Dongxing Mao

Homepage m962479949@gmail.com Google Scholar GitHub

Education

National University of Singapore *2021.08 – 2023.01*
M.Sc. in Electrical and Computer Engineering
- GPA: 4.25/5.00

Nanjing University of Science and Technology *2017.09 – 2021.06*
B.Eng. in Telecommunications Engineering
- GPA: 88.3/100

University of California, Los Angeles *2019.06 – 2019.08*
Summer Exchange Student

Research Interest

Unified Model, Multimodal LLMs (Experience in **Training on 256 GPUs**)

Publications

First-authored

- **Dongxing Mao**, Alex Jinpeng Wang, Jiawei Zhang, Weiming Han, Zhuobai Dong, Linjie Li, Yiqi Lin, Zhengyuan Yang, Libo Qin, Fuwei Zhang, Lijuan Wang and Min Li, "TextAtlas5M: A Large-scale Dataset for Dense Text Image Generation" . Dataset released with *65K+ downloads*, accepted at **ICML 2026**
- **Dongxing Mao**, Alex Jinpeng Wang, Jiahao Tang, Kevin Qinghong Lin, Linjie Li, Zhengyuan Yang, Lijuan Wang, Min Li and Jingru Tan, "Residual Decoder Adapter: ID-Preserving Tokenizer Adaption for Autoregressive Text Rendering", accepted at **CVPR 2026**
- **Dongxing Mao**, Yilin Wang, Linjie Li, Zhengyuan Yang, Alex Jinpeng Wang, "TextGround4M: A Prompt-Aligned Dataset for Layout-Aware Text Rendering", accepted at **AAAI 2026**

Co-authored

- Yilin Wang*, Xiangxi Zheng*, **Dongxing Mao**, Linjie Li, Ping Yu, Rui Yan, Yuan Yao, Zhengyuan Yang, Lijuan Wang, Alex Jinpeng Wang, " Efficient Frame Selection for Long Videos at Test Time with Early-Layer Attention", submitted to ECCV 2026
- Yuhao Zheng*, Hangyu Ran*, **Dongxing Mao**, Linjie Li, puzhen zhang, Guohao Li, Linyuan Lü, Philip Torr, Alex Jinpeng Wang, Kevin Qinghong Lin, " VidCode: Benchmarking Multimodal Code Generation for Video Animation", submitted to ECCV 2026
- Guanqiao Chen*, Jingru Tan*, **Dongxing Mao**, Libo Qin, Hu Jian Guo, Alex Jinpeng Wang, "DuetGen: Bridging Autoregressive and Diffusion Models for Text Render Image Generation", submitted to ECCV 2026
- Kevin Qinghong Lin*, Yuhao Zheng*, Hangyu Ran*, **Dongxing Mao**, Dantong Zhu, Linjie Li, Philip Torr, Alex Jinpeng Wang, "VCode: a Multimodal Coding Benchmark with SVG as Symbolic Visual Representation", preprint.
- Yilin Wang* , Heng Zhou*, **Dongxing Mao**, Linjie Li, Jingru Tan, Haochen Han, Zhengyuan Yang, Alex Jinpeng Wang, Min Li, "OR-PRM: A Process Reward Model for Algorithmic Problem in Operations Research", accepted at **ICLR 2026**
- Difei Gao, Lei Ji, Zechen Bai, Mingyu Ouyang, Peiran Li, **Dongxing Mao**, Qinchen Wu, Weichen Zhang, Peiyi Wang, Xiangwu Guo, Hengxu Wang, Luowei Zhou, Mike Zheng Shou, "AssistGUI: Task-oriented desktop graphical user interface automation", accepted at **CVPR 2024**
- Joya Chen, Zhaoyang Lv, Shiwei Wu, Kevin Qinghong Lin, Chenan Song, Difei Gao, Jia-Wei Liu, Ziteng Gao, **Dongxing Mao**, Mike Zheng Shou, "VideoLLM-online: Towards Large Video-Language Model for Streaming Video", accepted at **CVPR 2024**
- Stan Weixian Lei, Yuxuan Wang, **Dongxing Mao**, Difei Gao, Mike Zheng Shou, "AssistSR: Task-oriented Video Segment Retrieval for Personal AI Assistant", accepted at **EMNLP 2022**
- Benita Wong*, Joya Chen*, You Wu*, Stan Weixian Lei, **Dongxing Mao**, Difei Gao, Mike Zheng Shou, "AssistQ: Affordance-centric Question-driven Task Completion for Egocentric Assistant", accepted at **ECCV 2022**

Research Experience

Remote Research Intern, Microsoft Research, Seattle *2024.10 - Present*
Collaborative research with JPG Lab, Central South University Supervisor: Linjie Li

- Conduct research on text-centric multimodal generation, focusing on autoregressive image generation, visual tokenizers, and dense text rendering.
- Proposed Residual Decoder Adapter (RDA), an ID-preserving tokenizer adaptation method that improves text rendering in autoregressive models without changing the original token space; accepted at CVPR 2026.
- Led the construction of TextAtlas5M and TextGround4M, large-scale datasets for dense text image generation and prompt-grounded layout-aware text rendering; accepted at ICML 2026 and AAAI 2026.
- Built large-scale data processing, training, and evaluation pipelines for multimodal generation, including OCR-based evaluation, layout-aware metrics, and distributed training.

Research Intern, Showlab, National University of Singapore *2021.08 - 2024.09*
Supervisor: Mike Shou Zheng

- Conducted research on multimodal understanding and video-language models, with a focus on streaming, interactive, and assistant-oriented scenarios.
- Contributed to multiple systems and benchmarks, including AssistGUI, VideoLLM-online, AssistSR, and AssistQ.
- Worked on GUI automation, streaming video understanding, task-oriented video retrieval, and egocentric assistant reasoning.
- Co-authored papers accepted at CVPR 2024, EMNLP 2022, and ECCV 2022.

Services

- Reviewer: ECCV, AAAI, ICLR, CVPR.
- Organizer: CVPR 2023 Workshop-LOVEU.

Skills

- **Research Expertise**: Generative AI (Diffusion & Autoregressive Models), Multimodal Understanding
- **Deep Learning Frameworks**: PyTorch, HuggingFace (Diffusers, Transformers, Accelerate), DeepSpeed
- **Large-scale Computing**: Distributed Training, Linux, Performance Profiling (WandB, Tensorboard).
- **Programming Tools**: Python, C++, MATLAB, LaTeX, Git.
- **Languages**: Mandarin (Native), English (IELTS 7)