LEI ZHOU

✉: zhoulei0426@gmail.com in : www.linkedin.com/in/lei-zhou-nusarc : https://zray26.github.io

EDUCATION

National University of Singapore **Jan 2021 - Present**

Ph.D. in Mechanical Engineering (Advisor: Prof. Marcelo H. Ang Jr.)

- GPA: 4.33/5.00
- Relevant Coursework: Machine Vision; Deep Learning for Robotics; Digital Human

Huazhong University of Science and Technology **Sep 2014 - Jun 2018**

B.E. in Mechanical Engineering

- GPA: 3.80/4.00

WORK EXPERIENCE

Research Intern **May 2025 – Present**

Xiaomi Technology (Mentor: Dr. Long Chen)

- Leveraging **Xiaomi AI Glass** as a low-cost, scalable platform to collect diverse **egocentric human manipulation videos**, creating a massive dataset for **Vision-Language-Action (VLA)** model training.
- Validating **data scaling laws** by using this glass-collected human data to mitigate **real-world robot data scarcity**, exploring **World Models** and **Latent Action Models** to transfer human dexterity to generalist robots.
- **Core contributor** to the **MiMo-Embodied Foundation Model** project; leading **data curation** and **evaluation** for the **Embodied AI** domain, including **Chain-of-Thought (CoT)** and **Reinforcement Learning (RL)** data
- Achieved **SOTA** performance across **17 benchmarks**, validating the efficacy of the scaled data and training strategies.

Research Intern **May 2024 – Present**

Microsoft Research Asia – Beijing (Mentor: Dr. Jiaolong Yang)

- Core author of the **VITRA** project; developed the first fully-automated **human 3D VLA data** generation pipeline that transforms unstructured, real-life human videos into robot-aligned training data.
- Led the **3D Motion Labeling** system; engineered a high-fidelity pipeline integrating **HaWoR** for **hand reconstruction** and a modified **MegaSAM** for **metric-scale camera tracking**.
- Successfully processed **26 million frames** to curate **1 million atomic VLA episodes** with **dense 3D supervision**, enabling the model to achieve strong **zero-shot generalization** in unseen environments.
- Engineered a **Transformer-based reinforcement learning model** for dexterous robotic grasping, achieving a **91.2%** success rate on **seen objects** (**88.3%** on **unseen objects**) in simulation, and developed a **real-sim-real** pipeline integrating **NeRF-based object reconstruction**.

RESEARCH & PROJECT HIGHLIGHTS

Hand Mesh Reconstruction from Egocentric Videos **Nov 2024 – May 2025**

- Designed the **3D Motion Labeling** stage of the **VITRA** pipeline to recover **metric-scale world-space hand motion** from unscripted monocular egocentric videos.
- Integrated a modified **MegaSAM** visual SLAM system with **MoGe-2** depth estimation to achieve robust camera tracking, attaining **1.4 cm Absolute Trajectory Error (ATE)**.
- Employed **HaWoR** for per-frame **camera-space 3D hand reconstruction**, utilizing **spline smoothing** to ensure temporal coherence and remove outliers in world-space trajectories.
- Built the transformation pipeline to fuse camera poses and hand meshes into **world-space trajectories**, enabling the creation of **1 million atomic VLA episodes** for large-scale robot learning.

Reinforcement Learning for Dexterous Robotic Grasping **Aug 2024 – Nov 2024**

- Created a **universal Transformer-based model** via **offline distillation** from individually trained RL policies on **3,200 objects**.
- In simulation, it achieves up to **a 91.2%** success rate on seen objects (**88.3%** on new categories) in a **state-based** setting and **88.9%** (**86.8%**) in a **vision-based** setting.

- Designed a **real-sim-real pipeline** integrating **NeRF-based** object reconstruction into RL environments. Captured real-world objects as meshes, trained grasping policies in simulation, and transferred them back for real-world testing.
- Achieved state-of-the-art performance, with a paper accepted by **CVPR 2025**.

Robust Multiview Hand Mesh Reconstruction **May 2024 – Aug 2024**

- Reconstructed **hand meshes** for each camera view in the multiview setup.
- Applied **triangulation-based optimization** for multiview consistency, improving the robustness of the reconstruction.
- Incorporated **smoothing techniques** to ensure **temporal coherence** across video frames, yielding high precision in **3D keypoint annotation**.
- Validated **70%+** of frames (10 million total) for **4D hand pose** estimation and **VLA** model pretraining.

Diffusion-based Multi-Hands Robotic Grasp Generation **Nov 2023 - Mar 2024**

- Developed a **diffusion-based model** to generate grasp poses for multiple robotic dexterous hands.
- Introduced **visual affordance detection** and **open-vocabulary** analysis to filter functional grasp candidates.
- Achieved **44.73%** overall success rate on the **MultiDex Dataset**, improving generalization across multiple hand types.

Dynamic Scene Reconstruction for Robotic Grasping **Jun 2023 - Sep 2023**

- Utilized **SDF-based methods** to reconstruct novel objects from multi-view images.
- Designed a **dynamic scene reconstruction pipeline** that completed object point clouds in real time (**9.2 FPS**) by leveraging tracked object poses.
- Achieved **state-of-the-art performance** on the **GraspNet-1Billion benchmark** for robotic grasping tasks, with a paper accepted by **ICRA 2024**.

Category-Level Object Pose Estimation **Jan 2022 - Mar 2023**

- Developed a network that **completes partial point clouds** from depth camera data and **reconstructs objects** in canonical space by **deforming a shape prior**.
- Designed a robust **3D shape-matching module** to align reconstructed objects with observed partial point clouds for **pose and size estimation**.
- Achieved **state-of-the-art performance**, with a paper accepted by **IROS 2023**.

SELECTED PUBLICATIONS

- Xiaoshuai Hao, **Lei Zhou**, et al., “MiMo-Embodied: X-Embodied Foundation Model Technical Report”, 2025
- Qixiu Li, Yu Deng, Yaobo Liang, Lin Luo, **Lei Zhou**, Chengtang Yao, Lingqi Zeng, Zhiyuan Feng, Huizhi Liang, Sicheng Xu, Yizhong Zhang, Xi Chen, Hao Chen, Lily Sun, Dong Chen, Jiaolong Yang, Baining Guo, “Scalable Vision-Language-Action Model Pretraining for Robotic Manipulation with Real-Life Human Activity Videos”, under review, 2025
- Zhengshen Zhang, Hao Li, Yalun Dai, Zhengbang Zhu, **Lei Zhou**, Chenchen Liu, Dong Wang, Francis E. H. Tay, Sijin Chen, Ziwei Liu, Yuxiao Liu, Xinghang Li, Pan Zhou, “From Spatial to Actions: Grounding Vision-Language-Action Model in Spatial Foundation Priors”, under reiview, 2025
- Wenbo Wang, Fangyun Wei, **Lei Zhou**, Xi Chen, Lin Luo, Xiaohan Yi, Yizhong Zhang, Yaobo Liang, Chang Xu, Yan Lu, Jiaolong Yang, and Baining Guo, “UniGraspTransformer: Simplified Policy Distillation for Scalable Dexterous Robotic Grasping”, CVPR 2025
- **Lei Zhou**, Haozhe Wang, Zhengshen Zhang, Zhiyang Liu, Francis EH Tay, and Marcelo H. Ang Jr., “You Only Scan Once: A Dynamic Scene Reconstruction Pipeline for 6-DoF Robotic Grasping of Novel Objects”, ICRA 2024
- Zhengning Zhou, **Lei Zhou**, Shengxin Sun, and Marcelo H. Ang Jr., “A Robust and Efficient Robotic Packing Pipeline with Dissipativity-Based Adaptive Impedance-Force Control”, IROS 2024
- Zhiyang Liu, Ruiteng Zhao, **Lei Zhou**, Chengran Yuan, Yuwei Wu, Sheng Guo, Zhengshen Zhang, and Marcelo H. Ang Jr., “3D Affordance Keypoint Detection for Robotic Manipulation”, IROS 2024

- Zhengshen Zhang, **Lei Zhou**, Chenchen Liu, Chengran Yuan, Sheng Guo, Ruiteng Zhao, Marcelo H. Ang Jr., and Francis EH Tay, “DexGrasp-Diffusion: Diffusion-Based Unified Functional Grasp Synthesis Method for Multi-Dexterous Robotic Hands”, ISRR 2024
- **Lei Zhou**, Zhiyang Liu, Runze Gan, Haozhe Wang, and Marcelo H. Ang Jr., “DR-Pose: A Two-stage Deformation-and-Registration Pipeline for Category-level 6D Object Pose Estimation”, IROS 2023

CORE SKILLS

- Programming: Python, PyTorch, ROS.
- Computer Vision: Egocentric Human Pose Estimation, 3D Reconstruction, Object Pose Estimation.
- Robotics: Robotic Grasping, Reinforcement Learning.
- Language: Mandarin (native), English (fluent).