

Jiting Cai

[Personal Page](#) [Google Scholar](#) [Semantic Scholar](#)

(+1) 857-706-9289 caijiting@sjtu.edu.cn jitingcai515@gmail.com

800 Dongchuan Road, Shanghai Jiao Tong University, P.R.C.

EDUCATION

Shanghai Jiao Tong University (SJTU), Shanghai, China

Sep 2021 - Jun 2025 (Expected)

Bachelor of Computer Science and Engineering, Department of Computer Science and Engineering

- **Overall GPA: 4.1222/4.30 (Ranking: 1/103) — Major GPA: 4.1410/4.30 (Ranking: 1/103)**
- **Core Courses:** Calculus(1)&(2)(A&A+), Linear Algebra(A+), Probability and Statistics(A+), Computer Organization Principle(A+), Principles of Assembly and Compilation(A+), Computer Graphics(A+), Mathematical Foundations of Computer Science (A+), Experiments in Computer Organization(A+), Machine Learning (A+), Principal of Data Science (A+)

PUBLICATIONS & MANUSCRIPTS

1. **Take a Step Back: Rethinking the Two Stages in Visual Reasoning** [\[arXiv\]](#)
*Mingyu Zhang**, **Jiting Cai***, Yue Xu, Mingyu Liu, Cewu Lu, Yong-Lu Li. proceeding in *ECCV 2024*
2. **Architect: Generating Vivid and Interactive 3D Scenes with Hierarchical 2D Inpainting** [\[arXiv\]](#)
Yian Wang, Xiaowen Qiu, Jiageng Liu, Zhehuan Chen, Jiting Cai, Yufei Wang, Tsun-Hsuan Wang, Zhou Xian, Chuang Gan. proceeding in *NeurIPS 2024*

RESEARCH EXPERIENCES (SELECTED)

Project: Differential Physics and Scene Simulation

Mar 2024 - Now

MIT-IBM Watson AI Lab, Massachusetts Institute of Technology [\[website\]](#)

Advisor: Prof. Chuang Gan [\[personal page\]](#)

- proposed a zero-shot generative pipeline that creates diverse, complex, and realistic 3D interactive scenes to advance Embodied AI agents and Robotics research
- proposed to leverage 2D prior from vision generative models to facilitate the 3D interactive scene generation process, and make such process controllable by initializing from simulation-rendered image for hierarchical inpainting, ensuring consistent spatial features and controllable camera parameters and depth scale, allowing accurate 2D to 3D lifting
- Attempted to control the regions of rooms that the model can draw using masks and employed a pre-trained diffusion model to generate furniture and small items in available drawing areas based on corresponding prompts
- Used a multimodal large model and an open-source 3D object dataset to find suitable items based on the 2D diffusion generation results, arranged furniture and small objects in a 3D scene, and constrained object placement to prevent errors
- Helped improve the grounding of inpainting results for small objects by controlling their Euler angles and the scaling ratios in the x, y, and z directions, making them compatible with bounding boxes calculated from depth information, semantic segmentation masks, and camera intrinsic parameters
- Assisted in writing a paper on the relevant findings and published the paper at **NeurIPS 2024** [\[arXiv\]](#)
- Conducted extensive experiments during the rebuttal process of the paper, including providing the generated results of the text-to-3D model, testing the degree of matching between the generated scenes and the text prompts, and validating the necessity of blurring the inpainting masks. This led to the reviewers' unanimous inclination to accept the paper, ultimately resulting in its successful acceptance

Project: Visual Reasoning Framework Design

Apr 2023 - Now

MVIG Lab, RHOS Group, Shanghai Jiao Tong University [\[website\]](#)

Advisor: Prof. Yong-Lu Li [\[personal page\]](#)

- Identified that current visual reasoning technique all lacks generalization ability and would be restricted within a single task
- Analyzed the difference between symbolization stage and the reasoning stage of current visual reasoning model and determined that the reasoning stage is more general compared to the symbolization stage

*These authors contributed equally to this work

- Introduced a new visual reasoning framework based on the analysis and used different encoder for different tasks' symbolization stage while used the unique reasoning network for reasoning process
- Introduced the "approximation principle", which showcases the reasoner's ability could be improved by adding more independent tasks that belong to different domains and modalities
- Contributed to the analysis of disentanglement of symbolization and reasoning, the design of the framework structure, conducting of the experiments
- Contributed to the conclusion and paper writing of the project, publish paper at ECCV 2024 as the joint first author [\[arXiv\]](#)

Project: Multi-modal Learning for Medical Diagnosis

July 2024 - Sep 2024

Department of Computer Science and Engineering, Shanghai Jiao Tong University [\[website\]](#)

Advisor: Prof. Yi Hong [\[personal page\]](#)

- Conducted an in-depth survey of cutting-edge papers on the use of multi-modal data to assist medical diagnosis of complicated disease like Alzheimer
- Researched the currently mainstream medical image and text datasets, seeking compatible MRI, CT, and X-ray diagnostic images along with their corresponding textual diagnostic reports
- Leveraged existing Alzheimer's disease research reports to extract key information and used a large language model (LLM) to generate a large-scale knowledge graph that can be embedded into the model, aiming to design a knowledge-driven multimodal diagnostic model
- Developed a new multimodal network that utilizes a Mixture of Experts (MoE) and a hierarchical feature extraction approach to progressively align and fuse data from images and text. Experimental results demonstrate the network's decent performance in Alzheimer's diagnosis

Project: Applications of Differential Privacy

Oct 2022 - Mar 2023

John Hopcroft Center for Computer Science, Shanghai Jiao Tong University [\[website\]](#)

Advisor: Prof. Liyao Xiang [\[personal page\]](#)

- Conducted the survey considering differential privacy as the crucial technique for privacy protection and network security
- Helped to design a special kind of differential privacy method which favors comparatively efficiency
- Participated in mathematical analysis and conducting the experiments

ACADEMIC SERVICE

- **Conference Reviewer: ICLR 2025**

HONORS (SELECTED)

Merit Student (top 5%) <i>Shanghai Jiao Tong University</i>	2022, 2024
Academic Excellence Award Scholarship - A Level (top 3%) <i>Shanghai Jiao Tong University</i>	2022
Excellent Project in Engineering Practice (top 5th place) <i>Shanghai Jiao Tong University</i>	2022
Yang Yuanqing Undergraduate Honors Scholarship (5 persons per year) <i>Department of Computer Science and Engineering, Shanghai Jiao Tong University</i>	2024

SKILLS

Programming: C, C++, Python, PyTorch, MATLAB, Assembly Language, Verilog, L^AT_EX

Language: English (fluent), Mandarin Chinese (native)

- GRE **333+4.5**: Verbal 163 + Quantitative 170 + Analytical Writing 4.5
- TOEFL **108** : Reading 30 + Listening 28 + Speaking 23 + Writing 27

Teaching: Study Mentor of SJTU

Music: an excellent flute player