Sixun Dong

Email: sdong46@asu.edu

Homepage: cv.ironieser.cc

Mobile(US): +1-3239045968

Github: github.com/Ironieser

About Me

I am a Ph.D. student at KDD Lab, Arizona State University, supervised by Professor Yanjie Fu. I completed my Master's at ShanghaiTech University in SVIP-Lab under Professor Shenghua Gao, and my Bachelor's from Dalian University of Technology. My research focuses on video understanding, time series data, weakly supervised learning, multimodal learning, and vision large language models.

Education

Present	Ph.D. Student in Arizona State University, USA
	Focus: Multimodal Learning, Computer Vision, LLM Agent
July 2024	M.S. in Computer Science, ShanghaiTech University, China
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

Publications

$\dagger$ = Co-first author

1. *Under Review.* **Teaching Time Series to See and Speak: Forecasting with Aligned Visual and Textual Perspectives.**
   Sixun Dong, Wei Fan, Teresa Wu, Yanjie Fu.

2. *Under Review.* **TimesFrame: Multi-Variable Time Series is a Video of Numerical Data**
   Sixun Dong, Nanxu Gong, Haoyeu Bai, Xinyuan Wang, Wangyang Ying, Wei Fan, Yanjie Fu.

3. *Under Review.* **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]

4. *Under Review.* **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]

5. *Under Review.* **MECT: From Multimodal Knowledge Acquisition To Contrastive Embedding Construction For Generative Feature Transformation**
   Nanxu Gong, **Sixun Dong**, Haoyue Bai, Wangyang Ying, Yanjie Fu.

6. 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]

7. 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].

8. 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].

9. WACV 2024. **MLLM-Tool: A Multimodal Large Language Model For Tool Agent Learning.**
   Chenyu Wang, Weixin Luo, **Sixun Dong**, Xiaohua (Michael) Xuan, Zhengxin Li, Lin Ma, Shenghua Gao. [Paper] [Code].

10. 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]

Academic Community Service

Reviewer：CVPR 2023-2025, ICCV 2023-2025, ACM MM 2023-2025, ECCV 2024, ACCV 2024, KDD 2024, TMM, Neural Networks, TKDD

Internship Experience

May. 2025	**Efficient Vision-Language Modeling and Pretraining Alignment**
	GenAI Research Intern — Zoom Inc., GenAI Research Group, Seattle, WA, US
Present	1. Designing efficient token pruning strategies to accelerate VLM inference while maintaining alignment fidelity.
	2. Designing alignment strategies for better multimodal representation learning in pretraining.

Jan. 2024	**Co-Speech Gesture and Head Motion Generation**	
	Research Intern (Team Leader)	DGene, Digital Human Algorithm Department, Shanghai, China
Nov. 2023	Focused on improving digital avatar realism by integrating co-speech gestures with corresponding head motions. Developed techniques to enhance synchronization and quality of gesture and head motion in avatar generation.	

Oct. 2023	**Human Body Reconstruction and Anthropometric Measurements Based on Multi-view Camera Systems**	
	Research Intern (Team Leader)	DGene, Digital Human Algorithm Department, Shanghai, China
Aug. 2023	Led a project focusing on the development of 3D human body models from multi-view photographs for precise anthropometric measurements.	

· Developed an algorithm to generate parameterized human body models from images, ensuring accurate body measurements.
· Achieved a parameterized model capable of body pose adjustments with less than **7%** measurement error in under **3 minutes**.

Aug. 2023	**Audio Driven Talking Head Video Generation**	
	Research Intern (Team Leader)	Transsion Holdings, Audio-Video Generation Department, Shanghai, China
Apr. 2023	Led efforts to improve realism and lip-sync accuracy in talking head video generation through innovative algorithmic strategies.	

· Developed new optimization techniques for model training and architecture specifically for audio-driven talking head videos.
· Implemented facial restoration techniques, including advanced image blending, to enhance video output quality.
· Fine-tuned models achieved superior performance, surpassing existing SoTA in both commercial and academic benchmarks.

Technical Skills

Programming:	Python, Pytorch, C/C++, Linux, Git
Research Topics:	Video Understanding, Video & Motion Generation, Time Series Analysis
	Multimodal Learning, Weakly Supervised Learning, Contrastive Learning;