Zihui (Sherry) Xue

sherryxue@utexas.edu
https://zihuixue.github.io
GITHUB | SCHOLAR

Education	**The University of Texas at Austin**, Austin, USA	2020 - Present
	Ph.D. in Electrical and Computer Engineering	
	GPA 4.00/4.00, Advisor: Prof. Kristen Grauman	
	Fudan University, Shanghai, China	2016 - 2020
	B.E. in Electronic and Information Science and Technology	
	GPA: 3.66/4.00, Elite Engineering Program (top 5%)	
	University of California, Santa Barbara, Santa Barbara, USA	2018 - 2018
	UCEAP Exchange Program	
	GPA: 4.00/4.00, The College of Engineering Dean's Honors List	

Research Interests

Computer vision and machine learning
Video understanding,, multimodal learning

Experience	**Google DeepMind**	London, UK
	Student Researcher	May 2025 - Dec. 2025
	FAIR, Meta AI	Austin, USA
	Visiting Student Researcher	Jan. 2023 - Jan. 2025
	FAIR, Meta AI	New York, USA
	Research Intern to Dr. Lorenzo Torresani	May 2022 - Dec. 2022
	Shanghai Qi Zhi Institute	Shanghai, China
	Research Assistant to Prof. Hang Zhao	Aug. 2020 - Aug. 2021

Preprints

Seeing without Pixels: Perception from Camera Trajectories
arXiv, 2025
Zihui Xue, Kristen Grauman, Dima Damen, Andrew Zisserman, Tengda Han

Publications

Seeing the Arrow of Time in Large Multimodal Models
NeurIPS, 2025
Zihui Xue, Mi Luo, Kristen Grauman

When Thinking Drifts: Evidential Grounding for Robust Video Reasoning
NeurIPS, 2025
Mi Luo, **Zihui Xue**, Alex Dimakis, Kristen Grauman

Vid2Coach: Transforming How-To Videos into Task Assistants
The ACM Symposium on User Interface Software and Technology (UIST), 2025
Mina Huh, **Zihui Xue**, Ujjaini Das, Kumar Ashutosh, Kristen Grauman, Amy Pavel

REG: Rectified Gradient Guidance for Conditional Diffusion Models
ICML, 2025

Zhengqi Gao, Kaiwen Zha, Tianyuan Zhang, **Zihui Xue**, Duane S Boning

Progress-Aware Video Frame Captioning
CVPR, 2025
Zihui Xue, Joungbin An, Xitong Yang, Kristen Grauman

Viewpoint Rosetta Stone: Unlocking Unpaired Ego-Exo Videos for View-invariant Representation Learning
CVPR, 2025 (Oral)
Mi Luo, **Zihui Xue**, Alex Dimakis, Kristen Grauman

HOI-Swap: Swapping Objects in Videos with Hand-Object Interaction Awareness
NeurIPS, 2024
Zihui Xue, Mi Luo, Changan Chen, Kristen Grauman

Put Myself in Your Shoes: Lifting the Egocentric Perspective from Exocentric Videos
ECCV, 2024
Mi Luo, **Zihui Xue**, Alex Dimakis, Kristen Grauman

Action2Sound: Ambient-Aware Generation of Action Sounds from Egocentric Videos
ECCV, 2024 (Oral)
Changan Chen*, Puyuan Peng*, Ami Baid, **Zihui Xue**, Wei-Ning Hsu, David Harwath, Kristen Grauman

Learning Object State Changes in Videos: An Open-World Perspective
CVPR, 2024
Zihui Xue, Kumar Ashutosh, Kristen Grauman

Detours for Navigating Instructional Videos
CVPR, 2024 (Highlight)
Kumar Ashutosh, **Zihui Xue**, Tushar Nagarajan, Kristen Grauman

Ego-exo4d: Understanding skilled human activity from first-and third-person perspectives
CVPR, 2024 (Oral)
Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, ..., **Zihui Xue**, et al.

Learning Fine-grained View-Invariant Representations from Unpaired Ego-Exo Videos via Temporal Alignment
NeurIPS, 2023
Zihui Xue, Kristen Grauman

Egocentric Video Task Translation
CVPR, 2023 (Highlight)
Zihui Xue, Yale Song, Kristen Grauman, Lorenzo Torresani

The Modality Focusing Hypothesis: Towards Understanding Crossmodal Knowledge Distillation
ICLR, 2023 (Oral)
Zihui Xue*, Zhengqi Gao*, Sucheng Ren*, Hang Zhao

Dynamic Multimodal Fusion
CVPR MULA Workshop, 2023

Zihui Xue, Radu Marculescu

SUGAR: Efficient Subgraph-level Training via Resource-aware Graph Partitioning
IEEE Transactions on Computers, 2023
Zihui Xue, Yuedong Yang, Radu Marculescu

Anytime Depth Estimation with Limited Sensing and Computation Capabilities on Mobile Devices
CoRL, 2022
Yuedong Yang, **Zihui Xue**, Radu Marculescu

Co-advise: Cross Inductive Bias Distillation
CVPR, 2022
Sucheng Ren, Zhengqi Gao, Tianyu Hua, **Zihui Xue**, Yonglong Tian, Shengfeng He, Hang Zhao

What makes multi-modal learning better than single (provably)
NeurIPS, 2021
Yu Huang*, Chenzhuang Du*, **Zihui Xue**, Xuanyao Chen, Hang Zhao, Longbo Huang

On Feature Decorrelation in Self-Supervised learning
ICCV, 2021 (Oral)
Tianyu Hua*, Wenxiao Wang*, **Zihui Xue**, Sucheng Ren, Yue Wang, Hang Zhao

Multimodal Knowledge Expansion
ICCV, 2021
Zihui Xue, Sucheng Ren, Zhengqi Gao, Hang Zhao

Sampling graphlets of multiplex networks: a restricted random walk approach
ACM Transactions on the Web, 2021
Simiao Jiao, **Zihui Xue**, Xiaowei Chen, Yuedong Xu

Talks	Video Reasoning in Large Multimodal Models	Dec. 2025
	Invited talk at Multimodal Algorithmic Reasoning Workshop, NeurIPS 2025	
	Progress-Aware Video Frame Captioning	July 2025
	Invited talk at Weekly Reading Session, Runway	
	Invited talk at Visual Recognition course, UT Austin	
	Swapping Objects in Videos with Hand-Object Interaction Awareness	Feb. 2025
	Invited talk at Multimodal Weekly, TwelveLabs	
	Learning Fine-grained Activities from Videos	July 2024
	Invited talk at Stanford Medical AI and Computer Vision Lab	
	Learning Object State Changes in Videos: An Open-World Perspective	May 2024
	Invited talk at EAISeminar, FAIR, Meta	
	Egocentric Video Task Translation	Nov. 2022
	Invited talk at Egocentric Multimodal Activity Recognition Workshop, FAIR, Meta	

Professional Service	**Journal Reviewer**: IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Multimedia, International Journal of Computer Vision

Conference Reviewer: CVPR, ECCV, ICCV, ICML, NeurIPS, ICLR, CoRL

Honors & Awards		
	MIT EECS Rising Star	2025
	Graduate Engineering Council Travel Grant, UT Austin	2024
	1st Place in Talking-To-Me, Ego4D@ECCV 23 Challenge	2023
	3rd Place in PNR Keyframe Localization, Ego4D@ECCV 23 Challenge	2023
	Second Prize of the Scholarship for Outstanding Student, Fudan University	2020
	Nation Second Prize in China Undergraduate Mathematical Contest in Modeling	2018
	8/2115 (1st round) and 13/2116 (2nd round) in Alibaba Global Scheduling Algorithm	2018
	Tung OOCL Scholarship (the First Prize), Fudan University	2018
	Third Prize of the Scholarship for Outstanding Students, Fudan University	2017

Media Coverage

Meta AI Blog, Introducing Ego-Exo4D: A foundational dataset for research on video learning and multimodal perception   2024

MarkTechPost, Meet Ego-Exo4D: A Foundational Dataset and Benchmark Suite to Support Research on Video Learning and Multimodal Perception   2024

Meta AI Research, Advancing foundational research: overcoming challenges of video learning and multimodal perception   2023

Analytics India Magazine, Top 8 papers by Meta AI   2023

Meta AI Blog, Egocentric Video Task Translation   2023