

# Chao Huang

*Prepared on May. 2025*

<b>CONTACT</b>	Wegmans Hall, 2207 University of Rochester Rochester, NY, 14627	chuang65@cs.rochester.edu +1 (585) 910-9360 <a href="http://wikichao.github.io">http://wikichao.github.io</a>
<b>EDUCATION</b>	<b>University of Rochester</b> Ph.D. in Computer Science	Jan. 2021 – Present Rochester, NY
	<b>Nanjing University</b> B.Eng. in Electronic Science and Engineering	Sept. 2015 – Jun. 2019 Nanjing, China
<b>RESEARCH AREA</b>	Multimodal Learning and Generation, Audio-Visual Learning, Video Understanding	
<b>WORK EXPERIENCE</b>	<b>Meta Reality Labs Research</b> <i>Research Intern</i> Mentor: <i>Dr. Sanjeel Parekh, Dr. Ruohan Gao, Dr. Anurag Kumar</i>	May. 2024 – Aug. 2024 Cambridge, United Kingdom
	<b>Meta Reality Labs Research</b> <i>Research Intern</i> Mentor: <i>Dr. Dejan Markovic, Dr. Alexander Richard</i>	May. 2023 – Nov. 2023 Pittsburgh, PA
	<b>The Chinese University of Hong Kong</b> <i>Research Assistant</i> Mentor: <i>Prof. Chi-Wing Fu</i>	Jul. 2019 – Dec. 2020 Shatin, Hong Kong
<b>RESEARCH EXPERIENCE</b>	<b>University of Rochester</b> Advisor: Chenliang Xu	Jan. 2021 – Present Rochester, NY
	<b>Nanjing University</b> Advisor: Zhan Ma	Sept. 2018 – Jun. 2019 Nanjing, China
<b>PUBLICATIONS</b>	<ul style="list-style-type: none"><li>• Takusen: Asynchronous Temporal Modeling with Small-Large VLM for Streaming Dense Video Captioning. (<b>In Submission</b>)</li><li>• <b>Chao Huang</b>, Ruohan Gao, J. M. F. Tsang, Jan Kurcius, Cagdas Bilen, Chenliang Xu, Anurag Kumar, Sanjeel Parekh. Learning to Highlight Audio by Watching Movies. In <i>IEEE/CVF Conference on Computer Vision and Pattern Recognition</i>, 2025. (<b>CVPR</b>)</li><li>• <b>Chao Huang</b>, Susan Liang, Yapeng Tian, Anurag Kumar, and Chenliang Xu. High-Quality Audio-Visual Separation with Generative Diffusion Models. In <i>17th Asian Conference on Computer Vision</i>, 2024. (<b>ACCV</b>) <i>Best Paper, Runner-Up</i></li><li>• <b>Chao Huang</b>, Dejan Markovic, Chenliang Xu, Alexander Richard. Modeling and Driving Human Body Soundfields through Acoustic Primitives. In <i>The 18th European Conference on Computer Vision</i>, 2024. (<b>ECCV</b>)</li></ul>	

- **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Egocentric Audio-Visual Object Localization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. (CVPR)
- **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Audio-Visual Object Localization in Egocentric Videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022. (CVPR Workshop)
- **Chao Huang**, Haojie Liu, Tong Chen, Qiu Shen, and Zhan Ma. Extreme Image Compression via Multiscale Autoencoders with Generative Adversarial Optimization. In *IEEE Visual Communications and Image Processing*, 2019. **Oral Presentation** (VCIP)
- Yunlong Tang, Jing Bi, Siting Xu, ..., **Chao Huang**, ..., Ping Luo, Jiebo Luo, and Chenliang Xu. Video Understanding with Large Language Models: A Survey. In *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 2025.
- Yunlong Tang, Junjia Guo, ..., **Chao Huang**, Jing Bi, Zeliang Zhang, Pooyan Fazli, and Chenliang Xu. VIDCOMPOSITION: Can MLLMs Analyze Compositions in Compiled Videos? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. (CVPR)
- Susan Liang, **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Language-Guided Joint Audio-Visual Editing Via One-Shot Adaptation. In *17th Asian Conference on Computer Vision*, 2024. (ACCV)
- Susan Liang, **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Neural Acoustic Context Field: Rendering Realistic Room Impulse Response With Neural Fields. In *IEEE International Conference on Computer Vision Workshops*, 2023. (ICCV Workshop)
- Susan Liang, **Chao Huang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. AV-NeRF: Learning Neural Fields for Real-World Audio-Visual Scene Synthesis. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. (NeurIPS)
- Luchuan Song, Jing Bi, **Chao Huang**, and Chenliang Xu. Audio-visual action prediction with soft-boundary in egocentric videos. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2023. (CVPR Workshop)
- Xuefei Yan, David J Brady, Weiping Zhang, Changzhi Yu, Yulin Jiang, Jianqiang Wang, **Chao Huang**, Zian Li, Zhan Ma. Compressive Sampling for Array Cameras. *SIAM Journal on Imaging Science*, 2021. (SIIMS)

## ARXIV PREPRINTS

- **Chao Huang**, Susan Liang, Yunlong Tang, Li Ma, Yapeng Tian, and Chenliang Xu. FreSca: Scaling in Frequency Space Enhances Diffusion Models. In *arXiv preprint*, 2025.
- Yunlong Tang, Junjia Guo, Pinxin Liu, ..., **Chao Huang**, ..., Zeliang Zhang, and Chenliang Xu. Generative AI for Cel-Animation: A Survey. In *arXiv preprint*, 2025.
- **Chao Huang**, Susan Liang, Yunlong Tang, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Scaling Concept with Text-Guided Diffusion Models. In *arXiv preprint*, 2024.
- **Chao Huang\***, Ruihui Li\*, Xianzhi Li, and Chi-wing Fu. Non-local Part-Aware Point Cloud Denoising. *arXiv preprint*, 2020. (\* joint 1st authors)

## PATENTS

- “First-person audio-visual object localization systems and methods.” XU Chenliang, **Chao Huang**, Yapeng Tian, FNU Anurag Kumar - US Patent App. 18/599,398, 2024.
- “Neural radiance field systems and methods for synthesis of audio-visual scenes.” XU Chenliang, Susan Liang, **Chao Huang**, Yapeng Tian, FNU Anurag Kumar - US Patent App. 18/431,491, 2024.

## AWARDS

**Best Paper Award, Honorable Mention at ACCV 2024**

*for the paper “High-Quality Visually-Guided Sound Separation from Diverse Categories”*

## INVITED TALKS

**Egocentric Audio-Visual Localization**

*Invited paper talk at The Joint International 3rd Ego4D and 11th EPIC Workshop, CVPR 2023*

## SERVICES

**Reviewer:** AAAI, CVPR, ICCV, ACM MM, TMM, TIP, SIGGRAPH