

# Mingdian Liu

241 Florence St, Sunnyvale CA 94086

☎ 515-357-3600 | 🏠 mingdianliu.github.io | ✉ lmdvigor@gmail.com | 🔗 linkedin.com/in/mingdian-liu-205804110 | 📄 Google Scholar

## Education

### Iowa State University

Ph.D. in Electrical Engineering, Advisor: Meng Lu and Jiming Song (IEEE Fellow)

Iowa, USA

Aug. 2020 - May 2024

### Iowa State University

M.Sc in Computer Science, Artificial Intelligence. Advisor: Carl K Chang (IEEE Fellow)

Iowa, USA

Jan. 2021 - May 2023

### Iowa State University

M.Eng. in Electrical Engineering, Microelectronics and Semiconductor (PhD program transferred)

Iowa, USA

Aug. 2017 - May 2020

### Shandong University

B.Sc in Physics

Jinan, China

Sep. 2012 - June 2016

## Skills

### Computer Vision (CV)

Generative Model, Multimodal Learning, Human Pose Estimation

### Natural Language Processing (NLP)

Large Language Model, LLM Finetuning, Retrieval Augmented Generation, LLM Agents

### Deep Learning Model

Diffusion Model, GPT, VQ-VAE, ResNet, RNN/LSTM, GAN, Transformer

### Programming Language

Python, Java, C/C++, R, Matlab

### Platform and Tools

Pytorch, TensorFlow/Keras, AWS, GCP, CUDA, Linux, Kubernetes

## Work Experience

### [CV][NLP] Video search engine and highlight moment generation [Website]

Sunnyvale, California

Research Scientist (startup founding team), Cutlabs.ai

Feb 2024 - Now

- Design and implement a video search engine that enables efficient understanding of audio, text, and images, achieving a processing time of 0.1x the input video duration.
- Spearhead the launch of a highlight moment recommendation system, capable of generating viral short clips from long streaming videos.
- Develop a highlight extraction pipeline using Retrieval-Augmented Generation (RAG) and Large Language Model (LLM) multi-agent techniques.
- Drive a significant increase in monthly active users from 1,000 to over 100,000, contributing to \$200K in Annual Recurring Revenue (ARR) as a result of these innovations.

### [CV][NLP] Branching GPT model for text to body motion, hand motion, and face motion generation [Webpage] [Demo] [Paper]

New York City, New York

Research Intern, Snap Research

May 2023 - Dec 2023

- Developed a pipeline to extract motion data for human body, hands, and face from wide videos, constructing a private Text-to-Motion-X dataset using Vision-Language models and GPT-4.
- Proposed a novel multimodal GPT-based model for Text-to-Motion-X generation, using joint space mapping to ensure consistency across body, hand, and face motions.
- Currently refining the model for deployment in Snapchat's 3D Bitmoji. Submitted two patents and one academic paper related to this work.

### [CV] Multimodal GPT model for Text-guided music to dance generation [Demo] [Slides]

Sunnyvale, California

Applied Scientist Intern, Amazon Alexa AI

May 2022 - Aug 2022

- Designed a Multimodal Learning baseline for music-to-dance generation using Transformer, extracting music features with *Librosa* and representing dance motions with a 6-dimensional rotation matrix in *SMPL*. Trained the model on a dataset of 1.2 million frames.
- Developed a novel model based on VQ-VAE (for dance embedding), MotionCLIP (to align dance embedding space with text), and GPT (to predict future dance motions), enabling text-guided dance editing.
- Our model demonstrated comparable results to state-of-the-art (SOTA) models in both motion quality and diversity. Submitted one patent.

### [CV] MobileNet model for smart watch-based hand gesture recognition system on AR glasses [Demo] [Patent] [News Report]

Palo Alto, California

Research Intern, OPPO US Research Center, XR Interaction Lab

June 2021 - Dec. 2021

- Prototyped a hand gesture recognition model for AR glasses using a modified MobileNetV2 architecture, trained on 0.4 million IMU and PPG sensor data. Contributed to the full product development lifecycle, from initial prototyping to Beta release.
- Enhanced model robustness (achieving 96% recall and 94% precision) and reduced power consumption by 35% compared to existing solutions, by integrating handcrafted signal features. The model has been successfully deployed in OPPO Air Glass. Submitted two patents as part of this work.
- Collaborated closely with Product Managers, Technical Program Managers, and Software Engineers, to provide key technical insights that influenced leadership decisions regarding the design of future products, including the next generation of OPPO Watch and Air Glass.

## Research Projects

## [NLP] Enhancing LLM Grounding for Multi-Agent Collaboration [Paper]

Sep. 2023 - Feb 2024

- Curated a synthetic dataset for multi-agent collaboration using GPT-4 and rule-based filtering. Fine-tuned the LLaMA 2 model with LoRA on this dataset.
- Implemented a Theory of Mind framework (predicting the next action of an agent) using retrieval-augmented generation for LLM agent collaboration. Experiments were conducted on fine-tuned LLaMA 7B, 14B, ChatGPT, and GPT-4.
- Results showed a 23.5% improvement in multi-agent collaboration performance compared to GPT-4.

## [CV] Antenna Design using Conditional Generative Adversarial Network [Paper]

Aug. 2021 - May 2023

- Developed an automated pipeline for model building and data collection on Ansys HFSS simulation software.
- Modified a conditional GAN to generate novel antenna designs matching specific performance criteria.
- Proposed an active learning framework, reducing the dataset by 29% while maintaining model accuracy.

## [CV] Indoor Activities Recognition with Radar Sensor [Homepage] [Demo] [Oral] [Code]

Oct. 2020 - Feb. 2022

- Deployed a TI mmWave sensor to generate real-time point clouds for human pose detection.
- Combined range-FFT and Doppler-FFT to track moving individuals, reducing average position error to 0.11 m.
- Applied BiLSTM for user identification, achieving 90.2% accuracy across 3 users.
- Voxelized point clouds within bounding boxes, extracted spatial features via CNN, and proposed a Transformer model for human pose estimation, achieving 92.3% classification accuracy.
- Deployed a sensor fusion system integrating mmWave sensors and depth cameras on NVIDIA Jetson for indoor activity monitoring.

## Selected Publications

---

**[Patent]** **“Generating Text-to-motion Animations From Partially Annotated Datasets.”** M. Liu, B. Zhou *Filed by Snap Research, under review*

**[Conference]** **“Enhancing Common Ground Alignment and Negotiation through Theory of Mind Modeling”** S. Qiu, M. Liu, et al. *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*

**[Conference]** **“T2M-X: Learning Expressive Text-to-Motion Generation from Partially Annotated Data.”** M. Liu, B. Zhou, et al. *Under Review*

**[Journal]** **“Using generative model for intelligent design of dielectric resonator antennas.”** M. Liu, H. Zhang, et al. *Microwave and Optical Technology Letters*

**[Patent]** **“Novel Finger Swipe-based Smart Watch User Interface Control.”** X. Li, M. Liu. *WO Patent 2022,221,781*

**[Journal]** **“An IoT-enabled paper sensor platform for real-time analysis of isothermal nucleic acid amplification tests.”** M. Liu, Y. Zhao, et al. *Biosensors and Bioelectronics*

**[Conference]** **“Generative adversarial network-based design of dielectric resonator antenna for mmWave 5G applications.”** M. Liu, H. Zhang, et al. *2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*