

# SHUYU GAN

(+86)17300987266 — derek.sygan@gmail.com — <https://shuyugan.github.io/>

## EDUCATION

### Zhejiang University

Sep. 2020 - Jun. 2024

Bachelor of Engineering with **Honors** in Computer Science, Turing Class, jointly cultivated by Chu Kochen Honors College and Department of Computer Science & Technology

GPA(overall): **3.94/4.00** (90.86/100)

### Core Coursework and Grades:

- *Mathematics*: Linear Algebra (98), Mathematical Analysis (95), Numerical Analysis (97), Convex Optimization(94)
- *Computer Science*: Fundamentals of Programming and Algorithms (99), Data Structures (96), Introduction to Theoretical Computer Science (98), Computer Architecture (95), Introduction to Cybersecurity (97), Introduction to Data Mining (96)

## RESEARCH INTERESTS

Computer Vision, Generative Models (e.g., Diffusion models for 2D image generation and editing), Multi-modal AI, Natural Language Processing, LLM/VLM-based Agent Systems.

## PUBLICATIONS

### Cultural Self-Adaptive Multimodal Gesture Generation Based on Multiple Culture Gesture Dataset

Jingyu Wu, Shi Chen, **Shuyu Gan**, Weijun Li, yang changyuan eric, Lingyun Sun

ACMMM 2023

### Agent S: An Open Agentic Framework that Uses Computers Like a Human

Saaket Agashe, Jiuzhou Han, **Shuyu Gan**, Jiachen Yang, Ang Li, Xin Eric Wang

Submitted to ICLR 2025

## RESEARCH EXPERIENCE

### Multi-cultural Co-Speech Gesture Generation

Feb. 2023 - May. 2023

Research Assistant@International Design Laboratory, Zhejiang University

Advisor: Prof. Lingyun Sun

- Tackled the noticable absence of cultural modalities in the multimodal co-speech generation field.
- Introduced the Multiple Culture Gesture Dataset(MCGD), consisting of 10 different cultures and 10000 segment sequences.
- Introduced a Cultural Self-adaptive Gesture Generation Network(CSGN) as well as a new indicator Culture Deception Rate (CDR) to examine the performance of culture-specific gesture generation.
- Achieved adaptive, culture-specific gesture generation with CSGN, improving the state-of-the-art average FGD from 53.7 to 48.0 and culture deception rate (CDR) from 33.63% to 39.87%, earning acceptance to present this work at ACMMM 2023.
- **Contribution**: Collected, processed and annotated the Dataset(MCGD); Implemented the text, audio encoder and culture characteristic extractor of CSGN; Proposed adding adversarial loss to the training objective; Evaluated the performance of other baseline models on MCGD.

### Refining Diffusion Model Loss with End-to-End Information

Jul. 2023 - Sept. 2023

Summer Intern@mlPC, University of California, San Diego

Advisor: Prof. Zhuowen Tu

- Injected conceptual guidance into Diffusion models' training process to improve the quality of generated images and avoid making semantic mistakes.
- Introduced a GAN to be concurrently trained with the Diffusion Model, leveraging the end-to-end conceptual guidance.
- Demonstrated via experiments that the unconditional transformer-based Diffusion model(DiT) generated images of higher fidelity, reducing the FID from 13.77 to 11.63 when trained with end-to-end conceptual GAN guidance on CelebA-HQ Dataset.
- **Contribution**: Trained the DiT baseline using the selected dataset; Fine-tuned the DiT with perceptual loss; Fine-tuned DiT using the GAN guidance

## Feature-level Image Editing via Diffusion Models

Research Assistant@ReLER, CCAI, Zhejiang University

Dec. 2023 - May. 2024

Advisor: Prof. Yi Yang

- Developed a feature-blending image editing approach for diffusion models. Used the UNets to predict noise at each step for both the edited and original images at matching noise levels. Blended their feature maps from each UNet block using a scaled mask and fed the result into the next block, achieving a more coherent edited noise predictions for high-quality image editing.
- Proved that this editing method can perform various editing tasks, including object replacement, background replacement and modification of object attributes.
- Experiments showed that feature level editing can generate edited images of high quality(measured by FID) and coherence at the editing region boundary.
- **Contribution:** Earned an excellent rating as the undergraduate graduation project.

## A Robust General-Purpose OS Agent

Summer Intern@ERIC Lab, University of California, Santa Cruz

Jun. 2024 - Present

Advisor: Prof. Xin Eric Wang

- Introduced Agent S, an MLLM-based agent that autonomously perform complex computer tasks.
- Proposed a new agentic framework which integrates experience-augmented hierarchical planning, self-supervised continual memory update, and an Agent-Computer Interface.
- Showed that on the OSWorld benchmark our agent outperforms the baseline by 9.37% on the success rate (a relative improvement of 83.6%) and achieved a new state-of-the-art.
- **Contribution:** Built and maintained the Agent S codebase; Designed the ID-Based ACI to perform accurate interactions; Incorporated Perplexica into the framework to facilitate online search capabilities; Implemented Tree Search and World Model strategies to improve the agent's effectiveness in real-time decision-making; Enhance agent's visual grounding by integrating auxiliary accessibility tree's information.

## AWARDS AND HONORS

- |  |                      |
|--|----------------------|
| • Second Prize in National College Mathematics Competition                   | 2021                 |
| • First-Prize Scholarship of Zhejiang University&Outstanding Student, Top 2% | 2022-2023            |
| • Zhejiang University's Top-notch Scholarship for Basic Disciplines, Top 5%  | 2020-2021, 2021-2022 |
| • Outstanding Undergraduate Graduate of Zhejiang University                  | 2024                 |
| • Honorary Certificate from Chu Kochen Honors College, Zhejiang University   | 2024                 |

## SKILLS

- |                            |   |
|----------------------------|---|
| <b>English Fluency</b>     | - TOEFL iBT 110 (Reading 29, Listening 28, Speaking 23, Writing 30)<br>- GRE Verbal 154, Quantitative 170, Analytical Writing 3.5   |
| <b>Computer Skills</b>     | - C/C++, Java, Python(PyTorch), Verilog, Assembly Language,<br>Linux(Ubuntu)/Windows OS operation   |
| <b>Economics Knowledge</b> | - Excellent academic performance in Macroeconomics and Microeconomics<br>earning scores of 95 and 96, respectively.<br>- Active participant in the KPMG Business Competition. |

## MISCELLANEOUS

- |                  |   |
|------------------|---|
| <b>Volunteer</b> | - Volunteer teacher at Bailuzhou Middle School, Ji'an, Jiangxi Province.<br>- Museum docent at Liangzhu Museum, Hangzhou, Zhejiang.<br>- Accumulated over <b>250 hours</b> of volunteer service experience. |
| <b>Sports</b>    | - Middle and long-distance running.<br>- Participated in 1500m race and Half Marathon at Zhejiang University  |