MINGSHENG LI

Fudan University | H: +86 199 4623 3989 | limc22@m.fudan.edu.cn
Homepage

About Me

I am currently a second-year Master's student in Artificial Intelligence at Fudan University, advised by Prof. Tao Chen. I am also fortunate to work with Dr. Hongyuan Zhu from A*STAR, Singapore, Dr. Gang Yu, Dr. Xin Chen, and Dr. Chi Zhang from Tencent, and Dr. Bo Zhang from Shanghai AI Lab. Before this, I received my bachelor's degree in Electronic Engineering from Fudan University in 2022.

I work in the fields of deep learning and computer vision, with particular focuses on **large models**, **multi-modal learning**, and **embodied AI**. My research pursues to develop robust and scalable general-purpose AI systems to solve complex problems. Besides research, I also love sports, music, and board games.

Research Interests

Multi-modal Learning, Vision and Language, Large Language Models, and Continual Learning.

Education

Masters in Artificial Intelligence (GPA 3.51/4.00) — Sep. 2022 - Jun. 2025
Fudan University. — Shanghai, China
Bachelor in Electronic Engineering — Sep. 2018 - Jun. 2022
Fudan University. — Shanghai, China

Publications and Pre-prints

- WI3D: Weakly Incremental 3D Detection via Vision Foundation Models
  Mingsheng Li, Sijin Chen, Shengji Tang, Hongyuan Zhu, Yanyan Fang, Xin Chen, Zhuoyuan Li, Fukun Yin, Tao Chen.
  [Under Review | paper]
  [Summary]: Introducing new categories to pre-trained 3D detectors with 2D foundation models.
- M3DBench: Let's Instruct Large Models with Multi-modal 3D Prompts.
  Mingsheng Li, Xin Chen, Chi Zhang, Sijin Chen, Hongyuan Zhu, Fukun Yin, Gang Yu, Tao Chen.
  [ECCV 2024 | project | arxiv | github]
  [Summary]: A large-scale dataset instructing 3D VLMs with interleaved multi-modal prompts.
- Lightweight Model Pre-training via Language Guided Knowledge Distillation.
  Mingsheng Li, Lin Zhang, Mingzhen Zhu, Zilong Huang, Gang Yu, Tao Chen, Jiayuan Fan.
  [T-MM 2024 | paper | arxiv | github]
  [Summary]: Using language to refine knowledge distillation between teacher and student.
- Vote2Cap-DETR++: Decoupling Localization and Describing for End-to-End 3D Dense Captioning.
  Sijin Chen, Hongyuan Zhu, **Mingsheng Li**, Xin Chen, Peng Guo, Yinjie Lei, Gang Yu, Taihao Li, Tao Chen.
  [T-PAMI 2024| paper | arxiv | github]
  [Summary]: Decoupled feature extraction for localizing and describing objects in 3D scenes.
- LL3DA: Visual Interactive Instruction Tuning for Omni-3D Understanding, Reasoning, and Planning.
  Sijin Chen, Xin Chen, Chi Zhang, **Mingsheng Li**, Gang Yu, Hao Fei, Hongyuan Zhu, Jiayuan Fan, Tao Chen.
  [CVPR 2024 | project | paper | github]
  [Summary]: 3D-LLMs respond to visual and text interactions in complex 3D scenes.

Projects

- **End-to-end 3D object detection with Mamba.** Jan. 2024 - May. 2024
  Proposed 3DET-Mamba, a state-space model for indoor 3D object detection, integrating local- and global- feature learning with a novel decoding module, under review as a conference paper.
- **Language for 3D Scenes.** Aug. 2022 - Jan. 2024
  Proposed **Vote2Cap-DETR++**, a method that decouples the learnable queries into localization and caption queries to capture task-specific features for describing objects in 3D scenes, accepted to T-PAMI 2024. Presented **LL3DA**, a large language 3D assistant responding to both text and visual interactions with complex 3D scenes, accepted to CVPR 2024. Put forward **M3DBench**, a comprehensive 3D-language dataset covering 327k interleaved multi-modal instructions for 10 tasks related to 3D perception, understanding, reasoning, and planning, , accepted to ECCV 2024.
- **Continual Learning for 3D Detection.** Mar. 2023 - Sep. 2023
  Proposed WI3D, a label-efficient approach for continual learning that utilizes cost-effective 2D foundation models to introduce new categories to 3D detectors without forgetting previously learned knowledge, now under review as a journal paper.
- **Earlier Projects.** Before Sep. 2022
  Multi-modal 3D Object Detection. Proposed a Transformer-based multi-modal 3D detection model that performs semantic alignment and feature fusion between images and point clouds.
  Mask-Wearing Detection System. Developed a smart system that infers spatial relationships of facial key points (e.g., mouth, nose) to detect incorrect mask-wearing and recommend proper mask usage.

Scholarships and Awards

Second Prize of Graduate Academic Scholarship.	2023
Outstanding Graduate of Fudan University.	2022
Second Prize of China Undergraduate Mathematical Contest in Modeling, Shanghai.	2020
STEM (Science, Technology, Engineering, and Mathematics) Scholarship.	2020
First Prize of Chinese Mathematics Competitions, Shanghai (Top 20).	2019
National Encouragement Scholarship.	2019

Skills

Languages:	Chinese (native), English
Programming:	Python (primary), C, Matlab, SQL
Tools:	PyTorch, Visual Studio, MeshLab, Jupyter Notebook

References

Prof. Tao Chen	Supervisor	Fudan University	eetchen@fudan.edu.cn
Dr. Hongyuan Zhu	Co-author	A*STAR, Singapore	hongyuanzhu.cn@gmail.com