Mingdian Liu

227 Raphael Ave, Unit 16, Ames IA 50014

515-357-3600 | mingdianliu.github.io | lmdvigor@gmail.com | linkedin.com/in/mingdian-liu-205804110

Education

Iowa State University	*Iowa, USA*
Ph.D. in Electrical Engineering	*Sep. 2020 - Dec.2023*
Iowa State University	*Iowa, USA*
Master in Computer Science	*Jan. 2021 - May 2023*

Skills

Computer Vision (CV)	Generative Model, Multimodal Learning, Human Pose Estimation
Natural Language Processing (NLP)	LLM Finetuning, Retrieval Augmented Generation, AutoGPT
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

Work Experience

[CV][NLP] Conditional diffusion model for text to motion generation and retargeting *New York City, New York*

Research Intern, Snap Research *May 2023 - Dec 2023*

- Developed a conditional VAE model to unify motion data of different body shape.
- Modified GLIGEN model for share-aware text to motion generation.
- The generated motion can be retargeted to the body shape of any skeleton length.

[CV][NLP] Branching GPT model for text to body motion, hand motion, and face motion generation[webpage][demo] *New York City, New York*

Research Intern, Snap Research *May 2023 - Dec 2023*

- Built a pipeline for extracting the motion of human body, hand, and face from the wide videos. Constructed a private Text-to-Motion-X dataset by Vison-Language model and GPT-4.
- Firstly proposed a novel multimodal model based on the GPT model for text to Motion-X generation. Used joint space mapping to improve the consistency of the generative motion across human body, hand, and face. The paper was submitted to CVPR 2024.
- The model is under refinement for deployment in the 3D Bitmoji of Snapchat.

[CV] Mutimodal GPT model for Text-guided music to dance generation [demo] [slides] *Sunnyvale, California*

Applied Scientist Intern, Amazon Alexa AI *May 2022 - Aug 2022*

- Utilized Transformer as the backbone to design a Multimodal Learning baseline for music-to-dance generation. Extracted music features by *Librosa*. Presented dance motions by 6-dim rotation matrix in *SMPL*. Employed a dataset of 1.2 million frames for training model.
- Proposed a novel model based on VQ-VAE model (generate dance embedding), MotionCLIP (align dance embedding space to text embedding space), and GPT model (predict the future dance) to make dance motion editable by text guidance.
- The experiments showed our model was comparable to the SOTA model in both motion quality and motion diversity. One patent was submitted.

[CV] MobileNet model for smart watch-based hand gesture recognition system on AR glasses [demo] [patent] *Palo Alto, California*

Research Intern, OPPO US Research Center, XR Interaction Lab *June 2021 - Dec. 2021*

- Prototyped a hand gesture recognition model for AR glass using a modified MobileNetV2 trained with 0.4 million data of IMU and PPG sensors. Involved in the complete development processes including prototype, data collection, and Beta version release.
- Further improved model robustness (averaged 96% recall and 94% precision) and reduced 35% power consumption over existing methods by fusing handcrafted signal features into the model. Also submitted two patents on it.
- Composed 3-DoF hand tracking code for VR controller grip.

Research Projects

[NLP] Belief retrieval augmented generation for LLM agent collaboration *current*

- Finetoned LLaMA model by agent collaboration dataset.
- Implemented belief (the estimated next action of opposite agent) retrieval augmented generation for LLM agent collaborations. The experiments are run on finetoned-LLaMA 7B, 14B, ChatGPT, and GPT4
- The experiments show that the agent collaboration performance can be improved by 23.5%.

[CV] Antenna Design using Conditional Generative Adversarial Network [paper] *Aug. 2021 - May 2023*

- Developed a pipeline to automate model build-up and data collection on Ansys HFSS simulation software.
- Modified a conditional GAN model for generating novel antenna designs to match the requested antenna properties.
- Proposed an Active Learning framework to automatically to reduce 29% dataset while reaching the same accuracy of the old system.

[CV] Indoor Activities Recognition with Radar Sensor [homepage] [demo] [oral] [code]   *Oct. 2020 - Feb. 2022*

- Deployed TI mmWave sensor for real-time point cloud generation of human poses.
- Integrated range-FFT and Doppler-FFT for moving people tracking and reduced averaged position error to 0.11 m.
- Utilized BiLSTM for user identification with the accuracy of 90.2% for 3 people.
- Voxelized point cloud in bounding box and extracted spatial feature by CNN. Proposed a Transformer model for human pose estimation and reached a high classification accuracy of 92.3%.
- Deployed a sensor fusion system by mmWave sensor and depth camera on NVIDIA Jetson to monitor indoor human activities.

Publications

"*Theory-of-Mind Enhanced Dialogue Generation for Multi-Agent LLMs*" S. Qiu, M. Liu, *et al. Under Review*

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

"*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 Propagatio*

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

"*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*