Sixun Dong

Email: sixundong.ai@gmail.com Mobile: (US) +1-323-904-5968 (CN) +86-136-4924-4805
Homepage: cv.ironieser.cc Github: github.com/Ironieser LinkedIn: sixun-dong

Research Profile

My research focuses on **Multimodal Generative AI**, **Efficient VLM/LLM Reasoning**, and **Agentic Systems**. I have published in CVPR, NeurIPS, AAAI, IJCAI, WACV, 3DV, and conducted GenAI Research Intern at Zoom(US).

Education

Present	Ph.D. Student in Computer Science, Arizona State University, US
July 2024	M.S. in Computer Science, ShanghaiTech University, China , Advisor: Prof. Shenghua Gao,
July 2020	B.E. in Process Equipment and Control Engineering, Dalian University of Technology, China
July 2020	B.E. (Dual Degree) in Computer Science, Dalian University of Technology, China

Selected Publications [My Google Scholar]

† Equal Contribution

Topic 1: Efficient Multimodal Understanding

1. *Under Review*. **Rethinking Model Efficiency: Multi-Agent Inference with Large Models.**
Sixun Dong, Juhua Hu, Steven Li, Wei Wen, Qi Qian.
2. *arXiv'2508*. **MMTok: Multimodal Coverage Maximization for Efficient Inference of VLMs.** *Zoom
Sixun Dong, Juhua Hu, Mian Zhang, Ming Yin, Yanjie Fu, Qi Qian. [Paper]

Topic 2: Visual Perception & Generative Models

1. NeurIPS 2025. **Sculpting Features from Noise: Reward-Guided Hierarchical Diffusion for Task-Optimal Feature Transformation.**
Nanxu Gong†, Zijun Li†, **Sixun Dong**, Haoyue Bai, Wangyang Ying, Xinyuan Wang, Yanjie Fu. [Paper]
2. *arXiv'2506*. **Teaching Time Series to See and Speak: Forecasting with Aligned Visual and Textual Perspectives.**
Sixun Dong, Wei Fan, Teresa Wu, Yanjie Fu. [Paper]
3. *Under Review*. **TimesFrame: Multi-Variable Time Series is a Video of Numerical Data.**
Sixun Dong, Nanxu Gong, Haoyue Bai, Xinyuan Wang, Wangyang Ying, Wei Fan, Yanjie Fu.
4. 3DV 2024. **RoomDesigner: Encoding Anchor-latents for Style-consistent and Shape-compatible Indoor Scene Generation.**
Yiqun Zhao, Zibo Zhao, Jing Li, **Sixun Dong**, Shenghua Gao. [Paper] [Code]
5. CVPR 2023. **Weakly Supervised Video Representation Learning with Unaligned Text for Sequential Videos.**
Sixun Dong†, Huazhang Hu†, Dongze Lian, Weixin Luo, Yicheng Qian, Shenghua Gao. [Paper] [Code]
6. 🏆 CVPR 2022 Oral. **TransRAC: Encoding Multi-scale Temporal Correlation with Transformers for Repetitive Action Counting.**
Huazhang Hu†, **Sixun Dong**†, Yiqun Zhao, Dongze Lian, Zhengxin Li, Shenghua Gao. [Paper] [Code]

Topic 3: LLM Agents, Reasoning & Evaluation

1. AAAI 2026. **Efficient Post-Training Refinement of Latent Reasoning in Large Language Models.**
Xinyuan Wang, Dongjie Wang, Wangyang Ying, Haoyue Bai, Nanxu Gong, **Sixun Dong**, Kunpeng Liu, Yanjie Fu. [Paper]
2. *Under Review*. **To Think or Not To Think, That is The Question for LLM Reasoning in Theory of Mind Tasks.**
Nanxu Gong, Haotian Li, **Sixun Dong**, Jianxun Lian, Yanjie Fu, Xing Xie.
3. *Under Review*. **Towards Robust Dysarthric Speech Recognition: LLM-Agent Post-ASR Correction Beyond WER.**
Xiuwen Zheng, **Sixun Dong**, Bornali Phukon, Mark Hasegawa-Johnson, Chang D. Yoo.
4. *arXiv'2508*. **LiveMCP-101: Stress Testing and Diagnosing MCP-enabled Agents on Challenging Queries.** *Zoom
Ming Yin, Dinghan Shen, Silei Xu, Jianbing Han, **Sixun Dong**, Mian Zhang, Yebowen Hu, Shujian Liu, Simin Ma, Song Wang, Sathish Reddy Indurthi, Xun Wang, Yiran Chen, Kaiqiang Song. [Paper]
5. *arXiv'2508*. **Complex Logical Instruction Generation.** *Zoom
Mian Zhang, Shujian Liu, **Sixun Dong**, Ming Yin, Yebowen Hu, Xun Wang, Steven Ma, Song Wang, Sathish Reddy Indurthi, Haoyun Deng, Zhiyu Zoey Chen, Kaiqiang Song. [Paper]

6. *arXiv'2505*. **Agentic Feature Augmentation: Unifying Selection and Generation with Teaming, Planning, and Memories.**
Nanxu Gong†, **Sixun Dong**†, Haoyue Bai, Xinyuan Wang, Wangyang Ying, Yanjie Fu. [Paper]
7. IJCAI 2025. **Unsupervised feature transformation via in-context generation, generator-critic llm agents, and duet-play teaming.**
Nanxu Gong, Xinyuan Wang, Wangyang Ying, Haoyue Bai, **Sixun Dong**, Haifeng Chen, Yanjie Fu. [Paper]
8. WACV 2024. **MLLM-Tool: A Multimodal Large Language Model For Tool Agent Learning.**
Chenyu Wang, Weixin Luo, **Sixun Dong**, Xiaohua Xuan, Zhengxin Li, Lin Ma, Shenghua Gao. [Paper] [Code]

Academic Community Service

Conference Reviewer: CVPR 2023-, ICCV 2023-, ECCV 2024-, ACM MM 2023-2024, ACCV 2024, KDD 2024
Journal Reviewer: TMM, Neural Networks, TKDD

Research Experience

Aug. 2025 | May 2025 **Zoom Communications** Seattle, WA, USA
GenAI Research Intern
Proposed **MMTok**, a training-free method for efficient VLM inference. Contributed to **LiveMCP-101** (Agent Evaluation & Diagnostics) and **Complex Logical Instruction Generation** (LLM Instruction Following Evaluation).

May 2025 | Aug. 2024 **Arizona State University (ASU)** Tempe, AZ, USA
Research Assistant
Proposed **"TimesCLIP"** (Under Review) for multimodal forecasting. Conducted research on **Generative Feature Transformation** and **Multi-Agent Inference**, resulting in publications at **NeurIPS 2025, AAAI 2026, and IJCAI 2025**.

Jan. 2024 | Aug. 2023 **DGene (Digital Human Algorithm Dept.)** Shanghai, China
Research Intern (Project Lead)
Led two key projects: 1) **Co-Speech Gesture Generation**: Addressed the lack of coordination between gestures and head movements. Developed a method to align co-speech gestures with head poses, significantly enhancing avatar realism. 2) **3D Body Reconstruction**: Built a pipeline taking multi-view RGB images to generate pose-controllable parameterized human bodies. Achieved **<7%** anthropometric error within **3 minutes** for business integration.

Aug. 2023 | Apr. 2023 **Transsion Holdings (Audio-Video Gen. Dept.)** Shanghai, China
Research Intern
Optimized audio-driven 2D talking head generation to resolve low realism and lip-sync issues. Proposed model architecture improvements and implemented facial restoration using **Gaussian image blending**. Fine-tuned models significantly outperformed industrial and academic SoTA benchmarks.

Technical Skills

Programming:	Python, Pytorch, C/C++, Linux, Git	
Research Topics:	Multimodal Understanding and Generation	Language Language Model Evaluation
	Efficient Vision Language Model	LLM-based Multi-Agent System