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, First year ***GPA****:* 4.00/4.00 09/2024-01/2026 (expected)
Courses: Machine Learning (100), Theory and Practice of Software Engineering (100), Game Programming (100)

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
Courses: Natural Language Processing (97), Machine Learning (92), Computer Vision (90), Design of AI Product (96)

PUBLICATIONS

- 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).*

RESEARCH EXPERIENCES

Test-time Scaling for Spatial Inference in VQA Umass Amherst, United States
Research Leader 02/2025-04/2025
- Test-time scaling is a strategy to enhance a model's capability and generalization during inference—without retraining—aimed at improving performance on novel scenarios, complex tasks, or large-scale data.
- VLMs often lack sufficient inherent spatial reasoning ability to support effective test-time scaling in many embodied tasks. Recent advances in action-conditioned video generation models have effectively addressed this limitation.
- Action-conditioned video generation as a world model involves learning to predict future visual observations based on past frames and a sequence of actions. It can enhance the reasoning capability of VLMs.
- This project was conducted in collaboration with Microsoft Research.

Virtual Community for Scalable 3D Scene and Embodied Agents Umass Amherst, United States
Research Participant, Advisor: Chuang Gan 07/2024-12/2024
- Built on Genesis Platform, the motivation of Virtual Community is to feature large-scale community scenarios derived from the real world and construct a social world simulation 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.

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

- Built on **Genesis Platform**, the project aims to create a general framework for generating 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 in 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.
- Wrote a paper that was published to NeurIPS2023 with the title Connecting Multi-modal Contrastive Representations.

WordDiffuser: Help Cognize Character Using Diffusion (ISCID2023 accepted) Zhejiang University, China
*Research Leader, Advisor: **Lingyun Sun*** 03/2023-06/2023

- Proposed WordDiffuser, an AI-assisted and LLM-augmented framework generating images to help beginners understand the relationship between abstract characters and their meanings.
- Designed two stage image generation directed by Latent Diffusion Model (LDM). Utilized three loss functions to balance the structure, semantic consistence of the generated image and implemented stable diffusion for generation.
- Led the project and Summarized the project in a paper published to ISCID2023, an IEEE&EI conference.

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
Second prize in Chinese High School Mathematics League Chongqing, China

SKILLS & INTERESTS

Research Interest: Embody AI, 3D Generation, Generative Models, LLM Agent
Skills: Python (mainly PyTorch, NumPy, TensorFlow), C++, C#, HTML, CSS, JavaScript, SQL, MATLAB
Physical Simulators (**Genesis**, **Blender**, **Habitat**, **AI2-THOR**, **Gazebo**)
Hobby: Jogging, Cooking, Piano, PingPong