

# HAN LIANG

Ph.D. candidate, ShanghaiTech University

+86 13540034874 [◇ lianghan@shanghaitech.edu.cn](mailto:lianghan@shanghaitech.edu.cn) [◇ www.liang-han.com](http://www.liang-han.com)

## ABOUT ME

---

I'm now a Ph.D. student at [VRVC Lab, ShanghaiTech University](#), working with [Prof. Lan Xu](#) and [Prof. Jinyi Yu](#) on digital humans. Prior to that, I obtained my B.E. in computer software engineering from [UESTC](#).

My research focuses on the intersection of graphics, vision, and robotics, especially digital humans and generative AI. My ultimate aspiration is to realize human-centric multi-modal AI agents that liberate us human beings from the tedious and heavy work.

I expect to defend my Ph.D. in 2025, and I'm now actively looking for intern and full-time positions. I'm also open to any opportunities for cooperation and discussions. If you are interested, please don't hesitate to contact me.

## RESEARCH INTERESTS

---

**Fields:** Computer Vision, Graphics, Robotics

**Topics:** Digital humans, Generative AI, Multi-modal LLM

## EDUCATION

---

**ShanghaiTech University**

Ph.D. in Computer Science

Sep 2020 - 2025 (expected)

*Advisor: Prof. Lan Xu & Prof. Jinyi Yu*

**University of Electronic Science and Technology of China**

B.E. in Computer Software Engineering

Sep 2014 - Jun 2018

Rank 7/134 *Advisor: Prof. Qiao Liu*

## SELECTED PUBLICATIONS ([COMPLETE LIST](#))

---

- [1] **Media2Face: Co-speech Facial Animation Generation With Multi-Modality Guidance**  
Qingcheng Zhao\*, Pengyu Long\*, Qixuan Zhang, Dafei Qin, **Han Liang**, Lan Xu<sup>†</sup>, Jinyi Yu<sup>†</sup>  
*ACM SIGGRAPH, 2024*
- [2] **OMG: Towards Open-vocabulary Motion Generation via Mixture of Controllers**  
**Han Liang**, Jiacheng Bao, Ruichi Zhang, Sihan Ren, Sibe Yang, Xin Chen, Jinyi Yu<sup>†</sup>, Lan Xu<sup>†</sup>  
*IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2024*
- [3] **InterGen: Diffusion-based Multi-human Motion Generation under Complex Interactions**  
**Han Liang**, Wenqian Zhang, Wenxuan Li, Jinyi Yu<sup>†</sup>, Lan Xu<sup>†</sup>  
*International Journal of Computer Vision (IJCV), 2024*
- [4] **HybridCap: Inertia-aid monocular capture of challenging human motions**  
**Han Liang**, Yannan He, Chengfeng Zhao, Mutian Li, Jingya Wang, Jinyi Yu<sup>†</sup>, Lan Xu<sup>†</sup>  
*AAAI Conference on Artificial Intelligence (AAAI), 2023 *Oral**
- [5] **LiDAR-aid Inertial Poser: Large-scale Human Motion Capture by Sparse Inertial and LiDAR Sensors**  
Yiming Ren\*, Chengfeng Zhao\*, Yannan He, Peishan Cong, **Han Liang**, Jinyi Yu, Lan Xu<sup>†</sup>, Yuexin Ma<sup>†</sup>  
*IEEE Transactions on Visualization and Computer Graphics (TVCG), 2023*
- [6] **ChallenCap: Monocular 3d capture of challenging human performances using multi-modal references**  
Yannan He, Anqi Pang, Xin Chen, **Han Liang**, Minye Wu, Yuexin Ma, Lan Xu  
*IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021 *Oral**

## PROJECTS

---

- [1] **SignGPT: Multi-modal Large Sign Language Models.** We propose Sign-X, the most comprehensive and largest multi-modal dataset for Chinese Sign Language (CSL), spanning over 1000 hours and incorporating diverse modalities such as videos, 2D keypoints, 3D SMPL-X skeletons, glosses, and texts. Based on rich data, we pioneer the development of the first Large Sign Language Models (LSLM) to date, equipped with up to 72B parameters. LSLMs cross a spectrum of complex downstream tasks including Sign Language Production (SLP), Translation (SLT), and Recognition (SLR), setting new standards in the field.
- [2] **RhyLive: Monocular Full-body Mocap for Real-time Streaming.** Achieving fine-grained capture of the upper body, face, and hands using a single camera. The system has been integrated into the Bilibili Live streaming pipeline. ([Demo](#))
- [3] **RhyCap: Sparse-view Real-time Full-body Mocap System.** We propose a lightweight real-time markerless mocap system. With even only three consumer-grade web cameras, the system achieves closing industry-level accuracy. This system is now integrated into the Bilibili virtual Live mocap pipeline. ([Demo](#))
- [4] **NIR+VIS+Depth Multi-modal 3D Face Recognition System.** A method based on a latent variable model was proposed, and the recognition top-1 hit ratio of 1:0.5 million closed-set tests was improved from 91.62% to 96.37%. This system has been applied to the Zhuhai-HongKong-Macao Bridge national project and railway stations in Hefei, Urumqi, and other cities.

## AWARDS

---

National Doctorial Scholarship	2024
National Encouragement Scholarship	2016

## EXPERIENCE

---

DGene Inc. Research Intern	Jun 2021 - May 2022 <i>Supervisor: Dr. Yingliang Zhang</i>
Dilusense Inc. 3D Vision R&D	Jul 2018 - Jun 2020 <i>Supervisor: Prof. Juyong Zhang</i>
Graphics&Geometric Computing Laboratory, USTC Visiting Student	Oct 2017 - Jun 2018 <i>Supervisor: Prof. Ligang Liu</i>

## PROFESSIONAL SERVICES

---

**Conference Reviewer:** AAAI, ICCV, CVPR, SIGGRAPH  
**Journal Reviewer:** IJCV, TVCG, TMM

## TEACHING ASSISTANCE

---

CS280 Deep Learning	<i>Instructor: Prof. Xuming He &amp; Prof. Lan Xu</i>
CS283 Robotics	<i>Instructor: Prof. Laurent Kneip &amp; Prof. Sören Schwertfeger</i>

## SKILLS

---

### Programming Languages

Python (Pytorch, Pyrender, RL games, Issac gym, and so on.)  
C++ (OpenCV, CUDA, and so on.)

### Softwares

Visual Studio, Pycharm, Jupyter Notebook, Latex  
Unity, Blender, Maya