Jiageng Liu

• Phone: +1-4132654677 • Email: jiagengliu@umass.edu • Page: https://jiagengliu02.github.io

EDUCATION

University of Massachusetts Amherst Massachusetts, United States
Master in Computer Science ***GPA****:* 4.00/4.00 09/2024-12/2025

Zhejiang University Hangzhou, China
B.E. in Artificial Intelligence ***GPA****:* 3.75/4.00 ***Major GPA*** *(Third year)****:*** 3.96/4.00 09/2020-07/2024

PUBLICATIONS

- Yuncong Yang*, **Jiageng Liu*,** Zheyuan Zhang et al. Test-Time Scaling with World Models for Spatial Reasoning. *Advances in Neural Information Processing Systems (NeurIPS2025).* Program Website
- Yian Wang*, Xiaowen Qiu*, **Jiageng Liu*** et al. Architect: Generating Vivid and Interactive 3D Scenes with Hierarchical 2D Inpainting. *Advances in Neural Information Processing Systems (NeurIPS2024).* Program Website
- Zeyuan Yang*, **Jiageng Liu***, Peihao Chen, Anoop Cherian, Tim K. Marks, Jonathan Le Roux, Chuang Gan. RILA: Reflective and Imaginative Language Agent for Zero-Shot Semantic Audio-Visual Navigation. *Advances in Computer Vision and Pattern Recognition (CVPR2024).* Program Website
- Wang Zehan, Yang Zhao, Xize Cheng, Haifeng Huang, **Jiageng Liu** et al. Connecting multi-modal contrastive representations. *Advances in Neural Information Processing Systems (NeurIPS2023).* Program Website
- Yuhang Xu, Wanxu Xia, Yipeng Chen, **Jiageng Liu** et al. WordDiffuser: Helping beginners cognize and memorize glyph-semantic pairs using diffusion. *In Proceedings of the* IEEE *International Symposium on Computational Intelligence and Design (ISCID2023).* Program Website

PREPRINT

- Qinhong Zhou**, et al.** Virtual Community: An Open World for Humans, Robots, and Society. *under review at the International Conference on Learning Representations(ICLR 2026).* Program Website
- **Jiageng Liu*,** Weijie Lyu*, Xueting Li, Ming-Hsuan Yang. Edit3r: Instant 3D Scene Editing from Sparse Unposed Images. *under review at the CVPR 2026.* Program Website

RESEARCH EXPERIENCES

Multi-Modal Spatial Intelligence and Navigation Honda Research Institute USA, United States
Research Intern, Advisor: Nirav Savaliya 12/2025-Now

- I recently began a research internship at the Honda Research Institution.
- Our work primarily focuses on multi-modal scene representation and understanding, and is applied to indoor scene navigation within the Habitat AI simulation platform.

Instant 3D Scene Editing from Sparse Images (CVPR2026 under review) UC, Merced, United States
Research Leader, Advisor: Ming-Hsuan Yang 06/2025-11/2025

- Reconstructs and edits 3D scenes in a single feed-forward pass from unposed, view-inconsistent, instruction-edited images, without per-scene optimization or pose estimation.
- Uses a SAM2-based recoloring pipeline and an asymmetric multi-view input strategy to yield reliable, cross-view-consistent supervision and to encourage the network to fuse and align disparate observations from unposed images.
- Generalizes at inference to 2D instruction-edited inputs despite not being trained on such edits directly.
- Introduces DL3DV-Edit-Bench, achieving superior semantic alignment, consistency and speed on it over baselines.
- First-authored the paper currently under review at CVPR2026.

Test-Time Scaling with World Models for Spatial Reasoning (NeurIPS2025 accepted) Microsoft, United States
Research Leader, Advisor: Jianwei Yang 02/2025-06/2025

- State-of-the-art VLMs often fail on 3D spatial reasoning (e.g., perspective shifts). We address this by pairing a VLM with a controllable video world model that imagines egocentric views from a single image.
- Proposed test-time scaling via spatial beam search: the VLM sketches camera trajectories; the world model paints

corresponding views; the VLM stores helpful information as multi-view evidence step-by-step to analyze the answers.

- On the SAT benchmark,Our method works with four different VLMs (GPT-4o/4.1, InternVL3-14B, o1) and two world models (our SWM and Stable-Virtual-Camera), and can even surpass RL-scaled baselines when combined.
- Co-authored the paper that was accepted as poster at NeurIPS2025.

Virtual Community for Scalable 3D Scene and Embodied Agents (ICLR2026 under review) UMass, United States
Research Participant, Advisor: Chuang Gan 07/2024-10/2024

- The Motivation of Virtual Community is to feature large-scale community scenarios derived from the real world and construct a social world simulation platform designed to support embodied AI research.
- Scalable 3D Scene creation: which supports the generation of expansive outdoor and indoor environments at any location and scale, addressing the lack of a large-scale, interactive, open world scene for embodied AI research.
- Embodied agents with grounded characters and social relationship networks: the first to simulate socially connected agents at a community level, that also have scene-grounded characters.
- I take responsibility of multi-room and multi-scene indoor scene automatic generation, which based on Architect.
- Co-authored the paper that was under review at ICLR2026.

Lifting 2D Inpaint to Interactive 3D Scene Generation (NeurIPS2024 accepted) MIT-IBM Watson AI lab, United States
Research Leader, Advisor: Chuang Gan 11/2023-06/2024

- Aimming at create a general framework to generate sufficient and delicate interactive 3D scenes for Robotics and Embodied AI research, such as RL training.
- Created a framework name Architect, which utilizes pre-trained 2D image generative models to inpaint, and visual foundation model for perception to lift 2D image into 3D scene. Implement the inpainting iteratively and hierarchically.
- Architect is capable to generate infinity context-rich, fine-grind, interactive, indoor scenes.
- Co-authored the paper that was accepted as poster at NeurIPS2024.

LLM Commonsense Assisted Navigation (CVPR2024 accepted) MIT-IBM Watson AI lab, United States
Research Leader, Advisor: Chuang Gan 06/2023-11/2023

- Researched the spatial reasoning and multi-modal information processing capabilities of LLM.
- Utilized a zero-shot framework on Semantic Audio-Visual Navigation (SAVN) embodied AI task in simulation environment Habitat. Integrated LLM to imagine room layout, give high-level guidance into our framework Reflective and Imaginative Language Agent (RILA).
- Achieved results surpassed previous baselines (SOTA) that requires training, using LLM's commonsense.
- Co-authored the paper that was accepted as poster in CVPR2024.

Connecting Multi-modal Contrastive Representations (NeurIPS2023 accepted) Zhejiang University, China
Research Participant, Advisor: Zhou Zhao 02/2023-05/2023

- Conducted research on Multi-modal Contrastive Representation (MCR) learning, which aims to encode different modalities into a semantically aligned shared space.
- Devised a novel training-efficient method Connecting-MCR (C-MCR) for learning MCR without paired data.
- Designed a framework of connecting frozen CLIP and CLAP using our C-MCR to train an audio-visual model, dataset comes from six cross-modal. Analyzed the model on zero-shot downstream tasks comparing with SOTA methods.

HONORS & AWARDS

National Scholarship (Top 0.2%) Ministry of Education, China
Top ten raising star in College of CS College of Computer Science & Technology, Zhejiang University, China
First-class Scholarship (Top 3%) Zhejiang University, China
Learnings' Innovation Scholarship (Top 1.5%) College of Computer Science & Technology, Zhejiang University, China

SKILLS & INTERESTS

Research Interest: Spatial Intelligence, World Model, Computer Vision, 3D Generation, Embodied AI.
Skills: Python(mainly, pytorch, numpy, tensorflow), C++, C#, HTML, CSS, JavaScript, SQL, MATLAB.