

Chenliang Xu

Prepared on January 16, 2026

CONTACT	3005 Wegmans Hall 250 Hutchison Road University of Rochester Rochester, NY 14627	WORK EMAIL: chenliang.xu@rochester.edu PERSONAL EMAIL: chenliang.xu@outlook.com WEB: https://www.cs.rochester.edu/~cxu22
APPOINTMENTS	Associate Professor Department of Computer Science University of Rochester, Rochester, NY	7/2022 - PRESENT
	Affiliated Faculty Goergen Institute for Data Science and Artificial Intelligence University of Rochester, Rochester, NY	9/2016 - PRESENT
	Assistant Professor Department of Computer Science University of Rochester, Rochester, NY	9/2016 - 6/2022
RESEARCH INTERESTS	My research originates in computer vision and tackles interdisciplinary topics, including video understanding, audio-visual learning, vision and language, and methods for trustworthy AI.	
EDUCATION	Doctor of Philosophy Computer Science and Engineering University of Michigan, Ann Arbor, MI	2016
	Master of Science Computer Science and Engineering University at Buffalo, Buffalo, NY	2012
	Bachelor of Science Information and Computing Science Nanjing University of Aeronautics and Astronautics, Nanjing, China	2010
DISTINCTIONS	Edmund A. Hajim Outstanding Faculty Award	2025
	Best Paper Award Runner Up, ACCV <i>for the paper "High-Quality Visually-Guided Sound Separation from Diverse Categories"</i>	2024
	Cisco Faculty Award	2024
	Outstanding Speaker Award, The Physical Retail AI Workshop in WACV	2024
	James P. Wilmot Distinguished Professorship	2021
	University of Rochester Research Award	2021
	Best Paper Award, ACM SIGGRAPH VRCAI <i>for the paper "3D Human Avatar Digitization from A Single Image"</i>	2019
	Tencent Faculty Award	2018
	Best Paper Award, The Sound and Music Computing Conference <i>for the paper "Audio-Visual Source Association for String Ensembles"</i>	2017
	University of Rochester AR/VR Pilot Award	2017
	Best Open Source Code Third Prize, CVPR <i>for the paper "Evaluation of Super-voxel Methods for Early Video Processing"</i>	2012
	Best Demo Prize, The Greater New York Multimedia and Vision Meeting <i>for the demo "Streaming Hierarchical Video Segmentation"</i>	2012

2026

130. Ayoub Shahnazari, Zeliang Zhang, Sachith Dissanayake, Chenliang Xu, and Niaz Abdolrahim. Machine learning approaches for crystallographic classification from synthetic 2d x-ray diffraction data. *Journal of Applied Crystallography*, 2026

2025

129. Jiani Liu, Zhiyuan Wang, Zeliang Zhang, Chao Huang, Susan Liang, Yunlong Tang, and Chenliang Xu. Harnessing the computation redundancy in vits to boost adversarial transferability. In *Neural Information Processing Systems*, 2025. **(NeurIPS)**
128. Chao Huang, Yuesheng Ma, Junxuan Huang, Susan Liang, Yunlong Tang, Jing Bi, Wenqiang Liu, Nima Mesgarani, and Chenliang Xu. Zerosep: Separate anything in audio with zero training. In *Neural Information Processing Systems*, 2025. **(NeurIPS)**
127. Yunlong Tang, Pinxin Liu, Zhangyun Tan, Mingqian Feng, Rui Mao, Chao Huang, Jing Bi, Yunzhong Xiao, Susan Liang, Hang Hua, et al. Mmperspective: Do mllms understand perspective? a comprehensive benchmark for perspective perception, reasoning, and robustness. In *Neural Information Processing Systems Datasets and Benchmarks Track*, 2025. **(NeurIPS)**
126. Zidian Xie, Nanda Kishore Korrapolu, Amisha Dubey, Luchuan Song, Chenliang Xu, Karen M Wilson, AnaPaula Cupertino, and Dongmei Li. Leveraging large language models to identify engagement-driving features in vaping-related tiktok videos: Cross-sectional study. *Journal of Medical Internet Research*, 27:e76265, Nov 2025.
125. Yunlong Tang, Junjia Guo, Pinxin Liu, Zhiyuan Wang, Hang Hua, Jia-Xing Zhong, Yunzhong Xiao, Chao Huang, Luchuan Song, Susan Liang, Yizhi Song, Liu He, Jing Bi, Mingqian Feng, Xinyang Li, Zeliang Zhang, and Chenliang Xu. Generative ai for cel-animation: A survey. In *IEEE/CVF International Conference on Computer Vision Workshops*, 2025.
124. Susan Liang, Chao Huang, Yunlong Tang, Zeliang Zhang, and Chenliang Xu. p-avas: Can physics-integrated audio-visual modeling boost neural acoustic synthesis? In *IEEE/CVF International Conference on Computer Vision*, pages 13942–13951, 2025. **(ICCV)**
123. Pinxin Liu, Luchuan Song, Junhua Huang, Haiyang Liu, and Chenliang Xu. Gestureism: Latent shortcut based co-speech gesture generation with spatial-temporal modeling. In *IEEE/CVF International Conference on Computer Vision*, 2025. **(ICCV)**
122. Matthew Xie and Chenliang Xu. Predicting the risk of asthma development in youth using machine learning models. *PloS one*, 20(11):e0336591, 2025.
121. Zeliang Zhang, Xiaodong Liu, Hao Cheng, Chenliang Xu, and Jianfeng Gao. Diversifying the expert knowledge for task-agnostic pruning in sparse mixture-of-experts. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 86–102, 2025. **(ACL-Findings)**
120. Luchuan Song, Yang Zhou, Zhan Xu, Yi Zhou, Deepali Aneja, and Chenliang Xu. Streamme: Simplify 3d gaussian avatar within live stream. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Papers*, pages 1–10, 2025. **(SIGGRAPH)**
119. Susan Liang, Dejan Markovic, Israel D Gebru, Steven Krenn, Todd Keebler, Jacob Sandakly, Frank Yu, Samuel Hassel, Chenliang Xu, and Alexander Richard. Binauralflow: A causal and streamable approach for high-quality binaural speech synthesis with flow matching models. In *International Conference on Machine Learning*, 2025. **(ICML)**
118. Yunlong Tang, Jing Bi, Siting Xu, Luchuan Song, Susan Liang, Teng Wang, Daoan Zhang, Jie An, Jingyang Lin, Rongyi Zhu, Ali Vosoughi, Chao Huang, Zeliang Zhang, Feng Zheng, Jianguo Zhang, Ping Luo, Jiebo Luo, and Chenliang Xu. Video understanding with large language models: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. **(TCSVT)**

117. Zeliang Zhang, Gaowen Liu, Charles Fleming, Ramana Rao Kompella, and Chenliang Xu. Targeted forgetting of image subgroups in clip models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. **(CVPR)**
116. Yunlong Tang, Junjia Guo, Hang Hua, Susan Liang, Mingqian Feng, Xinyang Li, Rui Mao, 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)**
115. Chao Huang, Ruohan Guo, J. M. F. Tsang, Jan Kurcius, Cagdas Bilen, Chenliang Xu, Anurag Kumar, and Sanjeel Parekh. Learning to highlight audio by watching movies. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. **(CVPR)**
114. Jing Bi, Lianggong Bruce Wen, Zhang Liu, Junjia Guo, Yunlong Tang, Bingjie Wang, and Chenliang Xu. Unveilling visual perception in language models: An attention head analysis approach. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. **(CVPR)**
113. Zeliang Zhang, Susan Liang, Daiki Shimada, and Chenliang Xu. Rethinking audio-visual adversarial vulnerability from temporal and modality perspectives. In *International Conference on Learning Representations*, pages 1–16, 2025. **(ICLR)**
112. Pinxin Liu, Luchuan Song, Daoan Zhang, Hang Hua, Yunlong Tang, Huaijin Tu, Jiebo Luo, and Chenliang Xu. Gaussianstyle: Gaussian head avatar via stylegan. In *International Conference on 3D Vision*, pages 1–10, 2025.
111. Yunlong Tang, Daiki Shimada, Jing Bi, Mingqian Feng, Hang Hua, and Chenliang Xu. Empowering llms with pseudo-untrimmed videos for audio-visual temporal understanding. In *AAAI Conference on Artificial Intelligence*, pages 1–15, 2025. **(AAAI)**
110. Yunlong Tang, Gen Zhan, Li Yang, Yiting Liao, and Chenliang Xu. Cardiff: Video salient object ranking chain of thought reasoning for saliency prediction with diffusion. In *AAAI Conference on Artificial Intelligence*, pages 1–10, 2025. **(AAAI)**
109. Hang Hua, Yunlong Tang, Chenliang Xu, and Jiebo Luo. V2xum-llm: Cross-modal video summarization with temporal prompt instruction tuning. In *AAAI Conference on Artificial Intelligence*, pages 1–13, 2025. **(AAAI)**
- 2024**
108. Susan Liang, Chao Huang, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Language-guided audio-visual editing via one-shot adaptation. In *Asian Conference on Computer Vision*, pages 1011–1027, 2024. **(ACCV)**
107. Chao Huang, Susan Liang, Yapeng Tian, Anurag Kumar, and Chenliang Xu. High-quality visually-guided sound separation from diverse categories. In *Asian Conference on Computer Vision*, pages 35–49, 2024. **Best Paper, Runner-Up (ACCV)**
106. Zeliang Zhang, Zhuo Liu, Mingqian Feng, and Chenliang Xu. Can clip count stars? an empirical study on quantity bias in clip. In *Findings of the Conference on Empirical Methods in Natural Language Processing*, pages 1081–1086, 2024. **(EMNLP-Findings)**
105. Luchuan Song, Lele Chen, Celong Liu, Pinxin Liu, and Chenliang Xu. Texttoon: Real-time text toonify head avatar from single video. In *SIGGRAPH Asia*, pages 1–11, 2024. **(SIGGRAPH Asia)**
104. Jing Bi, Yunlong Tang, Luchuan Song, Ali Vosoughi, Nguyen Nguyen, and Chenliang Xu. Eagle: Egocentric aggregated language-video engine. In *ACM International Conference on Multimedia*, pages 1682–1691, 2024. **(Multimedia)**
103. Luchuan Song, Pinxin Liu, Lele Chen, Guojun Yin, and Chenliang Xu. Tri2-plane: Thinking head avatar via feature pyramid. In *European Conference on Computer Vision*, pages 1–20, 2024. **(ECCV)**

102. Chao Huang, Dejan Markovic, Chenliang Xu, and Alexander Richard. Modeling and driving human body soundfields through acoustic primitives. In *European Conference on Computer Vision*, pages 1–17, 2024. **(ECCV)**
101. Nguyen Nguyen, Jing Bi, Ali Vosoughi, Yapeng Tian, Pooyan Fazli, and Chenliang Xu. Oscar: Object state captioning and state change representation. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3565–3576, 2024. **(NAACL-Findings)**
100. Zeliang Zhang, Mingqian Feng, Zhiheng Li, and Chenliang Xu. Discover and mitigate multiple biased subgroups in image classifiers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10906–10915, 2024. **(CVPR)**
99. Rongyi Zhu, Zeliang Zhang, Susan Liang, Zhuo Liu, and Chenliang Xu. Learning to transform dynamically for better adversarial transferability. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24273–24283, 2024. **(CVPR)**
98. Ali Vosoughi, Shijian Deng, Yapeng Tian, Chenliang Xu, and Jiebo Luo. Cross modality bias in visual question answering: A causal view with possible worlds vqa. *IEEE Transactions on Multimedia*, 26:8609–8624, 2024. **(TMM)**
97. Zeliang Zhang, Wei Yao, Susan Liang, and Chenliang Xu. Random smooth-based certified defense against text adversarial attack. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1251–1265, 2024. **(EACL-Findings)**
96. Jinyang Jiang, Zeliang Zhang, Chenliang Xu, Zhaofei Yu, and Yijie Peng. One forward is enough for neural network training via likelihood ratio method. In *International Conference on Learning Representations*, pages 1–24, 2024. **(ICLR)**
95. Luchuan Song, Pinxin Liu, Guojun Yin, and Chenliang Xu. Adaptive super resolution for one-shot talking head generation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4115–4119, 2024. **(ICASSP)**
94. Ali Vosoughi, Luca Bondi, Ho-Hsiang Wu, and Chenliang Xu. Learning audio concepts from counterfactual natural language. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 366–370, 2024. **(ICASSP)**
93. Zidian Xie, Shijian Deng, Pinxin Liu, Xubin Lou, Chenliang Xu, and Dongmei Li. Characterizing anti-vaping posts for effective communication on instagram using multimodal deep learning. *Nicotine and Tobacco Research*, 26(Supplement.1):S43–S48, 2024.
- 2023**
92. Jerardo E Salgado, Samuel Lerman, Zhaotong Du, Chenliang Xu, and Niaz Abdolrahim. Automated classification of big x-ray diffraction data using deep learning models. *npj Computational Materials*, 9(1):214, 2023. **(npj)**
91. Susan Liang, Chao Huang, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Av-nerf: Learning neural fields for real-world audio-visual scene synthesis. In *Neural Information Processing Systems*, pages 1–19, 2023. **(NeurIPS)**
90. Luchuan Song, Guojun Yin, Zhenchao Jin, Xiaoyi Dong, and Chenliang Xu. Emotional listener portrait: Neural listener head generation with emotion. In *IEEE/CVF International Conference on Computer Vision*, pages 20839–20849, 2023. **(ICCV)**
89. Zhiheng Li, Ivan Evtimov, Albert Gordo, Caner Hazirbas, Tal Hassner, Cristian Canton Ferrer, Chenliang Xu, and Mark Ibrahim. A whac-a-mole dilemma: Shortcuts come in multiples where mitigating one amplifies others. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20071–20082, 2023. **(CVPR)**
88. Chao Huang, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Egocentric audio-visual object localization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 22910–22921, 2023. **(CVPR)**

87. Zheng Zhang, Zheng Ning, Chenliang Xu, Yapeng Tian, and Toby Jia-Jun Li. Peanut: A human-ai collaborative tool for annotating audio-visual data. In *ACM Symposium on User Interface Software and Technology*, pages 1–18, 2023. **(UIST)**
86. Joseph S. German, Guofeng Cui, Chenliang Xu, and Robert A. Jacobs. Rapid runtime learning by curating small datasets of high-quality items obtained from memory. *PLOS Computational Biology*, 19(10):1–32, 2023.
85. 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/CVF International Conference on Computer Vision Workshops*, pages 1–4, 2023.
84. Jing Bi, Nguyen Nguyen, Ali Vosoughi, and Chenliang Xu. Misar: A multimodal instructional system with augmented reality. In *IEEE/CVF International Conference on Computer Vision Workshops*, pages 1–4, 2023.
83. Yiyang Su, Ali Vosoughi, Shijian Deng, Yapeng Tian, and Chenliang Xu. Separating invisible sounds toward universal audio-visual scene-aware sound separation. In *IEEE/CVF International Conference on Computer Vision Workshops*, pages 1–4, 2023.
82. Susan Liang, Chao Huang, Yapeng Tian, Anurag Kumar, and Chenliang Xu. Av-nerf: Learning neural fields for real-world audio-visual scene synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–4, 2023.
81. 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*, pages 1–4, 2023.
- 2022**
80. Zhiheng Li, Anthony Hoogs, and Chenliang Xu. Discover and mitigate unknown biases with debiasing alternate networks. In *European Conference on Computer Vision*, pages 270–288, 2022. **(ECCV)**
79. Luchuan Song, Xiaodan Li, Zheng Fang, Zhenchao Jin, Yuefeng Chen, and Chenliang Xu. Face forgery detection via symmetric transformer. In *ACM International Conference on Multimedia*, pages 4102–4111, 2022. **(Multimedia)**
78. Zhiheng Li, Martin Renqiang Min, Kai Li, and Chenliang Xu. Stylet2i: Toward compositional and high-fidelity text-to-image synthesis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18197–18207, 2022. **(CVPR)**
77. Jing Shi, Ning Xu, Haitian Zheng, Alex Smith, Jiebo Luo, and Chenliang Xu. Spaceedit: Learning a unified editing space for open-domain image editing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19730–19739, 2022. **(CVPR)**
76. Guangyao Li, Yake Wei, Yapeng Tian, Chenliang Xu, Ji-Rong Wen, and Di Hu. Learning to answer questions in dynamic audio-visual scenarios. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19108–19118, 2022. **Oral Presentation (CVPR)**
75. Guangyuan Li, Jun Lv, Yapeng Tian, Dou Qi, Chengyan Wang, Chenliang Xu, and Jing Qin. Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast mri super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20636–20645, 2022. **(CVPR)**
74. Zhengyuan Yang, Jingen Liu, Jing Huang, Xiaodong He, Tao Mei, Chenliang Xu, and Jiebo Luo. Cross-modal contrastive distillation for instructional activity anticipation. In *International Conference on Pattern Recognition*, pages 5002–5009, 2022. **(ICPR)**
73. Tong Shan, Casper E. Wenner, Chenliang Xu, Zhiyao Duan, and Ross K. Maddox. Speech-in-noise comprehension is improved when viewing a deep-neural-network-generated talking face. *Trends in Hearing*, 26, 2022.

72. 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*, pages 1–4, 2022.

2021

71. Jing Bi, Jiebo Luo, and Chenliang Xu. Procedure planning in instructional videos via contextual modeling and model-based policy learning. In *IEEE/CVF International Conference on Computer Vision*, pages 15611–15620, 2021. **Oral Presentation (ICCV)**
70. Samuel Lerman, Charles Venuto, Henry Kautz, and Chenliang Xu. Explaining local, global, and higher-order interactions in deep learning. In *IEEE/CVF International Conference on Computer Vision*, pages 1224–1233, 2021. **(ICCV)**
69. Zhiheng Li and Chenliang Xu. Discover the unknown biased attribute of an image classifier. In *IEEE/CVF International Conference on Computer Vision*, pages 14970–14979, 2021. **(ICCV)**
68. Jing Shi, Yiwu Zhong, Ning Xu, Yin Li, and Chenliang Xu. A simple baseline for weakly-supervised scene graph generation. In *IEEE/CVF International Conference on Computer Vision*, pages 16393–16402, 2021. **(ICCV)**
67. Yiwu Zhong, Jing Shi, Jianwei Yang, Chenliang Xu, and Yin Li. Learning to generate image scene graph from natural language supervision. In *IEEE/CVF International Conference on Computer Vision*, pages 1823–1834, 2021. **(ICCV)**
66. Yapeng Tian and Chenliang Xu. Can audio-visual integration strengthen robustness under multimodal attacks? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5601–5611, 2021. **(CVPR)**
65. Yapeng Tian, Di Hu, and Chenliang Xu. Cyclic co-learning of sounding object visual grounding and sound separation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2745–2754, 2021. **(CVPR)**
64. Lele Chen, Chen Cao, Fernando De la Torre, Jason Saragih, Chenliang Xu, and Yaser Sheikh. High-fidelity face tracking for ar/vr via deep lighting adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13059–13069, 2021. **(CVPR)**
63. Jing Shi, Ning Xu, Yihang Xu, Trung Bui, Franck Deroncourt, and Chenliang Xu. Learning by planning: Language-guided global image editing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13590–13599, 2021. **(CVPR)**
62. Sizhe Li, Yapeng Tian, and Chenliang Xu. Space-time memory network for sounding object localization in videos. In *British Machine Vision Conference*, pages 1–13, 2021. **(BMVC)**
61. Ziyi Kou, Guofeng Cui, Shaojie Wang, Wentian Zhao, and Chenliang Xu. Improve cam with auto-adapted segmentation and co-supervised augmentation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3598–3606, 2021. **(WACV)**
60. Shaojie Wang, Wentian Zhao, Ziyi Kou, Jing Shi, and Chenliang Xu. How to make a blt sandwich? learning vqa towards understanding web instructional videos. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1130–1139, 2021. **(WACV)**
59. Zhong Li, Lele Chen, Celong Liu, Fuyao Zhang, Zekun Li, Yu Gao, Yuanzhou Ha, Chenliang Xu, Shuxue Quan, and Yi Xu. Animated 3d human avatars from a single image with gan-based texture inference. *Computers & Graphics*, 95:81–91, 2021
58. Yankun Gao, Zidian Xie, Li Sun, Chenliang Xu, and Dongmei Li. Characteristics and user engagement of anti-vaping posts on instagram: Observational study. *JMIR Public Health and Surveillance*, 7(11):e29600, 2021.

2020

57. Yapeng Tian, Dingzeyu Li, and Chenliang Xu. Unified multisensory perception: Weakly-supervised audio-visual video parsing. In *European Conference on Computer Vision*, pages 436–454, 2020. **Spotlight Presentation (ECCV)**
56. Lele Chen, Guofeng Cui, Celong Liu, Zhong Li, Ziyi Kou, Yi Xu, and Chenliang Xu. Talking-head generation with rhythmic head motion. In *European Conference on Computer Vision*, pages 35–51, 2020. **(ECCV)**
55. Jie Chen, Zhiheng Li, Jiebo Luo, and Chenliang Xu. Learning a weakly-supervised video actor-action segmentation model with a wise selection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9901–9911, 2020. **Oral Presentation (CVPR)**
54. Zhiheng Li, Wenxuan Bao, Jiayang Zheng, and Chenliang Xu. Deep grouping model for unified perceptual parsing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4053–4063, 2020. **(CVPR)**
53. Xiaoyu Xiang, Yapeng Tian, Yulun Zhang, Yun Fu, Jan Allebach, and Chenliang Xu. Zooming slow-mo: Fast and accurate one-stage space-time video super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3370–3379, 2020. **(CVPR)**
52. Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. Tdan: Temporally deformable alignment network for video super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3360–3369, 2020. **(CVPR)**
51. Jing Bi, Vikas Dhiman, Tianyou Xiao, and Chenliang Xu. Learning from interventions using hierarchical policies for safe learning. In *AAAI Conference on Artificial Intelligence*, pages 10352–10360, 2020. **Oral Presentation (AAAI)**
50. Jing Shi, Ning Xu, Trung Bui, Franck Deroncourt, Zheng Wen, and Chenliang Xu. A benchmark and baseline for language-driven image editing. In *Asian Conference on Computer Vision*, pages 636–651, 2020. **Oral Presentation (ACCV)**
49. Lele Chen, Justin Tian, Guo Li, Cheng-Haw Wu, Erh-Kan King, Kuan-Ting Chen, Shao-Hang Hsieh, and Chenliang Xu. Tailorgan: Making user-defined fashion designs. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3241–3250, 2020. **(WACV)**
48. Sefik Emre Eskimez, Ross K. Maddox, Chenliang Xu, and Zhiyao Duan. End-to-end generation of talking faces from noisy speech. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 1948–1952, 2020. **(ICASSP)**
47. Haitian Zheng, Lele Chen, Chenliang Xu, and Jiebo Luo. Pose flow learning from person images for pose guided synthesis. *IEEE Transactions on Image Processing*, 30:1898–1909, 2020. **(TIP)**
46. Baojie Fan, Yang Cong, Yandong Tang, Jiandong Tian, and Chenliang Xu. Structured and consistent multi-layer multi-kernel subtask correction filter tracker. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(6):2328–2342, 2020. **(TCSVT)**
45. Yankun Gao, Zidian Xie, Li Sun, Chenliang Xu, and Dongmei Li. Electronic cigarette-related contents on instagram: Observational study and exploratory analysis. *JMIR Public Health and Surveillance*, 6(4):e21963, 2020.
44. Zhiheng Li, Geemi P. Wellawatte, Maghesree Chakraborty, Heta A. Gandhi, Chenliang Xu, and Andrew D. White. Graph neural network based coarse-grained mapping prediction. *Chemical Science*, 11(9524-9531), 2020.
43. Yapeng Tian, Di Hu, and Chenliang Xu. Co-learn sounding object visual grounding and visually indicated sound separation in a cycle. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–4, 2020.
42. Yapeng Tian, Chenliang Xu, and Dingzeyu Li. Deep audio prior: Learning sound source separation from a single audio mixture. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–4, 2020.

41. Yapeng Tian, Dingzeyu Li, and Chenliang Xu. Weakly-supervised audio-visual video parsing toward unified multisensory perception. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–4, 2020.
40. Lele Chen, Guofeng Cui, Ziyi Kou, Haitian Zheng, and Chenliang Xu. What comprises a good talking-head video generation? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–4, 2020.

2019

39. Zhong Li, Lele Chen, Celong Liu, Yu Gao, Yuanzhou Ha, Chenliang Xu, Shuxue Quan, and Yi Xu. 3d human avatar digitization from a single image. In *ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 1–8, 2019.
Best Paper Award (VRCAI)
38. Jing Shi, Jia Xu, Boqing Gong, and Chenliang Xu. Not all frames are equal: Weakly-supervised video grounding with contextual similarity and visual clustering losses. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10444–10452, 2019. **(CVPR)**
37. Lele Chen, Ross K. Maddox, Zhiyao Duan, and Chenliang Xu. Hierarchical cross-modal talking face generation with dynamic pixel-wise loss. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7832–7841, 2019. **(CVPR)**
36. Wentian Zhao, Shaojie Wang, Zhihuai Xie, Jing Shi, and Chenliang Xu. Gan-em: Gan based em learning framework. In *International Joint Conference on Artificial Intelligence*, pages 4404–4411, 2019. **(IJCAI)**
35. Hao Huang, Luowei Zhou, Wei Zhang, Jason J. Corso, and Chenliang Xu. Dynamic graph models for modeling higher-order interactions in activity recognition. In *British Machine Vision Conference*, pages 1–12, 2019. **(BMVC)**
34. Yawen Lu, Sophia Kourian, Carl Salvaggio, Chenliang Xu, and Guoyu Lu. Single image 3d vehicle pose estimation for augmented reality. In *IEEE Global Conference on Signal and Information Processing*, pages 1–5, 2019.
33. Yan Yan, Chenliang Xu, Dawen Cai, and Jason J. Corso. A weakly supervised multi-task ranking framework for actor-action semantic segmentation. *International Journal of Computer Vision*, 128:1414–1432, 2019. **(IJCV)**
32. Robert A. Jacobs and Chenliang Xu. Can multisensory training aid visual learning?: A computational investigation. *Journal of Vision*, 19(11):1–12, 2019.
31. Sefik Emre Eskimez, Ross K. Maddox, Chenliang Xu, and Zhiyao Duan. Noise-resilient training method for face landmark generation from speech. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 28(1):27–38, 2019.
30. Bochen Li, Karthik Dinesh, Chenliang Xu, Gaurav Sharma, and Zhiyao Duan. Online audio-visual source association for chamber music performances. *Transactions of the International Society for Music Information Retrieval*, 2(1):29–42, 2019.
29. Yapeng Tian, Jing Shi, Bochen Li, Zhiyao Duan, and Chenliang Xu. Audio-visual event localization in the wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 5–8, 2019.
28. Yapeng Tian, Chenxiao Guan, Justin Goodman, Marc Moore, and Chenliang Xu. Audio-visual interpretable and controllable video captioning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 9–12, 2019.
27. Lele Chen, Haitian Zheng, Ross K. Maddox, Zhiyao Duan, and Chenliang Xu. Sound to visual: Hierarchical cross-modal talking face video generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–4, 2019.

2018

26. Yapeng Tian, Jing Shi, Bochen Li, Zhiyao Duan, and Chenliang Xu. Audio-visual event localization in unconstrained videos. In *European Conference on Computer Vision*, pages 252–268, 2018. **(ECCV)**
25. Lele Chen, Zhiheng Li, Ross K. Maddox, Zhiyao Duan, and Chenliang Xu. Lip movements generation at a glance. In *European Conference on Computer Vision*, pages 538–553, 2018. **(ECCV)**
24. Li Ding and Chenliang Xu. Weakly-supervised action segmentation with iterative soft boundary assignment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6508–6516, 2018. **(CVPR)**
23. Luowei Zhou, Chenliang Xu, and Jason J. Corso. Towards automatic learning of procedures from web instructional videos. In *Proc. of AAAI Conference on Artificial Intelligence*, pages 7590–7598, 2018. **Oral Presentation (AAAI)**
22. Tianlang Chen, Chenliang Xu, and Jiebo Luo. Improving text-based person search by spatial matching and adaptive threshold. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1879–1887, 2018. **(WACV)**
21. Lele Chen, Yue Wu, Adora M. DSouza, Anas Z. Abidin, Axel Wismüller, and Chenliang Xu. Mri tumor segmentation with densely connected 3d cnn. In *SPIE Medical Imaging 2018: Image Processing*, volume 105741, pages 105741F1–8, 2018. **Oral Presentation (SPIE)**
20. Sefik Emre Eskimez, Ross K. Maddox, Chenliang Xu, and Zhiyao Duan. Generating talking face landmarks from speech. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 372–381, 2018.
19. Lele Chen, Sefik Emre Eskimez, Zhiheng Li, Zhiyao Duan, Chenliang Xu, and Ross K. Maddox. Toward a visual assistive listening device: Real-time synthesis of a virtual talking face from acoustic speech using deep neural networks. *The Journal of the Acoustical Society of America*, 143(3):1813–1813, 2018.
18. Maghesree Chakraborty, Chenliang Xu, and Andrew D. White. Encoding and selecting coarse-grain mapping operators with hierarchical graphs. *The Journal of Chemical Physics*, 149(13):134106, 2018.

2017

17. Bochen Li, Chenliang Xu, and Zhiyao Duan. Audio-visual source association for string ensembles through multi-modal vibrato analysis. In *Sound and Music Computing Conference*, pages 159–166, 2017. **Best Paper Award (SMC)**
16. Yan Yan, Chenliang Xu, Dawen Cai, and Jason J. Corso. Weakly supervised actor-action segmentation via robust multi-task ranking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1022–1031, 2017. **(CVPR)**
15. Tingting Han, Hongxun Yao, Chenliang Xu, Xiaoshuai Sun, Yanhao Zhang, and Jason J. Corso. Dancelets mining for video recommendation based on dance styles. *IEEE Transactions on Multimedia*, 19(4):712–724, 2017. **(TMM)**
14. Amos Newswanger and Chenliang Xu. One-shot video object segmentation with iterative online fine-tuning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–5, 2017.
13. Lele Chen, Sudhanshu Srivastava, Zhiyao Duan, and Chenliang Xu. Deep cross-modal audio-visual generation. In *ACM International Conference on Multimedia Thematic Workshops*, pages 349–357, 2017.
12. Luowei Zhou, Chenliang Xu, Parker Koch, and Jason J. Corso. Watch what you just said: Image captioning with text-conditional attention. In *ACM International Conference on Multimedia Thematic Workshops*, pages 305–313, 2017.

2016 and Before

11. Chenliang Xu and Jason J. Corso. Actor-action semantic segmentation with grouping process models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3083–3092, 2016. **Spotlight Presentation (CVPR)**
10. Chenliang Xu and Jason J. Corso. Libsvx: A supervoxel library and benchmark for early video processing. *International Journal of Computer Vision*, 119(3):272–290, 2016. **(IJCV)**
9. Chenliang Xu, Shao-Hang Hsieh, Caiming Xiong, and Jason J. Corso. Can humans fly? action understanding with multiple classes of actors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2264–2273, 2015. **(CVPR)**
8. Chenliang Xu, Richard F. Doell, Stephen Jose Hanson, Catherine Hanson, and Jason J. Corso. A study of actor and action semantic retention in video supervoxel segmentation. *International Journal of Semantic Computing*, 7(4):353–375, 2013. **(IJSC)**
7. Chenliang Xu, Spencer Whitt, and Jason J. Corso. Flattening supervoxel hierarchies by the uniform entropy slice. In *IEEE/CVF International Conference on Computer Vision*, pages 2240–2247, 2013. **(ICCV)**
6. Chenliang Xu, Richard F. Doell, Stephen Jose Hanson, Catherine Hanson, and Jason J. Corso. Are actor and action semantics retained in video supervoxel segmentation? In *IEEE International Conference on Semantic Computing*, pages 286–293, 2013. **Oral Presentation (ICSC)**
5. Pradipto Das, Chenliang Xu, Richard F. Doell, and Jason J. Corso. A thousand frames in just a few words: Lingual description of videos through latent topics and sparse object stitching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2634–2641, 2013. **(CVPR)**
4. Chenliang Xu, Caiming Xiong, and Jason J. Corso. Streaming hierarchical video segmentation. In *European Conference on Computer Vision*, pages 626–639, 2012. **Oral Presentation (ECCV)**
3. Chenliang Xu and Jason J. Corso. Evaluation of super-voxel methods for early video processing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1202–1209, 2012. **(CVPR)**
2. S. Oh, A. Perera, I. Kim, M. Pandey, K. Cannons, H. Hajimirsadeghi, A. Vahdat, G. Mori, B. Miller, S. McCloskey, Y.-C. Cheng, Z. Huang, C.-H. Lee, C. Xu, R. Kumar, W. Chen, J. J. Corso, L. Fei-Fei, D. Koller, V. Ramanathan, K. Tang, A. Joulin, and A. Alahi. Trecvid 2013 genie: Multimedia event detection and recounting. In *NIST TRECVID Workshop*, 2013.
1. A. Perera, S. Oh, M. Pandey, T. Ma, A. Hoogs, A. Vahdat, K. Cannons, H. Hajimirsadeghi, G. Mori, S. McCloskey, B. Miller, S. Venkatesha, P. Davalos, P. Das, C. Xu, J. J. Corso, R. Srihari, I. Kim, Y.-C. Cheng, Z. Huang, C.-H. Lee, K. Tang, L. Fei-Fei, and D. Koller. Trecvid 2012 genie: Multimedia event detection and recounting. In *NIST TRECVID Workshop*, 2012.

PATENTS

3. System and method for generating videos depicting virtual characters. Luchuan Song and Chenliang Xu. *US-20240346735-A1*, 2024.
2. First-person audio-visual object localization systems and methods. Chenliang Xu, Chao Huang, Yapeng Tian, and FNU Anurag Kumar. *US-20240305944-A1*, 2024.
1. Neural radiance field systems and methods for synthesis of audio-visual scenes. Chenliang Xu, Susan Liang, Chao Huang, Yapeng Tian, and FNU Anurag Kumar. *US-20240267695-A1*, 2024.

Current Projects (sorted by the start date)

7. **PI: Leverage Large Language Models for Complex Robot Manipulation** **\$60,000**
SOURCE: Center of Excellence in Data Science and AI (COE) 09/2025-08/2026
OBJECTIVE: Develop vision-language-action models for efficient, transparent, and complex robot manipulation
6. **PI: Enhancing Video Accessibility for Blink and Low-Vision Individuals Using Large Language Models** **\$42,000**
SOURCE: University of Rochester Research Awards 7/2025-6/2026
OBJECTIVE: A seed grant for developing an AI-powered system that enhances video accessibility and engagement for BLV individuals
5. **Co-PI: VIGIL–Vectors Of Intelligent Guidance In Long-Reach Rural Healthcare** **\$982,906**
SOURCE: The Advanced Research Projects Agency for Health (ARPA-H) 11/2024-10/2029
OBJECTIVE: Improve rural healthcare responses through a continually-learning augmented intelligence that adaptively transforms generalists into specialists based on experience, skill, and context
4. **Co-PI: Artificial Intelligence for effective communication to promote vaping cessation on social media** **\$376,000**
SOURCE: National Institutes of Health (NIH) 9/2024-8/2029
OBJECTIVE: Provide methods effective in delivering the risks of e-cigarette use to social media users and the public, and reaching communities on the potential harms of e-cigarette use
3. **PI: Multimodal Video Understanding** **\$360,000**
SOURCE: Sony Group Corporation 10/2022-8/2026
OBJECTIVE: Develop video understanding methods that interact efficiently across vision, audio, and language modalities, with rich user interaction and spatiotemporal awareness capabilities
2. **Co-PI: CDS&E: Inferring Lattice Dynamics from Temporal X-ray Diffraction Data** **\$375,000**
SOURCE: National Science Foundation (NSF) 8/2022-7/2025
OBJECTIVE: Develop machine learning methods to detect rare lattice-level mechanisms responsible for phase transformation and plastic deformation under extreme conditions.
1. **Co-PI: Time-resolved classification of x-ray diffraction data using deep-learning-powered computer vision techniques** **\$574,050**
SOURCE: National Nuclear Security Administration (NNSA) 7/2022-6/2025
OBJECTIVE: Develop automated deep-learning-powered computer vision techniques for classifying crystal structures and phase fractions from synthetic X-ray diffraction data.

Completed Projects (sorted by the end date)

13. **Co-PI: MILLY–Multi-directional Loosely-Linked Archetype Models for Perceptually-enabled Task Guidance** **\$959,641**
SOURCE: Defense Advanced Research Projects Agency (DARPA) 1/2022-12/2024
OBJECTIVE: Develop AR/VR technologies that allow the AI assistant to perceive the user's intention and environment and provide them with just-in-time feedback through speech and aligned graphics
12. **Sub-PI: From Human-Powered to Automated Video Description for Blind and Low Vision Users** **\$493,380**
SOURCE: NIH 7/2023-10/2024
OBJECTIVE: Develop an AI-based tool in collaboration with sighted describers and blind and low-vision individuals that more efficiently produces video descriptions and offers on-demand access to visual information in online videos
11. **PI: Towards Robust and Fair AI Algorithms against Multiple Shortcuts** **\$120,000**
SOURCE: COE 10/2021-8/2024
OBJECTIVE: Design a novel, scalable human-machine collaborative system to assist humans in discovering unknown biases in any image classifiers

10. **PI: Egocentric 3D Audio-Visual Scene Understanding** **\$399,753**
 SOURCE: Meta 9/2021-8/2024
 OBJECTIVE: Develop AI algorithms that integrate surrounding 3D auditory and visual percepts for scene understanding via audio-visual grounding, separation, and generation with egocentric videos and AR applications as foci

9. **PI: III: Small: Collaborative Research: Scalable Deep Bayesian Tensor Decomposition** **\$199,134**
 SOURCE: NSF 10/2019-9/2023
 OBJECTIVE: Develop new tensor decomposition methods to discover various complicated entity relationships and handle the rapid data growth of real applications

8. **Co-PI: Artificial Intelligence for effective communication on health effects of electronic cigarettes through Instagram** **\$20,000**
 SOURCE: University of Rochester GIDS-AI 10/2021-10/2022
 OBJECTIVE: A seed project to identify potentially effective ways of communicating with the public about the health risks of electronic cigarette use on the most popular social media platform among youth, i.e., Instagram

7. **PI: BIGDATA: F: Audio-Visual Scene Understanding** **\$666,000**
 SOURCE: NSF 9/2017-8/2022
 OBJECTIVE: Develop algorithms to achieve a human-like audio-visual bimodal scene understanding that overcomes the limitations in single-modality analysis through big data analysis of Internet videos

6. **Co-PI: CDS&E: D3SC: Applying Video Segmentation to Coarse-grain Mapping Operators in Molecular Simulations** **\$488,605**
 SOURCE: NSF 8/2018-7/2022
 OBJECTIVE: Apply advances in computer vision, e.g., video segmentation and action recognition, to improve models of multiscale systems in chemistry

5. **Co-PI: Capturing nanoscale lattice variations by applying AI-powered computer vision techniques on synthetic x-ray diffraction data** **\$70,226**
 SOURCE: University of Rochester Research Awards 7/2021-6/2022
 OBJECTIVE: A seed project to generate a planning study and preliminary XRD synthetic data that can be used for future proposals to national programs for deep learning model developments.

4. **PI: RI: Small: Learning Dynamics and Evolution towards Cognitive Understanding of Videos** **\$465,990**
 SOURCE: NSF 9/2018-8/2021
 OBJECTIVE: Develop computational models to study higher-order inference in understanding web instructional videos.

3. **Sub-PI: BOCA – Body-worn Camera Analysis in Public Safety** **\$74,982**
 SOURCE: National Institute of Standards and Technology (NIST) 7/2019-11/2020
 OBJECTIVE: Investigate contextual information (e.g., detected objects, body parts, and their spatiotemporal relations, detected audio events in the scene) to improve the first-person-view activity recognition performance.

2. **PI: Anomaly Detection for Videos Applied in Manufacturing Settings** **\$40,259**
 SOURCE: COE 10/2019-3/2020
 OBJECTIVE: Develop a state-of-the-art deep learning-based computer vision framework for anomaly detection videos applied in manufacturing processes.

1. **Co-PI: Real-Time Synthesis of a Virtual Talking Face from Acoustic Speech** **\$50,000**
 SOURCE: University of Rochester AR/VR Pilot Program 8/2017-7/2018
 OBJECTIVE: Develop a real-time talking face by analyzing the correlations between auditory and visual signals to assist hearing-impaired people.

Research Gift Donations (total \$460,000)

16. Adobe, \$10,000 12/2024
15. Cisco Faculty Award, \$75,000 9/2024

14. Adobe, \$7,000	12/2021
13. Corning, \$70,000	11/2021
12. Adobe, \$10,000	9/2021
11. Wilmot Professorship, \$10,000	7/2021
10. InnoPeak Technology, \$30,000	3/2021
9. Adobe, \$7,000	3/2021
8. Corning, \$30,000	3/2021
7. Adobe, \$7,500	11/2020
6. Adobe, \$28,500	Fall 2019
5. Corning, \$30,000	10/2019
4. Viscovery, \$55,000	9/2018
3. Tencent AI Lab Rhino-Bird, \$60,000	6/2018
2. NVIDIA, GPU Donations	8/2017
1. Markable AI, \$30,000	8/2017

PROFESSIONAL
SERVICES

Organizing Committee

• CVPR Workshop on GenAI for Storytell	2026
• ICCV Trustworthy FMs Workshop	2025
• ICME Tutorial “Multimedia Deepfake Detection”	2024
• CVPR Tutorial “Audio-Visual Scene Understanding”	2021
• WACV Tutorial “Audio-Visual Scene Understanding”	2021
• CVPR Workshop “Brave New Ideas For Motion and Spatio-Temporal Representations”	2017

Conference Senior Roles

• Program Co-Chair, International Conference on Robotics and Automation Sciences	2026
• Area Chair, CVPR	2024—2026
• Area Chair, ACM Multimedia	2023—2025
• Senior PC, AAAI Conference on Artificial Intelligence	2022
• Senior PC, International Joint Conference on Artificial Intelligence (IJCAI)	2021
• Advisory Committee, International Conference on Computer Vision & Image Processing	2021

Conference Reviewer

• IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	2016—2023
• European Conference on Computer Vision (ECCV)	2018—2022
• IEEE/CVF International Conference on Computer Vision (ICCV)	2015—2025
• Neural Information Processing Systems (NeurIPS)	2016, 2021, 2025
• International Conference on Learning Representations (ICLR)	2022, 2025
• AAAI Conference on Artificial Intelligence (AAAI)	2017—2021
• Conference on Empirical Methods in Natural Language Processing (EMNLP)	2020
• IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)	2026
• Asian Conference on Computer Vision (ACCV)	2016—2018
• British Machine Vision Conference (BMVC)	2015—2019, 2021
• Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)	2019
• Annual Conference of the European Association for Computer Graphics	2019
• International Conference on Semantic Computing (ICSC)	2017
• International Symposium CompIMAGE’16	2016
• XXIIIrd ISPRS Congress	2016
• ISPRS Geospatial Week – Image Sequence Analysis	2015
• IEEE/ISPRS 3rd Joint Workshop on Multi-Sensor Fusion for Dynamic Scene Understanding	2015
• Indian Conference on Computer Vision, Graphics and Image Processing	2014

Journal Services

• Associate Editor, IEEE Transactions on Multimedia	2022—2024
• Reviewer, IEEE Transactions on Pattern Analysis and Machine Intelligence	2016—2021
• Reviewer, IEEE Transactions on Image Processing	2015—2021
• Reviewer, IEEE Transactions on Multimedia	2013—2021
• Reviewer, International Journal of Computer Vision	2016—2017, 2020
• Reviewer, Computer Vision and Image Understanding	2016—2020

- Reviewer, IEEE Transactions on Circuits and Systems for Video Technology 2014—2019
- Reviewer, Pattern Recognition 2016—2019
- Reviewer, Autonomous Robots 2018
- Reviewer, Signal Processing: Image Communication 2016—2017
- Reviewer, Journal of Visual Communication and Image Representation 2017
- Reviewer, ISPRS Journal of Photogrammetry and Remote Sensing 2017
- Reviewer, Machine Vision and Applications 2016
- Reviewer, IET Computer Vision 2014, 2016
- Reviewer, Electronic Letters on Computer Vision and Image Analysis 2016
- Reviewer, Image and Vision Computing 2014—2015
- Reviewer, OSA Biomedical Optics Express 2014
- Reviewer, IPSJ Transactions on Computer Vision and Applications 2013

Grant Panelist/Reviewer

- National Science Foundation (CISE/IIS, Panelist) 2021
- National Science Foundation (Ad-hoc Reviewer) 2020
- National Science Foundation (CISE/CNS, Panelist) 2019
- National Science Foundation (CISE/IIS, Panelist) 2019
- University Research Awards (UR, Reviewer) 2018

UNIVERSITY SERVICES

- UR CS PhD Admissions Committee Chair 2022—Present
- UR Data Science PhD Program Planning Committee 2022—Present
- UR CS Undergraduate Curriculum Committee 2019—2021
- UR CS Lab Committee 2018—2022
- UR CS PhD Admissions Committee 2017—2022
- UR GIDS Working Group Co-Chair 2020—2021
- UR Sproull Fellowship Review Panel 2021
- UR GIDS MS Admissions Committee 2018—2020
- UR Outstanding Dissertation Award Committee 2019
- UR CS Faculty Search Committee 2019
- UR CS Website Committee 2018

TALKS

Understand and Reconstruct Multimodal Egocentric Scenes

- Invited Talk at VinAI 1/2025
- Invited Talk at Hanoi University of Science and Technology 1/2025
- Invited Talk at Corning, Inc. 8/2024

Audio-Visual Synthesis, Separation, and Joint Editing by Language

- Invited Talk at the AVGenL Workshop in ECCV 2024 9/2024

Towards Fair Computer Vision: Discover the Hidden Biases of an Image Classifier

- Invited Talk at the AI and Computer Vision Meetup by Voxel51 2/2024

Audio-Visual Scene Synthesis for AI-Empowered Physical Retail

- Invited Talk at the 1st Physical Retail AI Workshop (PRAW) in IEEE/CVF WACV 1/2024

The Fragile Audio-Visual Integration under Adversarial Attacks

- CVPR Workshop on Sight and Sound 6/2021

Cyclic Co-Learning of Sounding Object Visual Grounding and Sound Separation

- Facebook Reality Labs 4/2021

Discover the Hidden Biases of an Image Classifier

- URCS Computer Vision Seminar 2/2021
- Aibee Inc. 1/2021

Audio-Visual Scene Understanding

- MIT Computer Vision Seminar 10/2020

Learning from Interventions using Hierarchical Policies for Safe Learning

- NIST PCSR PI Meeting 7/2020

Learning from Web Instructional Videos

- UR Warner School LiDA Colloquium Series 2/2020

GAN-EM: GAN Based EM Learning Framework

- International Joint Conference on Artificial Intelligence 8/2019

Audio-Visual Scene Understanding

- Corning Inc. 11/2019

• Viscovery Inc.	5/2019
• UR Laboratory for Laser Energetics	12/2018
• Western New York Image and Signal Processing Workshop	10/2018
<i>Applications of Multimodal Deep Learning</i>	
• URMC Clinical and Translational Science Institute Analytics Colloquium	4/2018
<i>MRI Tumor Segmentation with Densely Connected 3D CNN</i>	
• SPIE Medical Imaging	2/2018
<i>From Cross-Modal Generation to Audio-Visual Scene Understanding</i>	
• UR Data Science Summer Colloquium	6/2018
• Rochester Institute of Technology	11/2017
<i>Towards the What, Who and Where of Video Understanding</i>	
• CVPR Workshop “The DAVIS Challenge on Video Object Segmentation”	7/2017
<i>Actor-Action Semantic Segmentation with Grouping Process Models</i>	
• Midwest Vision Workshop	4/2016
<i>Scale-Adaptive Video Understanding</i>	
• University of Utah	3/2016
• Stevens Institute of Technology	3/2016
• University of Rochester	3/2016
• Michigan State University	2/2016
<i>Pulling Information from Scales</i>	
• UMich SPEECS Seminar	1/2016
<i>Action Understanding with Multiple Classes of Actors</i>	
• UMich AI Lab Mini-Symposium	5/2015
<i>Hierarchical Video Segmentation: Methods, Perception and Application</i>	
• UMich SPEECS Seminar	9/2014
<i>LIBSVX and Video Segmentation Evaluation</i>	
• IEEE CVPR Tutorial on Video Segmentation	6/2014
<i>Are actor and action semantics retained in video supervoxel segmentation?</i>	
• IEEE International Conference on Semantic Computing	9/2013

TEACHING

CSC 249/449 Machine Vision	
<i>UR: Spring 2018, Spring 2019, Spring 2020, Fall 2020, Spring 2022, Spring 2024</i>	
CSC 245/445 Deep Learning	
<i>UR: Spring 2017, Fall 2021, Fall 2022, Fall 2023, Fall 2024</i>	
CSC 577 Advanced Topics in Computer Vision	
<i>UR: Fall 2016, Fall 2017, Fall 2018, Fall 2019, Spring 2021, Spring 2023</i>	
CSC 413 Intro to Augmented and Virtual Reality (Co-Instructor)	
<i>UR: Fall 2020, Fall 2021, Fall 2022</i>	
CSC 414 Selected Topics in Augmented and Virtual Reality	
<i>UR: Spring 2023</i>	
NSF REU Site: Computational Methods for Understanding Music, Media, and Minds	
<i>UR: Summer 2017, Summer 2018</i>	
Taught a multi-day deep learning workshop to NSF REU site students.	
Pre-College Data Science Program	
<i>UR: Summer 2018</i>	
Taught a half-day introduction to AI to a group of international high school students visiting UR from China, South Korea, Spain, Haiti, etc.	
EECS 598 Foundations of Computer Vision (Grad TA)	
<i>UMich: Fall 2015</i>	

STUDENT ADVISING

Student Distinctins	
• UG Advisee, Jiani Liu, selected as Runner Up for the CRA Outstanding UG Awards	2025
• PhD Advisee, Chao Huang, received the NeurIPS Scholar Award	2025
• PhD Advisee, Chao Huang, selected for the ICCV Doctoral Consortium	2025

- High School Advisee, Matthew Xie, named among top scholars in the Regeneron Science Talent Search 2025
- PhD Advisee, Zhiheng Li, selected for the CVPR Doctoral Consortium 2023
- UG Advisee, Sizhe Li, selected as Finalist for the CRA Outstanding UG Awards 2023
- PhD Advisee, Yapeng Tian, selected for the AAAI New Faculty Highlights 2022
- PhD Advisee, Yapeng Tian, selected for the CVPR Doctoral Consortium 2022
- Ph.D. Advisee, Lele Chen, won the Donald M. and Janet C. Barnard Fellowship 2020

Ph.D. Advisees

12. Pinxin Liu (UR CS) 2025—Present
11. Mingqian Feng (UR CS) 2023—Present
10. Yunlong Tang (UR CS) 2023—Present
9. Zeliang Zhang (UR CS) 2022—Present
8. Susan Liang (UR CS) 2022—Present
7. Luchuan Song (UR CS) 2022—Present
6. Chao Huang (UR CS) PhD: Graduated 2026
DISSERTATION: *Controllable Generative Modeling for Multimodal Perception and Synthesis*
FIRST APPOINTMENT: *Research Scientist at Tencent America*
5. Jing Bi (UR CS) MS: 2018—2020; PhD: Graduated 2026
DISSERTATION: *Learn to Act in Pixels*
4. Zhiheng Li (UR CS) UG: Summer 2017; PhD: Graduated 2023
DISSERTATION: *Discover and Mitigate Biases in Discriminative and Generative Image Models*
FIRST APPOINTMENT: *Research Scientist at Amazon AWS*
3. Jing Shi (UR CS) PhD: Graduated 2022
DISSERTATION: *Vision Based Language to Action Mapping*
FIRST APPOINTMENT: *Research Scientist at Adobe Research*
2. Yapeng Tian (UR CS) PhD: Graduated 2022
DISSERTATION: *Audio-Visual Scene Understanding Towards Unified, Explainable, and Robust Multisensory Perception*
FIRST APPOINTMENT: *Assistant Professor at UT Dallas*
1. Lele Chen (UR CS) PhD: Graduated 2022
DISSERTATION: *High-Fidelity Talking Avatar Video Generation*
FIRST APPOINTMENT: *Staff Research Scientist at Oppo US Research*

Ph.D. Committee

36. Jingyang Lin (UR CS)
35. Yifan Li (UR CS)
34. Yifan Zhu (UR CS)
33. James Spann (UR CS)
32. Hang Hua (UR CS) Graduated 2025
31. Nathan Kent (UR CS) Graduated 2025
30. Jie An (UR CS) Graduated 2025
29. Max Wasserman (UR CS) Graduated 2025
28. Ruoyang Hu (UR BCS/CS) Graduated 2024
27. Yiming Gan (UR CS) Graduated 2023
26. Yu Feng (UR CS) Graduated 2023
25. Yu Chen (UR CS) Graduated 2023
24. Songyang Zhang (UR CS) Graduated 2023
23. Haitian Zheng (UR CS) Graduated 2023
22. Wei Zhu (UR CS) Graduated 2023
21. Ali K. Shargh (UR ME) Graduated 2022
20. Sabyasachi Shivkumar (UR BCS/CS) Graduated 2022

19. Weijian Li (UR CS)	Graduated 2022
18. Wei Xiong (UR CS)	Graduated 2022
17. Joseph German (UR BCS/CS)	Graduated 2022
16. Lisa Jin (UR CS)	Graduated 2022
15. Zhengyuan Yang (UR CS)	Graduated 2021
14. Shupeng Gui (UR CS)	Graduated 2021
13. Hanlin Tang (UR CS)	Graduated 2021
12. Samiha Samrose (UR CS)	Graduated 2021
11. Taylan Sen (UR CS)	Graduated 2021
10. Nabil Hossain (UR CS)	Graduated 2020
9. Haichuan Yang (UR CS)	Graduated 2020
8. Feng Yang (UR CS)	Graduated 2020
7. Christopher Bates (UR BCS/CS)	Graduated 2020
6. Bochen Li (UR ECE)	Graduated 2020
5. Kushal Kafle (Rochester Institute of Technology)	Graduated 2020
4. Haofu Liao (UR CS)	Graduated 2019
3. Jianbo Yuan (UR CS)	Graduated 2019
2. Yichi Zhang (UR ECE)	Graduated 2019
1. Sefik Emre Eskimez (UR ECE)	Graduated 2019

MS Advisees

27. Jingyan Chen (Fall 2024 -)
26. Xinyang Li (Fall 2024 -)
25. Yuesong Huang (Fall 2024 -)
24. Bingjie Wang (Fall 2024 -)
23. Jiacan Yu (Fall 2024 -)
22. Junjia Guo (Summer 20204 -)
21. Bowen Yang (Spring 2024 -)
20. Nguyen Manh Nguyen (Fall 2022 - Spring 2024)
19. Rongyi Zhu (Fall 2022 - Spring 2024, then a Ph.D. student at Stony Brook University)
18. Shijian Deng (Fall 2021 - Fall 2022, then a Ph.D. student at UT Dallas)
17. Guangyu Sun (Fall 2020 - Summer 2022, then a Ph.D. student at UCF)
16. Rohan Sharma (Summer - Fall 2020, then a Ph.D. student at SUNY Buffalo)
15. Guo Li (Summer 2019 - Spring 2020, then Amazon)
14. Guofeng Cui (Spring 2019 - Spring 2020, then a Ph.D. student at Rutgers University)
13. Ziyi Kou (Fall 2018 - Spring 2020, then a Ph.D. student at Univ. of Notre Dame)
12. Justin Tian (Fall 2018 - Spring 2019)
11. Jing Bi (Summer 2018 - Summer 2020, then a Ph.D. student in URCS)
10. Wei Zhang (Summer 2018 - Spring 2019, then Amazon)
9. Hao Huang (Summer 2018 - Spring 2019, then a Ph.D. student at NYU)
8. Wentian Zhao (Summer 2018 - Spring 2019, then Adobe Research)
7. Shaojie Wang (Summer 2018 - Spring 2019, then a Ph.D. student at WUSTL)
6. Jie Chen (Fall 2017 - Fall 2019, then Microsoft Research)
5. Li Ding (Summer 2017, then a Research Engineer at MIT)
4. Sudhanshu Srivastava (Spring - Summer 2017, then a Ph.D. student at UCSB)
3. Lele Chen (Spring 2017 - Spring 2018, then a Ph.D. student in URCS)
2. Mingyang Zhou (Spring 2016, Univ. of Michigan, then a Ph.D. student at UC Davis)
1. Yao Li (Spring 2013, SUNY Buffalo)

Undergraduate Advisees

33. Junhua Huang (Fall 2024 -)
32. Jiarui Wu (Summer 2024 -)
31. Jinxi He (Summer 2024 -)
30. Xin Liang (Summer 2024 - , Tongji University)
29. Pinxin (Andy) Liu (Spring 2023 - Summer 2024, then a Ph.D. student in URCS)
28. Ji-Ze Jang (Summer 2021 - Summer 2022)
27. Miranda Rublaitus (Summer 2022, NSF REU, Yale University)
26. Kalsey Colotl (Summer 2022, NSF REU, NYU)
25. Sizhe Li (Spring 2020 - Fall 2022, then a Ph.D. student at MIT)
Finalist for the CRA Outstanding UG Researcher Awards 2023
24. Phuong Vu (Fall 2021)
23. Yiyang Su (Summer 2020 - Spring 2021, then a Ph.D. student at MSU)
22. Logan Peters (Summer 2020, Carleton College)
21. Theodore Chapman (Summer 2020, NSF REU)
20. Patrick Phillips (Summer 2020, NSF REU)
19. Bryce Elizabeth Yahn (Summer 2020, NSF REU)
18. Tianyou Xiao (Summer 2018 - Fall 2019, then a master's student at Cornell Tech)
17. Wenxuan Bao (Summer - Fall 2019, Tsinghua University, then a Ph.D. student at UIUC)
16. Ariel Tello (Summer 2019, Xero Research Fellow)
15. Chenxiao Guan (Summer 2018, Xerox Research Fellow, then a master's student at CMU)
14. Qiuyue Sun (Summer 2018, then a master's student at Stanford)
13. Justin Goodman (Summer 2018, NSF REU, University of Maryland)
12. Marc Moore (Summer 2018, NSF REU, Mississippi State University)
11. Amos Newswanger (Summer - Fall 2017)
10. Max Torop (Summer 2017, then a master's student at WUSTL)
9. Zhiheng Li (Summer-Fall 2017, Wuhan University, then a Ph.D. student in URCS)
8. Wei Zhao (Summer 2017, USTC)
7. Wes Smith (Summer 2017, NSF REU, University of Edinburgh, then a Ph.D. student in URCS)
6. Moses Bug (Summer 2017, NSF REU, Brandeis University)
5. Shengqi Suizhu (Fall 2016 - Spring 2017)
4. Austin Schaffer (Spring 2014, SUNY Buffalo)
3. Libing Wu (Spring 2014, SUNY Buffalo)
2. Spencer Whitt (Summer 2013, SUNY Buffalo)
1. Tyler Ganter (Summer 2013, SUNY Buffalo)

Visiting Scholars

4. Ali Vosoughi Fall 2022—Present
Ph.D. Student of ECE, University of Rochester
3. Daiki Shimada Fall 2022—Summer 2023
Researcher at Sony R&D Center
2. Robert Jacobs Fall 2018
Professor of Brain and Cognitive Science, University of Rochester
1. Baojie Fan Summer 2018—Spring 2019
Associate Professor, Nanjing University of Posts and Telecommunications

6. **AVE** 2018
Audio-Visual Event (AVE) dataset is a large video dataset that consists of 4143 10-second videos with both audio and video tracks for 28 audio-visual events and their temporal boundary annotations. It is the largest dataset for sound event detection.
<http://www.cs.rochester.edu/~cxu22/d/ave/>
5. **YouCook2** 2018
YouCook2 is the largest task-oriented, instructional video dataset in the vision community. It contains 2000 long untrimmed videos from 89 cooking recipes; on average, each distinct recipe has 22 videos. The procedure steps for each video are annotated with temporal boundaries and described by imperative English sentences.
<http://youcook2.eecs.umich.edu/>
4. **A2D** 2015
A dataset and benchmark for action recognition and segmentation with multiple classes of actors. It considers seven actor classes (adult, baby, dog, etc.) and eight action classes (climb, crawl, eat, etc.). It contains 3782 videos with at least 99 instances per valid actor-action tuple.
<http://www.cs.rochester.edu/~cxu22/a2d/>
3. **Video2Text.net** 2013
A website and web-service for automatic conversion of videos to natural language sentences based on the video content. This website showcases our work in the vision+language domain.
<http://www.video2text.net>
2. **YouCook** 2013
Dataset of third-person cooking videos categorized into six styles of cooking and selected from open-source web videos of different kitchens and complexity levels. It contains object and action bounding boxes as well as multiple natural language descriptions of each video.
<http://www.cs.rochester.edu/~cxu22/youcook/>
1. **LIBSVX** 2012, 2013
Supervoxel library: a set of methods for early video processing by computing supervoxel segmentations as well as a quantitative benchmark for fair comparisons of those segmentations.
<http://www.supervoxels.com>
– *Winner Best Demo Prize at 2nd Greater New York Multimedia and Vision Meeting.* 6/2012
– *Winner Best Open Source Code 3rd Prize at IEEE CVPR 2012.* 6/2012