

# Sihan Ren

+1 5109184584    sihan\_ren@berkeley.edu

## Education

**University of California, Berkeley**, Visiting Student

Berkeley, California,  
United States  
Aug. 2024 - Jun. 2025

- **Major:** Computer Science
- Overall GPA: 4.0/4.0
- Related Coursework: LLM Agents, Computer Vision, Computer Graphics, Artificial Intelligence

**ShanghaiTech University**, Bachelor of Engineering

Shanghai, China  
Sept. 2022 - Present

- **Major:** Computer Science, **Minor:** Interactive Design
- Overall GPA: 3.76/4.0, Rank: 21/162 in CS Major
- Award: Merit Student (for top 15% of students), 2022-2023
- Related Coursework: Machine Learning, 3D Visual Concept Design(Blender), Unity Game Development, Critical Studies of VR Animation Expression, Character Animation

## Technologies

**Programming Skills:** Python(Proficient), C#(Proficient), C++(Familiar with), MATLAB(Familiar with), C(Familiar with)

**3D Software and Game Engine:** Blender(Proficient), Unity(Proficient)

**Misc:** Agentic Workflow Development, LLM Fine-tuning, Human Motion Generation, Human Motion Recognition

## Publications

**OMG: Towards Open-vocabulary Motion Generation via Mixture of Controllers**    [arXiv:\[2312.08985\]](#) 

Han Liang, Jiacheng Bao, Ruichi Zhang, **Sihan Ren**, Yuecheng Xu, Sibe Yang, Xin Chen, Jingyi Yu, Lan Xu

- **Accepted by CVPR 2024** (The IEEE/CVF Conference on Computer Vision and Pattern Recognition).
- **Research area:** Text-driven Human Motion Generation, text2motion, Motion Diffusion Model

**SCOPE: Sign Language Contextual Processing with Embedding from LLMs**    [arXiv:\[2409.01073\]](#) 

Yuqi Liu\*, Wenqian Zhang\*, **Sihan Ren**, Chengyu Huang, Jingyi Yu, Lan Xu

- **Accepted by AAAI 2025** (AAAI Conference on Artificial Intelligence).
- **Research area:** Sign Language Recognition/Translation, Large Language Models (LLMs), Contextual Understanding

## Research Experiences

**Virtual Reality and Visual Computing Center(VRVC)** | ShanghaiTech University | Mar. 2023 - Aug. 2024

*Undergraduate Research Intern*

- **Experience 1:** Text-Driven Human Motion Generation Using Mixture of Experts and Diffusion Models
- Processed SMPL-series 3D motion workflows and developed efficient Python data processing tools. Consolidated 20M segments of 3D motion data recorded using various methods and skeleton formats into a unified, learning-friendly format.
- Tested state-of-the-art motion generation algorithms such as MotionGPT and MLD using PyTorch. Conducted quantitative performance evaluations and systematically benchmarked algorithm effectiveness.
- Utilized Blender and motion retargeting tools to create high-quality human motion visualizations and renderings.
- Conducted comparative experiments on the HumanML3D and Mixamo datasets, demonstrating the superior performance of zero-shot text-driven motion generation models regarding FID, R-Precision, and Diversity metrics.
- **Experience 2:** Sign Language Recognition and Translation Based on Large Language Models (LLMs)
- Designed the pipeline, integrating contextual semantic information with 2D keypoint features using LLMs, leveraging transformer encoder architecture, improving the accuracy of sign language recognition and translation.
- Extracted high-precision 2D keypoints from sign language videos using DWPose, providing reliable input for the model.
- Fine-tuned the Qwen2 to integrate sign language vocabulary with contextual semantics, achieving high-quality translation.
- Developed and implemented algorithms using PyTorch, based on Transformer architectures, optimizing training efficiency with Distributed Data Parallel training.
- Conducted comparative experiments on datasets like CSL-Daily and Phoenix-2014T, achieving state-of-the-art translation performance on ROUGE and BLEU metrics.

## Project Experiences

---

**AI Nuggets: Agentic Research Assistant** | University of California, Berkeley | Feb. 2025 - Jun. 2025

- Developed agentic workflows using the OpenAI Agents SDK, integrating the MCP toolchain for scalable coordination between agents.
- Built a research investigation agent for the X platform using xAI APIs and Apify crawlers, demonstrating improved information extraction and analysis capabilities compared to Grok.
- Designed an academic knowledge graph building and leveraging pipeline using Neo4j to provide structured background information for the agent.
- Integrated the graph with other tools to improve the agent's understanding of research community, improving performance on tasks such as citation suggestion, collaborator recommendation, and literature retrieval.

**Motion Capture and Motion Prior** | University of California, Berkeley | Feb. 2025 - Jun. 2025 [Project Page](#)

- Proposed a novel pipeline for reconstructing human motion from noisy or occluded sequences by learning motion priors using diffusion models.
- Designed and implemented a transition-aware MotionVAE in PyTorch to model latent spaces for motion transitions, and trained a Transformer-based diffusion model to capture motion priors.
- Integrated DDIM-style inference and Score Distillation Sampling (SDS) to optimize latent transition representations, improving the plausibility and temporal consistency of reconstructed motions.
- Achieved superior performance over baselines on the AMASS dataset for occluded joint recovery and physically realistic motion generation; validated the effectiveness through ablation and comparative studies.

**3D Gaussian Splatting and Mesh extraction** | University of California, Berkeley | Feb. 2025 - Jun. 2025 [Project Page](#)

- Built an end-to-end pipeline from video input to 3D Gaussian Splatting training and final mesh extraction.
- Developed the system with Docker-based containerization to ensure reproducibility, ease of deployment, and system robustness.
- Modified and extended the LangSplat framework by integrating the CLIP model for text-conditioned extraction of object-specific point clouds from 3D Gaussian scenes.

**Multi-modal LLM Benchmark** | University of California, Berkeley | Aug. 2024 - Dec. 2024 [github](#)

- Won **3rd place** in the LLM Agents MOOC Hackathon, benchmark track.
- Designed a simulation environment using Python, with Google Street View APIs and Places APIs, creating realistic urban scenarios for agent testing.
- Built and tested multimodal LLM Agents using APIs from different companies, conducted comprehensive evaluations of over ten mainstream models, including GPT, Claude, and Gemini.
- Implemented WebUI visualization with Gradio. Displays visual and textual observations of the LLM agent, as well as reasoning processes, facilitating efficient experimental analysis.
- Evaluated various LLM models' performance in reasoning, spatial perception, and information integration, providing insights for understanding and improving LLM behavior in complex environments.

**Motion Generation Driven by Music** | Shanghai Tech University | Sep. 2023 - Jan. 2024

- Developed latent diffusion models within the PyTorch Lightning framework to generate music-driven human animation. Trained and tested the model on the AIST++ dataset, enhancing animation quality and synchronization with music.
- Led preliminary survey and algorithm implementation, optimized model training, and designed comparative experiments to evaluate performance.
- Created charts and showcased results, wrote the paper using LaTeX, rendered effect visualization in Blender, and presented experimental results.

## Work Experiences

---

**Shanghai DGene Digital** | Shanghai, China | Mar. 2023 - Aug. 2024

*R&D Engineer Intern*

- Built and trained sign language recognition models using PyTorch and developed a sign language translation demo application with Unity. Integrated the algorithm into mobile devices for real-time translation driven by camera data, improving communication efficiency in dental clinic scenarios.
- Developed a Gradio WebUI tool to optimize the recording and annotation workflow for sign language motion data, enhancing data collection and processing efficiency, and supporting seamless operations for deaf volunteers.
- Explored and deployed image generation algorithms, implementing style transfer for cartoon-styled 3D model animations using Comfy UI. Generated realistic avatar videos, expanding the application scenarios of sign language translation technology.