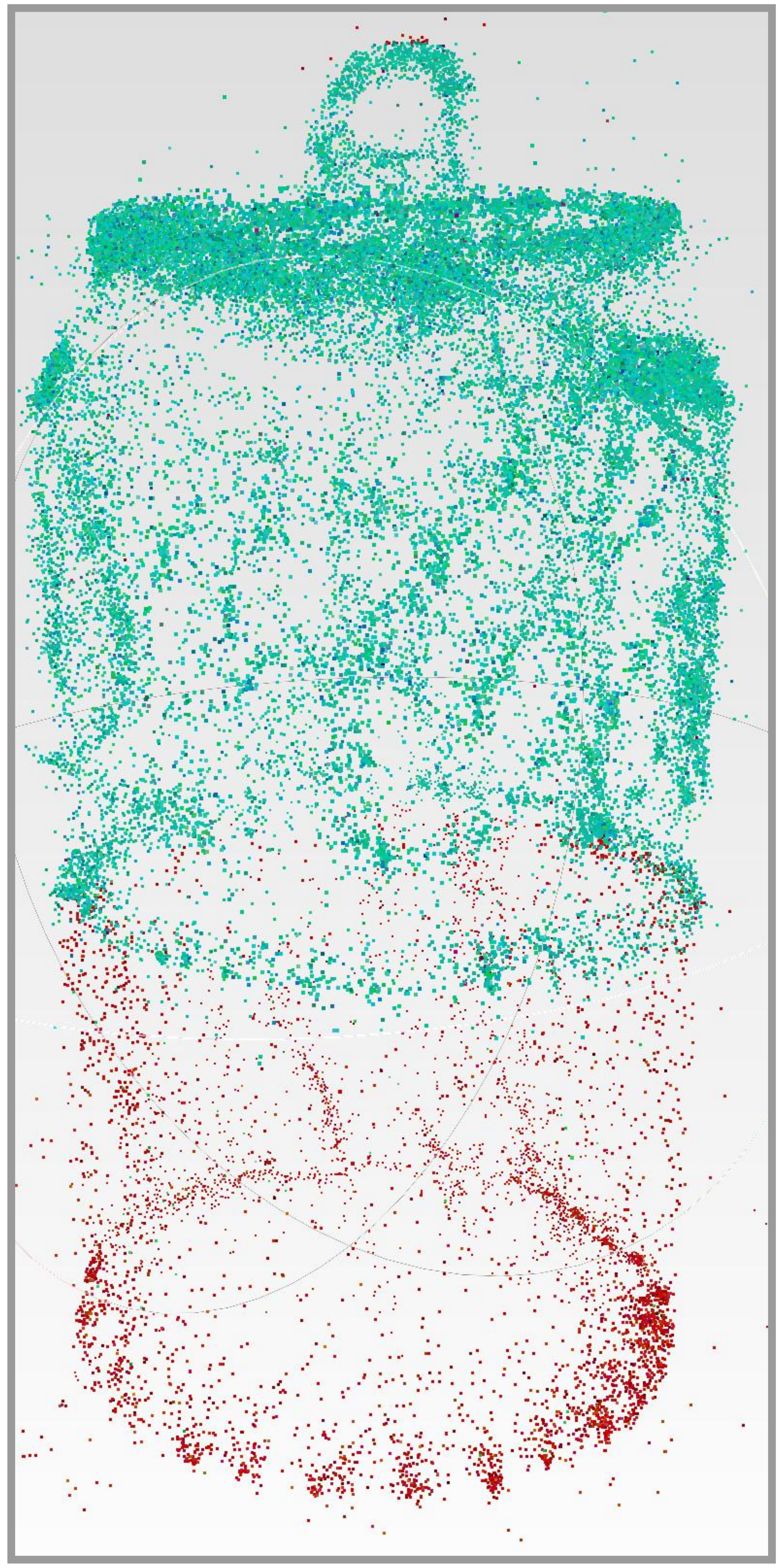
- Noise from low-resolution thermal images is inevitable.
- Denoise based on maximum negative transitions from positive thermal values (undersurface, warm) to negative values (above-surface, cool).
  - Assume a planar boundary separates liquid and air regions.

Coke can’ s top hole introduces spurious hot signals.





Origina
l

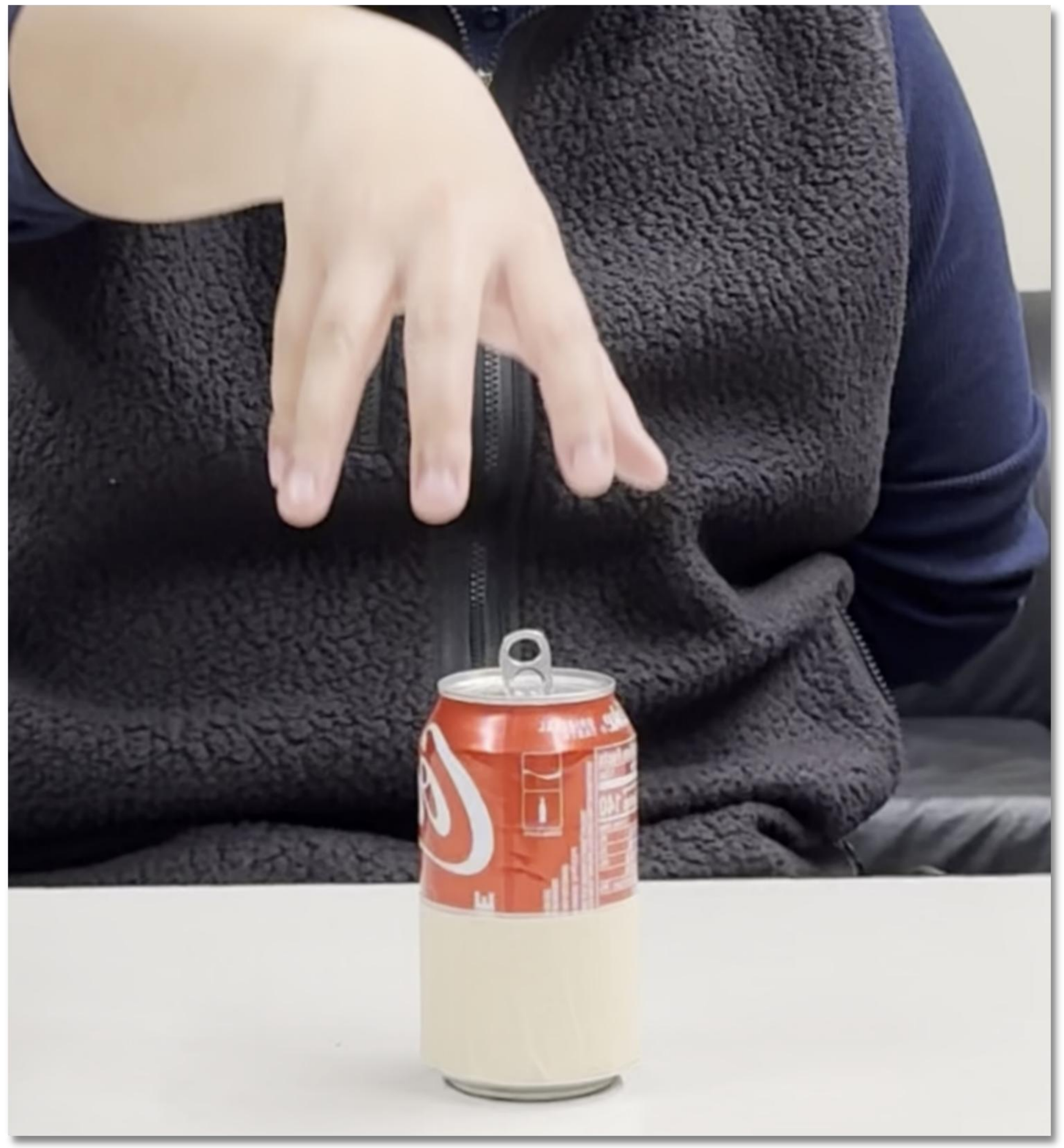


Denoised

Thermal Sensing Pipeline

Grasping Conditioned on Thermal Observations

- Learn from human demonstrations using imitation learning and checking whether the contact points fall into the high-temperature zone.



Why do we add Finger-Tip Acceleration Sensing?

- Dense tactile coverage **across** all finger surfaces is challenging
- Distinguish between different weights and textures
  - When tactile sensors cannot cover the entire finger surface
  - Contact forces can be inferred through: $\vec{F} = m\vec{a}$
  - IMU sensors provide acceleration data for force estimation

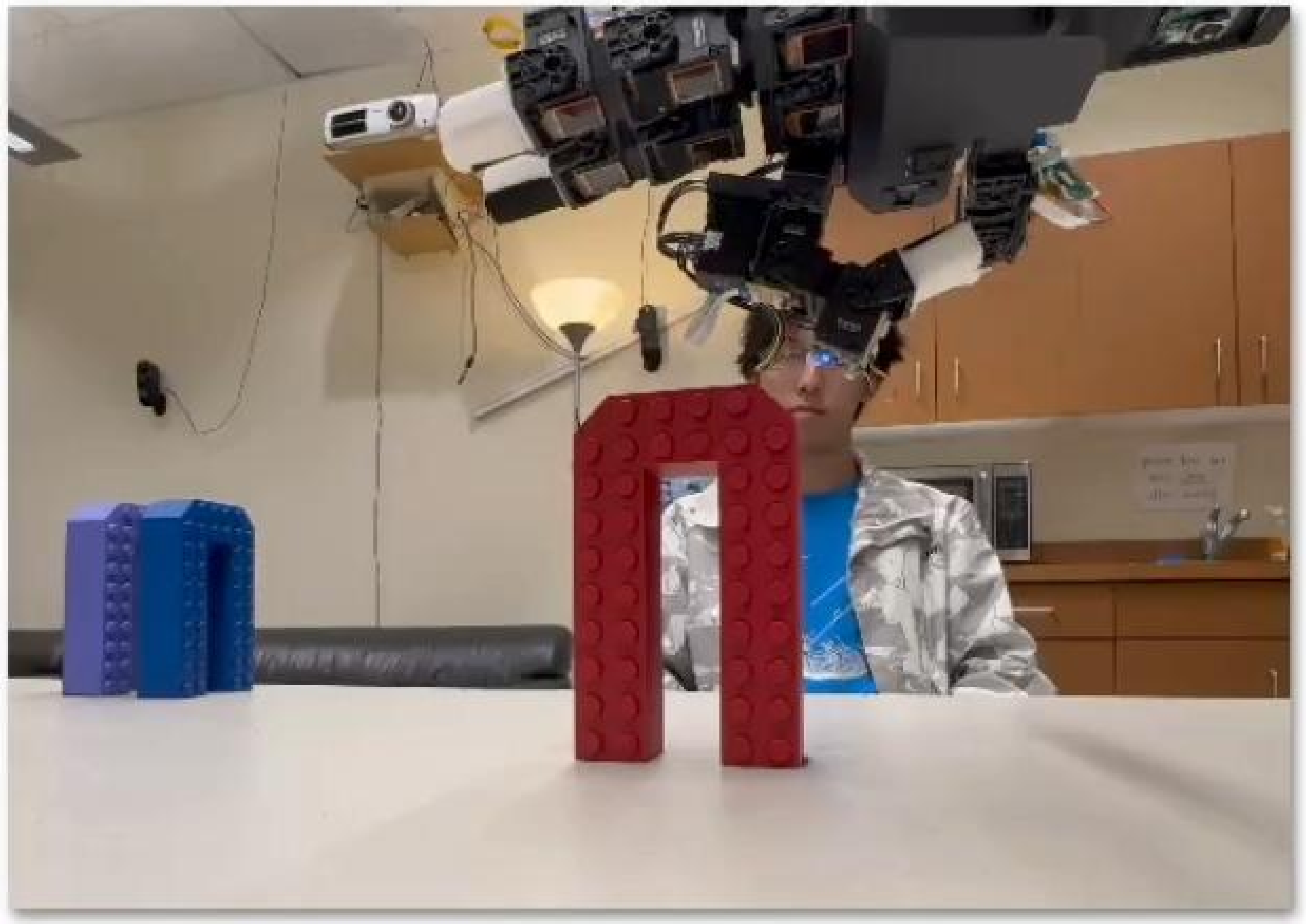
Mode:
Serial
Port:
/dev/tty.usbser
Baud Rate:
115200
Disconnect
Read Interval (ms):
10
Set
Enable EKF
Process Noise:
0.01
Measurement Noise:
0.1
Apply EKF Settings
Zero IMU
Reset View
IMU Orientation
X
Y
Z
Accel
Acceleration X
Acceleration Y
Acceleration Z
Gyroscope X
Gyroscope Y
Gyroscope Z

Design Motivation & Experimental Implementation

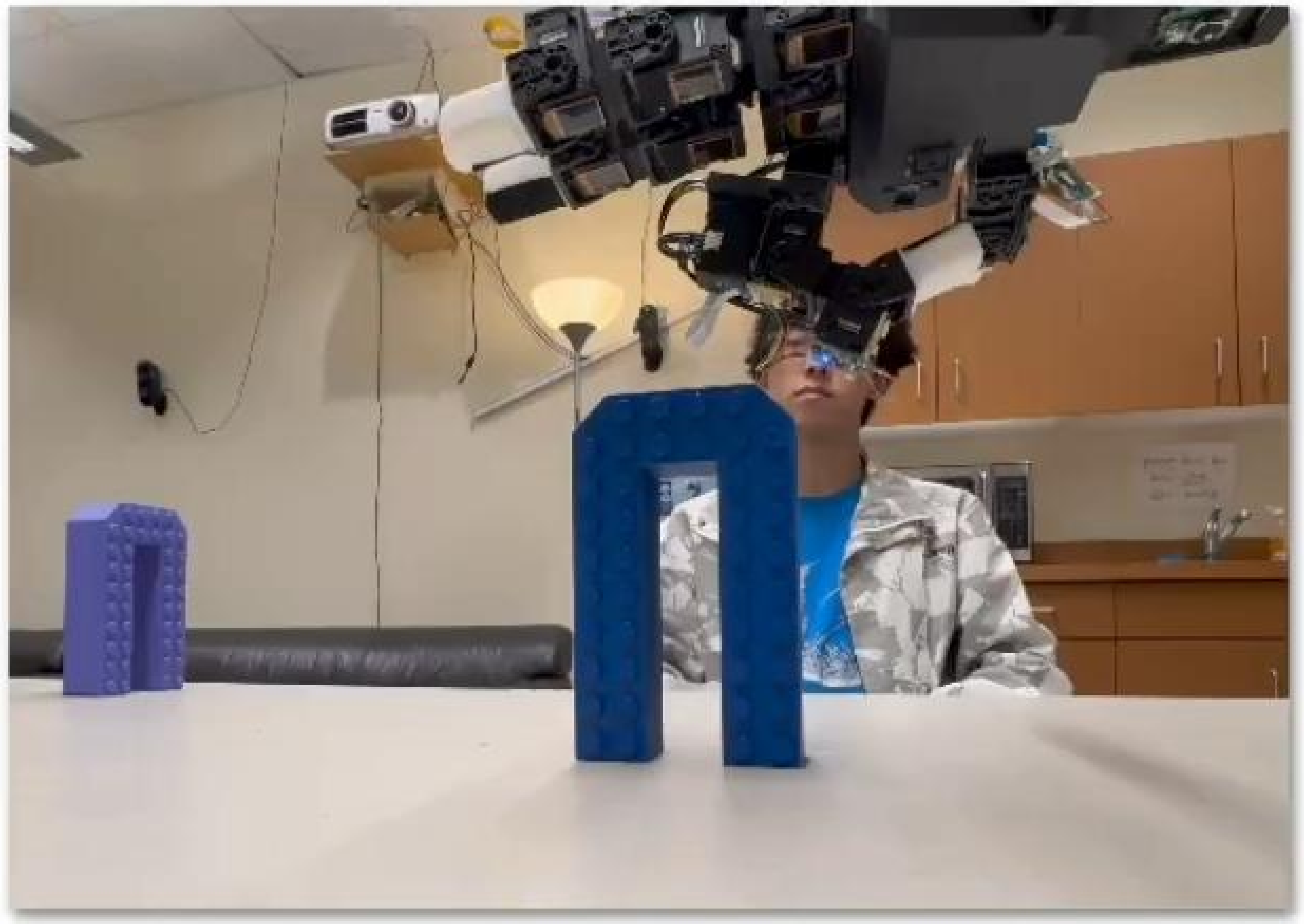
Mass-Aware Classification via Inertial Sensing

IMU Object Mass Classification

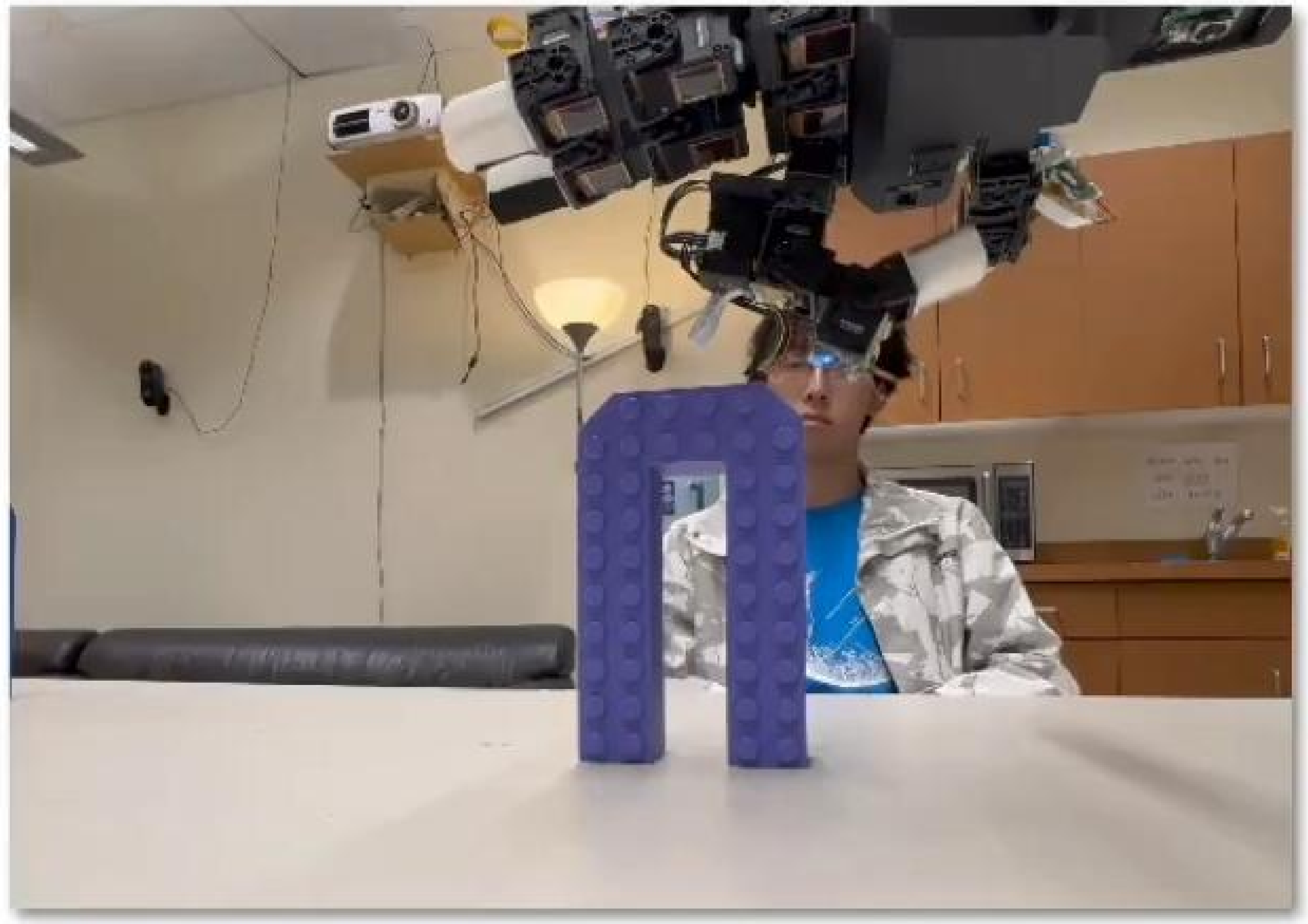
Experiment Setup



82 g



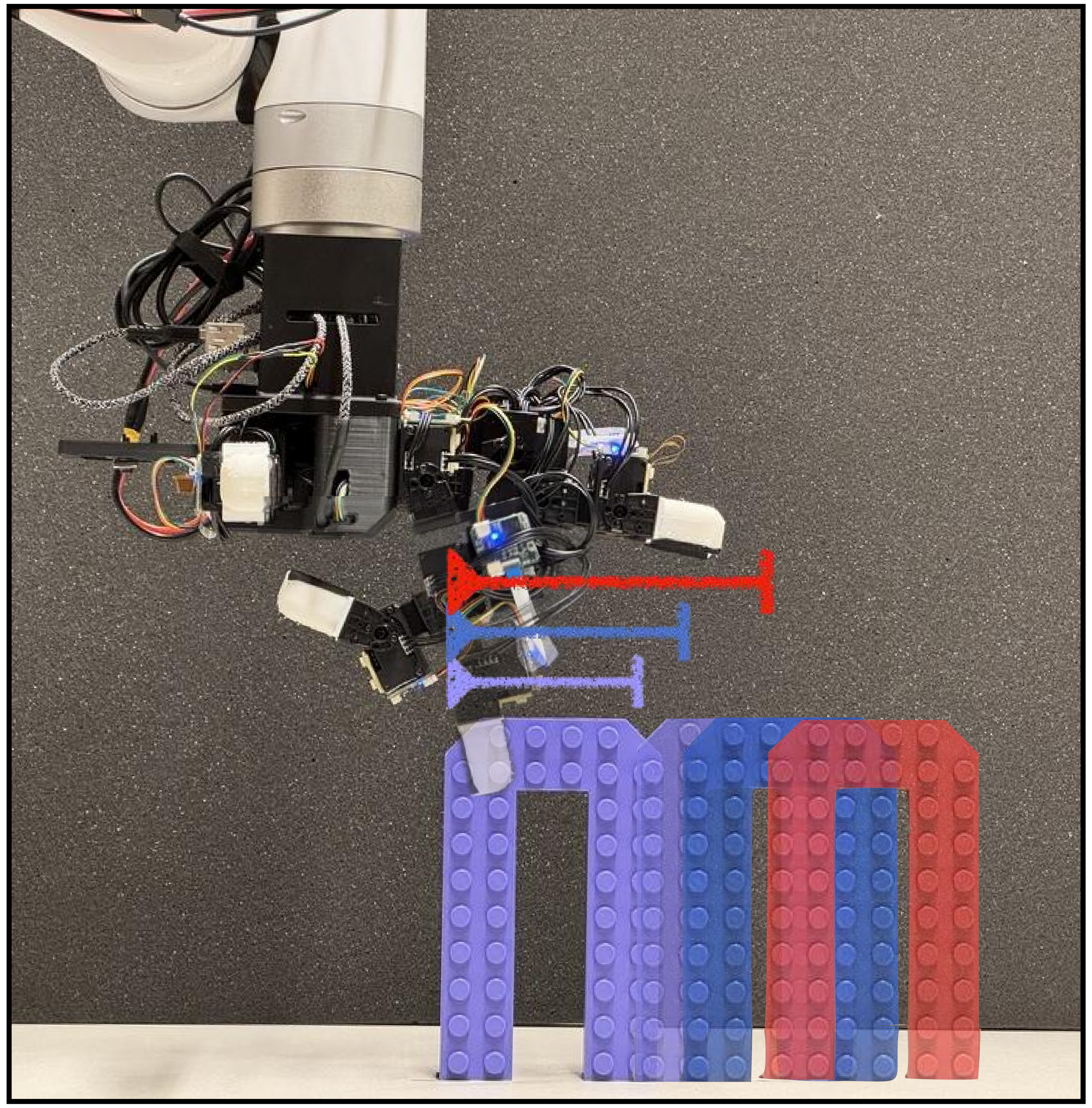
125 g



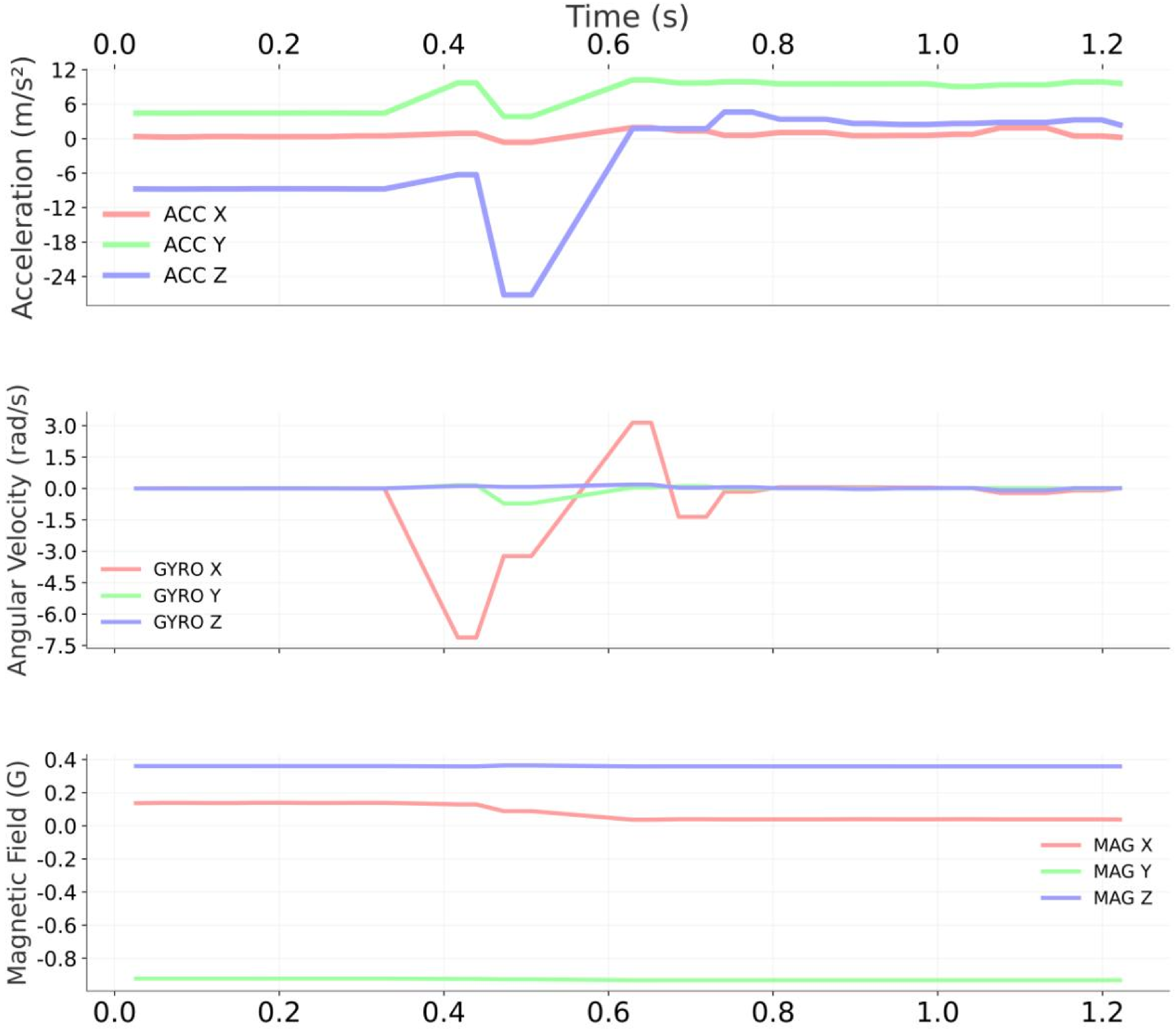
219 g

IMU Object Mass Classification

Experiment Setup



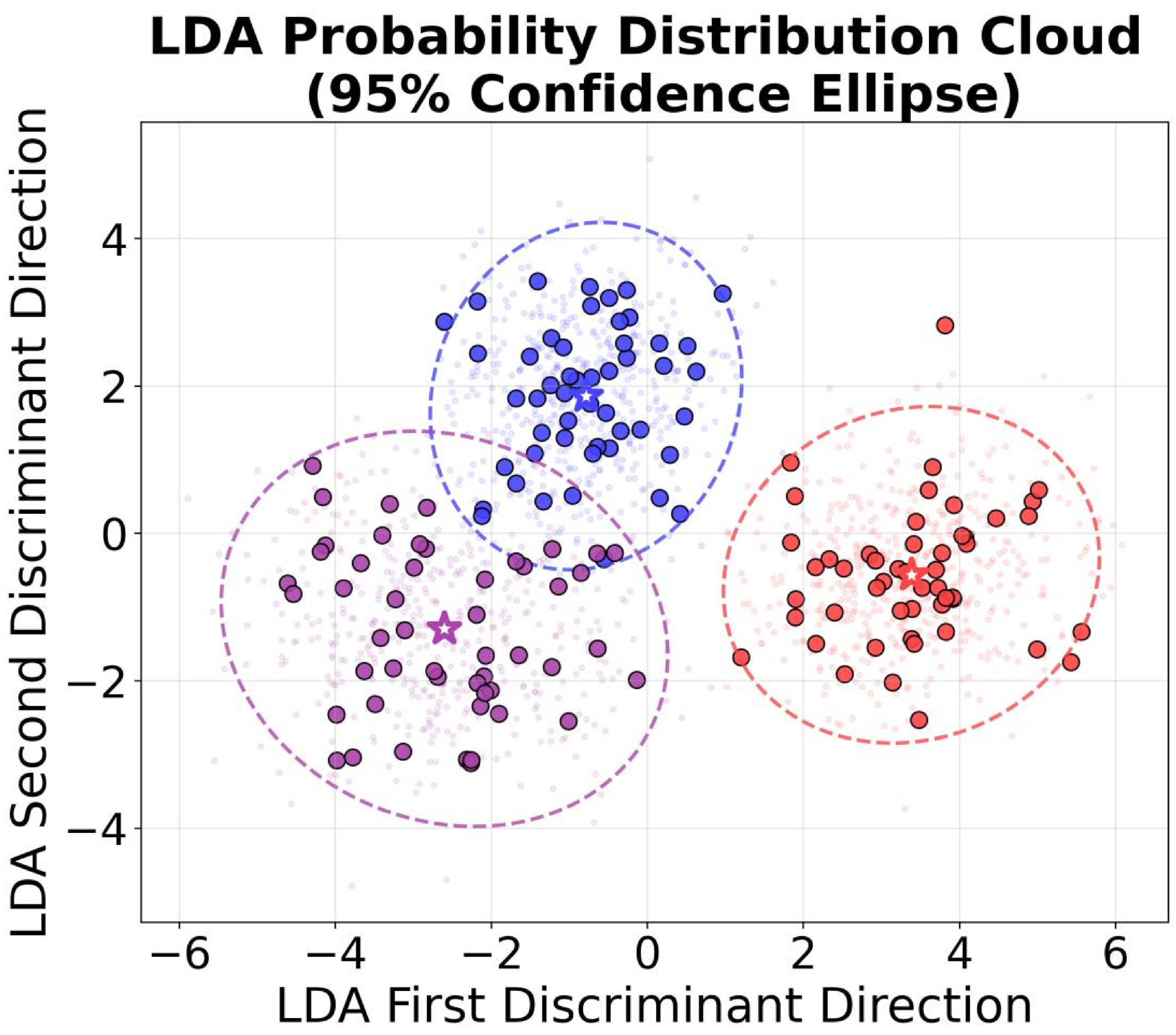
(a) Fingertip flicking experiment setup



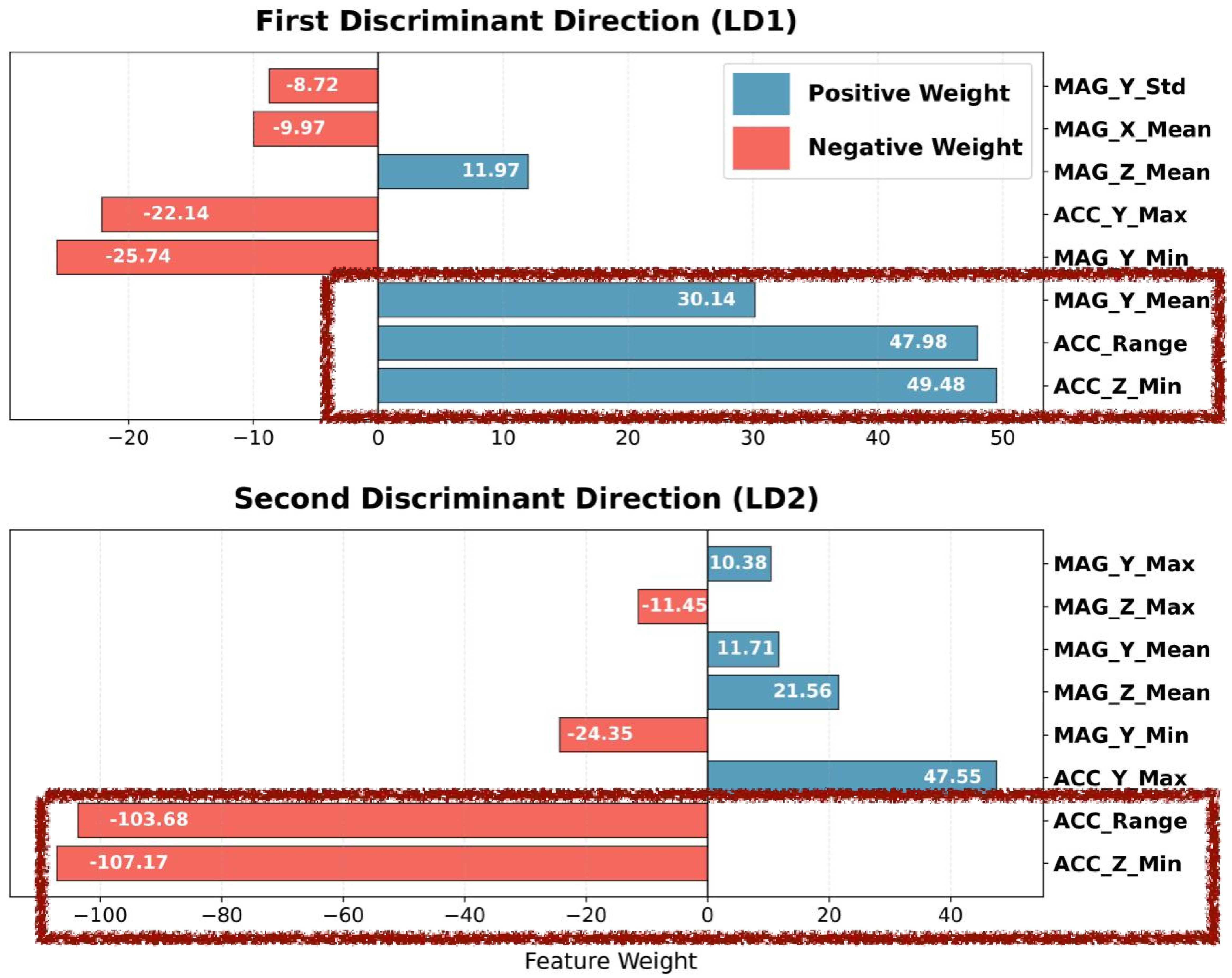
(b) Fingertip flicking sensor data acquisition

IMU Object Mass Classification

Result Analysis



(a) Two-dimensional LDA feature space showing separation of three object classes



(b) Feature contribution weights for the first and second discriminant directions

IMU Object Mass Classification

Result Analysis

(c) Confusion Matrix of Random Forest showing separation of three object classes

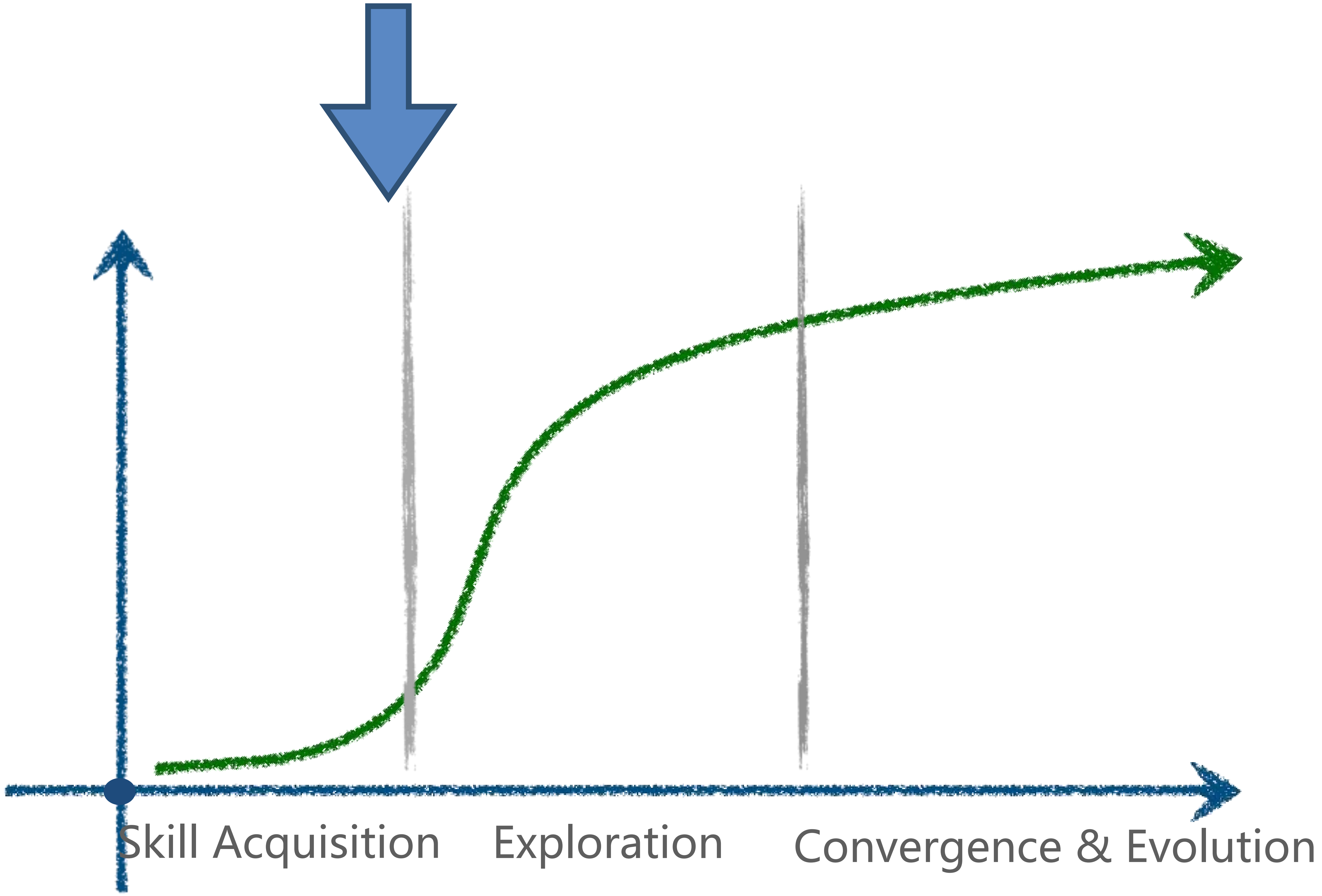
	Precision	Recall	F1-score	Support
Blue	0.94	1.00	0.97	15
Purple	1.00	0.93	0.97	15
Red	1.00	1.00	1.00	15
Accuracy			**0.98**	45
Macro avg	0.98	0.98	0.98	45
Weighted avg	0.98	0.98	0.98	45

(d) Classification accuracy for three object classes

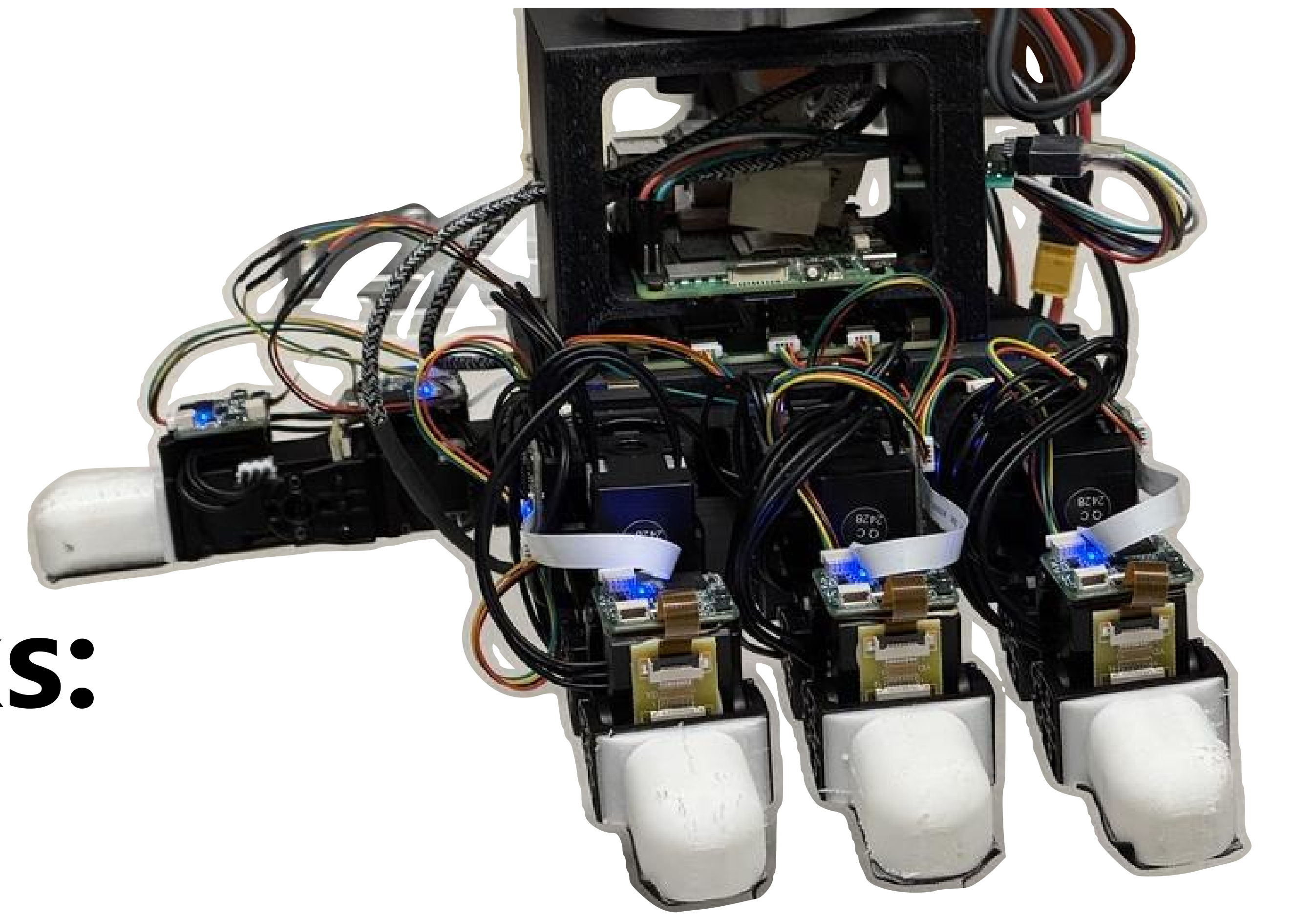
Wrap Up & Future Directions

Wrap Up

We Learn by Feeling, Just Like Babies Taking First Steps



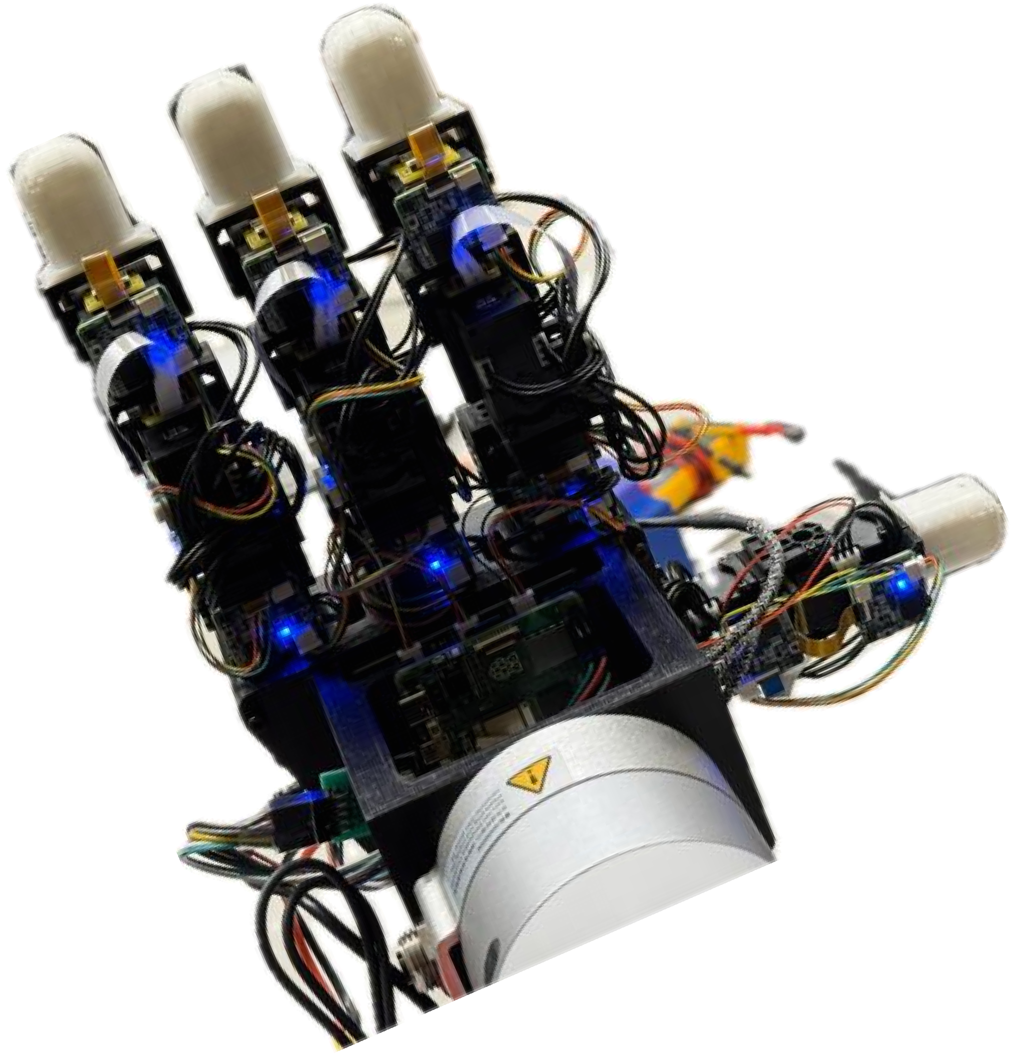
Future Directions and Takeaways

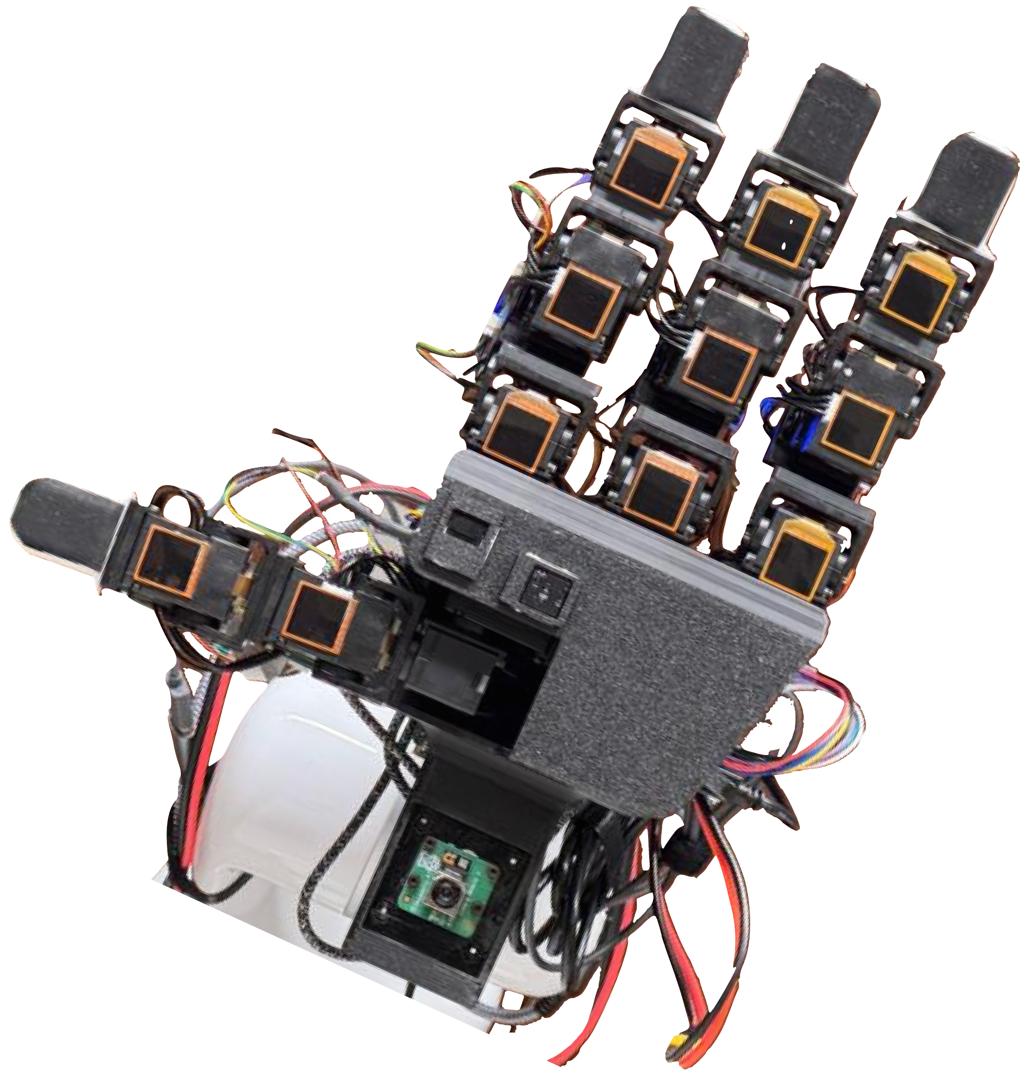


- **We can Leverage Multi-modal Sensing in More Tasks:**
  - Real-time thermal-aware sensing for Enhanced Manipulation:
    - Integrate point cloud-based SLAM with thermal distribution mapping. Then plug in real-time temperature perception into standard manipulation pipelines.
  - Material Classification through contact sensing:
    - Explore and distinguish object textures and material properties.
- Multi-modal sensing is essential for real-world manipulation and robot environment interaction.
- We're in the early exploration phase - like babies learning to walk, robotic manipulation is taking its steps toward a richer, more natural world understanding.

- **We Will Open-source the Design and Keep Improving the MOTIF Hand.**
  - We hope this inspires future work in developing novel methods for Multi-modal sensory robotic dexterous hands for complex tasks.

USC Viterbi
School of Engineering

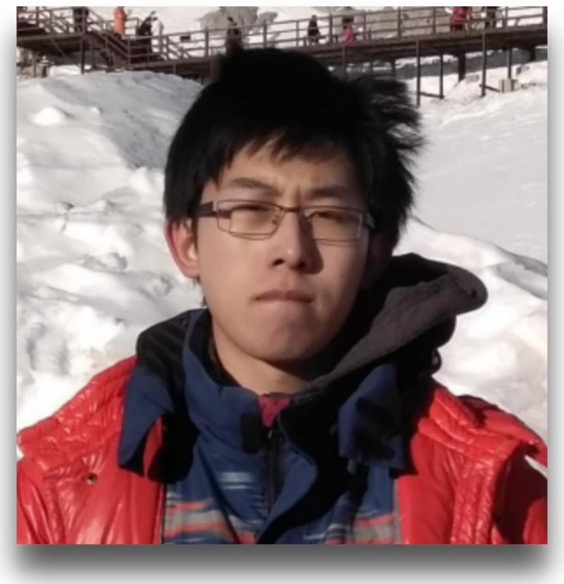
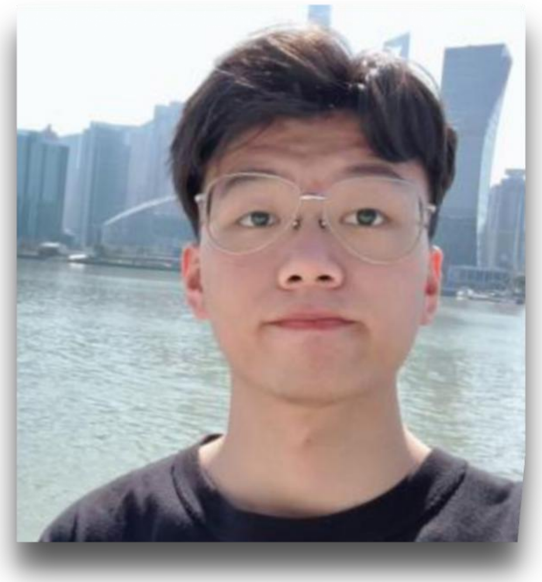
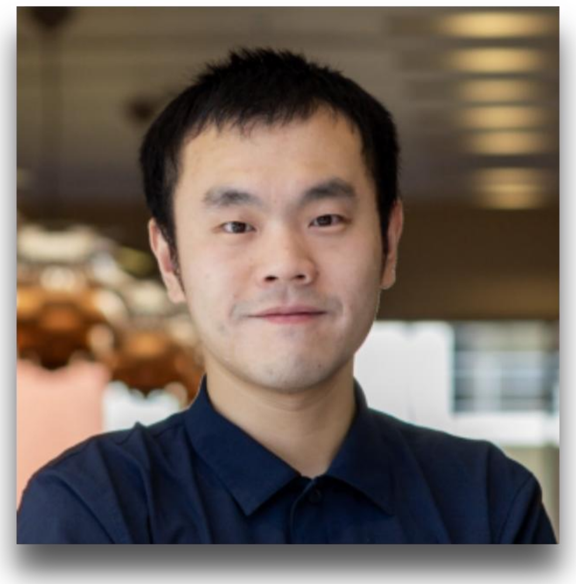




MOTIF Hand

A Robotic Hand for Multimodal Observations with Thermal, Inertial and Force Sensors.




Hanyang Zhou*,[1], Haozhe Lou*, Wenhao Liu*, Enyu Zhao, Yue Wang, and Daniel Seita

*Equal Contribution
[1] Applying to Ph.D