

# Chenliang Xu

Prepared on April 9, 2024

CONTACT	3005 Wegmans Hall 250 Hutchison Road University of Rochester Rochester, NY 14627	EMAIL: chenliang.xu@rochester.edu WEB: <a href="https://www.cs.rochester.edu/~cxu22">https://www.cs.rochester.edu/~cxu22</a>
APPOINTMENTS	<b>Associate Professor</b> Department of Computer Science University of Rochester Rochester, NY	7/2022 - PRESENT
	<b>Assistant Professor</b> 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	<b>Doctor of Philosophy</b> Computer Science and Engineering University of Michigan Ann Arbor, MI	2016
	<b>Master of Science</b> Computer Science and Engineering University at Buffalo Buffalo, NY	2012
	<b>Bachelor of Science</b> Information and Computing Science Nanjing University of Aeronautics and Astronautics Nanjing, China	2010
DISTINCTIONS	<b>James P. Wilmot Distinguished Professorship</b> <i>for "some of the most promising young men and women in the early stages of their academic careers"</i>	2021
	<b>University of Rochester Research Awards</b> <i>for the project "applying AI-powered computer vision techniques on synthetic x-ray diffraction data"</i>	2021
	<b>Best Paper Award at the 17th ACM SIGGRAPH VRCAI Conference</b> <i>for the paper "3D human avatar digitization from a single image"</i>	2019
	<b>Best Paper Award at the 14th Sound and Music Computing Conference</b> <i>for the paper "audio-visual source association for string ensembles through multi-modal vibrato analysis"</i>	2017
	<b>University of Rochester AR/VR Pilot Award</b> <i>for the project "real-time synthesis of a virtual talking face from acoustic speech"</i>	2017
	<b>Best Open Source Code Third Prize at IEEE CVPR</b> <i>for the paper "evaluation of super-voxel methods for early video processing"</i>	2012
	<b>Best Demo Prize at the 2nd Greater New York Multimedia and Vision Meeting</b> <i>for the demo "streaming hierarchical video segmentation"</i>	2012
PUBLICATIONS	<b>Google Scholar:</b> <a href="https://scholar.google.com/citations?user=54HfyDIAAAAJ">https://scholar.google.com/citations?user=54HfyDIAAAAJ</a>  <b>Peer-Reviewed Articles and Proceedings</b> CVPR, ICCV, and ECCV are premier computer vision conferences. According to Google Scholar Metrics, as of 8/2023, CVPR has h5-index 422, ICCV 228, and ECCV 238. CVPR is also ranked 1st of all journals and conferences in Engineering and Computer Science and 4th when considering everything else.	

## 2024

- P101. 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*, 2024 (NAACL)
- P100. 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*, 2024 (CVPR)
- P99. 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*, 2024 (CVPR)
- P98. 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*, pages 1–16, 2024. (TMM)
- P97. 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*, 2024. (EACL)
- P96. Jinyang Jiang, Zeliang Zhang, Chenliang Xu, Zhaoifei Yu, and Yijie Peng. One forward is enough for neural network training via likelihood ratio method. In *International Conference on Learning Representations*, 2024. (ICLR)
- P95. 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*, 2024. (ICASSP)
- P94. 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*, 2024. (ICASSP)
- P93. 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

- P92. 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 Comp. Mater.)
- P91. 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)
- P90. 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)
- P89. 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)
- P88. 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)

- P87. 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)**
- P86. 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. **(PLOS Comp. Biology)**
- P85. 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. **(AV4D workshop)**
- P84. 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. **(AV4D workshop)**
- P83. 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. **(AV4D workshop)**
- P82. 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. **(Sight and Sound Workshop)**
- P81. 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. **(Sight and Sound Workshop)**

## 2022

- P80. 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)**
- P79. 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. **(ACM Multimedia)**
- P78. 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)**
- P77. 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)**
- P76. 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)**
- P75. 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)**
- P74. 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)**

- P73. 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. **(Trends in Hearing)**
- P72. 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. **(Sight and Sound Workshop)**

## 2021

- P71. 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)**
- P70. 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)**
- P69. 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)**
- P68. 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)**
- P67. 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)**
- P66. 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)**
- P65. 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)**
- P64. 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)**
- P63. 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)**
- P62. 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)**
- P61. 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)**
- P60. 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)**
- P59. 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. **(CG)**
- P58. Yankun Gao, Zidian Xie, Li Sun, Chenliang Xu, and Dongmei Li. Characteristics and user

## 2020

- P57. 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)**
- P56. 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)**
- P55. 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)**
- P54. 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)**
- P53. 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)**
- P52. 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)**
- P51. 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)**
- P50. 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)**
- P49. 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)**
- P48. 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)**
- P47. 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)**
- P46. 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)**
- P45. 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.
- P44. 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.

- P43. 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. **(Sight and Sound Workshop)**
- P42. 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. **(Sight and Sound Workshop)**
- P41. 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. **(Sight and Sound Workshop)**
- P40. 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. **(Sight and Sound Workshop)**

## 2019

- P39. 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)**
- P38. 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)**
- P37. Lele Chen, Ross K. Maddox, Zhiyao Duan, and Chenliang Xu. Hierarchical cross-modal talking face generation with dyanmic pixel-wise loss. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7832–7841, 2019. **(CVPR)**
- P36. 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)**
- P35. 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)**
- P34. 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. **(GlobalSIP)**
- P33. 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)**
- P32. Robert A. Jacobs and Chenliang Xu. Can multisensory training aid visual learning?: A computational investigation. *Journal of Vision*, 19(11):1–12, 2019. **(JOV)**
- P31. 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. **(TASLP)**
- P30. Bochen Li, Karthik Dinesh, Chenliang Xu, Gaurav Sharma, and Zhiyao Duan. Online audio-visual source association for chamber music performances. *Transcations of the International Society for Music Information Retrieval*, 2(1):29–42, 2019. **(TISMIR)**
- P29. 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. **Oral Presentation (Sight and Sound Workshop)**



- P28. 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. **(Sight and Sound Workshop)**
- P27. 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. **(Sight and Sound Workshop)**

## 2018

- P26. 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)**
- P25. 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)**
- P24. 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)**
- P23. 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)**
- P22. 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)**
- P21. 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 MI)**
- P20. 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. **(LVA/ICA)**
- P19. 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.
- P18. 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

- P17. 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)**
- P16. 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)**
- P15. 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)**

- P14. 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. **(CVPR Workshop)**
- P13. 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. **(ACM MM Workshop)**
- P12. 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. **(ACM MM Workshop)**

## 2016 and Before

- P11. 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)**
- P10. 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)**
- P9. 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)**
- P8. 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)**
- P7. 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)**
- P6. 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)**
- P5. 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)**
- P4. 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)**
- P3. 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)**
- P2. 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.
- P1. 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.



- T8. Lele Chen, Guofeng Cui, Ziyi Kou, Haitian Zheng, and Chenliang Xu. What comprises a good talking-head video generation?: A survey and benchmark. *arXiv preprint arXiv:2005.03201*, 2020.
- T7. Yapeng Tian, Chenxiao Guan, Justin Goodman, Marc Moore, and Chenliang Xu. An attempt towards interpretable audio-visual video captioning. *arXiv preprint arXiv:1812.02872*, 2018.
- T6. Yapeng Tian, Chenliang Xu, and Dingzeyu Li. Deep audio prior. *arXiv preprint arXiv:1912.10292*, 2019.
- T5. Burak Donderici, Caleb New, and Chenliang Xu. Assembling semantically-disentangled representations for predictive-generative models via adaptation from synthetic domain. *arXiv preprint arXiv:2002.09818*, 2020.
- T4. Jing Bi, Tianyou Xiao, Qiuyue Sun, and Chenliang Xu. Navigation by imitation in a pedestrian-rich environment. *arXiv preprint arXiv:1811.00506*, 2018.
- T3. Li Ding and Chenliang Xu. Tricorner: A hybrid temporal convolutional and recurrent network for video action segmentation. *arXiv preprint arXiv:1705.07818*, 2017.
- T2. Chenliang Xu, Caiming Xiong, and Jason J. Corso. Action understanding with multiple classes of actors. *arXiv preprint arXiv:1704.08723*, 2017.
- T1. Chenliang Xu. *Scale-adaptive video understanding*. PhD thesis, University of Michigan, 2016.

## FUNDING

### Current Projects (sorted by the start date)

- F18. **PI: Audio-Visual Referential Dialogue** **\$100,000**  
 SOURCE: Sony Group Corporation 9/2023-8/2024  
 OBJECTIVE: Develop models that interact efficiently across vision, audio, and language modalities, with rich user interaction and spatiotemporal awareness capabilities.
- F17. **PI: Towards Robust and Fair AI Algorithms against Multiple Shortcuts** **\$60,000**  
 SOURCE: The New York State Center of Excellence in Data Science 9/2023-8/2024  
 OBJECTIVE: A seed project to generate preliminary results that can be used for future proposals to design a novel, scalable human-machine collaborative system to assist humans in discovering unknown biases in any image classifiers.
- F16. **Sub-PI: From Human-Powered to Automated Video Description for Blind and Low Vision Users** **\$493,380**  
 SOURCE: National Institutes of Health (NIH) 7/2023-6/2028  
 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.
- F15. **Co-PI: CDS&E: Inferring Lattice Dynamics from Temporal X-ray Diffraction Data** **\$375,000**  
 SOURCE: National Science Foundation (NSF) #2202124 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.
- F14. **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.
- F13. **Sub-PI: MILLY-Multi-directional Loosely-Linked Archetype Models for Perceptually-enabled Task Guidance** **\$959,641**  
 SOURCE: Defense Advanced Research Projects Agency (DARPA) 1/2022-1/2026  
 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.

- F12. **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.

**Completed Projects** (*sorted by the end date*)

- F11. **PI: III: Small: Collaborative Research: Scalable Deep Bayesian Tensor Decomposition** **\$199,134**  
 SOURCE: National Science Foundation (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.
- F10. **PI: Explore Audio-Visual Model Robustness under Common Corruptions** **\$60,000**  
 SOURCE: Sony Group Corporation 10/2022-9/2023  
 OBJECTIVE: Study flexibility and robustness of audio-visual multimodal recognition models under common corruptions.
- F9. **PI: Identify Hidden Biases of AI Algorithms via Human-Machine Collaboration** **\$60,000**  
 SOURCE: The New York State Center of Excellence in Data Science 10/2021-6/2023  
 OBJECTIVE: Design a novel, scalable human-machine collaborative system to assist humans in discovering unknown biases in any image classifiers.
- F8. **Co-PI: Artificial Intelligence for effective communication on health effects of electronic cigarettes through Instagram** **\$20,000**  
 SOURCE: UR Goergen Institute for Data Science 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 in youth, i.e., Instagram.
- F7. **PI: BIGDATA: F: Audio-Visual Scene Understanding** **\$666,000**  
 SOURCE: National Science Foundation (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.
- F6. **Co-PI: CDS&E: D3SC: Applying Video Segmentation to Coarse-grain Mapping Operators in Molecular Simulations** **\$488,605**  
 SOURCE: National Science Foundation (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.
- F5. **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.
- F4. **PI: RI: Small: Learning Dynamics and Evolution towards Cognitive Understanding of Videos** **\$465,990**  
 SOURCE: National Science Foundation (NSF) 9/2018-8/2021  
 OBJECTIVE: Develop computational models to study higher-order inference in understanding web instructional videos.
- F3. **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.
- F2. **PI: Anomaly Detection for Videos Applied in Manufacturing Settings** **\$40,259**  
 SOURCE: The New York State Center of Excellence in Data Science 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.

- F1. 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 \$368,000)

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

#### Participated Major Efforts

* <i>NRT-HDR: Interdisciplinary Graduate Training in the Science, Technology, and Applications of Augmented and Virtual Reality</i> SOURCE: NSF #1922591 OBJECTIVE: The project will train a new cohort of Ph.D. students with a unique set of competencies in the AR/VR domain and human-data-system interfaces.	\$1,560,000 9/2019-8/2024
* <i>HDR TRIPODS: Collaborative Research: Foundations of Greater Data Science</i> SOURCE: NSF #1934962 OBJECTIVE: The University of Rochester and Cornell University jointly establish the Greater Data Science Cooperative Institute (GDSC).	\$550,266 9/2019-8/2022
* <i>REU Site: Computational Methods for Understanding Music, Media, and Minds</i> SOURCE: NSF #1659250 OBJECTIVE: Students in the University's REU will engage in interdisciplinary research that combines machine learning, audio engineering, music theory, and cognitive science.	\$323,950 3/2017-2/2020

#### PROFESSIONAL SERVICES

##### ASSOCIATE EDITOR

* IEEE Transactions on Multimedia (IEEE TMM)	2022—Present
--	--------------

##### ORGANIZING COMMITTEE

* 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

##### AREA CHAIR

* IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	2024
* ACM Multimedia	2023, 2024

##### SENIOR PROGRAM COMMITTEE/META-REVIEWER

* AAAI Conference on Artificial Intelligence	2022
* International Joint Conference on Artificial Intelligence (IJCAI)	2021

##### PROGRAM COMMITTEE

* 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—2021
* AAAI Conference on Artificial Intelligence (AAAI)	2017—2021
* Neural Information Processing Systems (NeurIPS)	2016, 2021
* International Conference on Learning Representations (ICLR)	2022
* Conference on Empirical Methods in Natural Language Processing (EMNLP)	2020

	* British Machine Vision Conference (BMVC)	2015—2019, 2021
	* Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)	2019
	* Asian Conference on Computer Vision (ACCV)	2016—2018
	* 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
	INTERNATIONAL ADVISORY COMMITTEE	
	* International Conference on Computer Vision & Image Processing (CVIP)	2021
	JOURNAL REVIEWER	
	* IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)	2016—2021
	* IEEE Transactions on Image Processing (IEEE TIP)	2015—2021
	* IEEE Transactions on Multimedia (IEEE TMM)	2013—2021
	* International Journal of Computer Vision (IJCV)	2016—2017, 2020
	* Computer Vision and Image Understanding (CVIU)	2016—2020
	* IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT)	2014—2019
	* Pattern Recognition (PR)	2016—2019
	* Autonomous Robots	2018
	* Signal Processing: Image Communication	2016—2017
	* Journal of Visual Communication and Image Representation	2017
	* ISPRS Journal of Photogrammetry and Remote Sensing	2017
	* Machine Vision and Applications	2016
	* IET Computer Vision	2014, 2016
	* Electronic Letters on Computer Vision and Image Analysis	2016
	* Image and Vision Computing (IVC)	2014—2015
	* OSA Biomedical Optics Express	2014
	* 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—Present
	* 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	<i>MILLY: Multi-Directional Loosely-Linked Archetype Models for PTG</i>	
	• DARPA PTG PI Meeting	11/2022
	• DARPA PTG PI Meeting	5/2022
	• DARPA PTG Kickoff Meeting	11/2021
	<i>The Fragile Audio-Visual Integration under Adversarial Attacks</i>	
	• CVPR Workshop on Sight and Sound	6/2021
	<i>Cyclic Co-Learning of Sounding Object Visual Grounding and Sound Separation</i>	
	• Facebook Reality Labs	4/2021

<i>Discover the Hidden Biases of an Image Classifier</i>	
• URCS Computer Vision Seminar	2/2021
• Aibee Inc.	1/2021
<i>Audio-Visual Scene Understanding</i>	
• MIT Computer Vision Seminar	10/2020
<i>Learning from Interventions using Hierarchical Policies for Safe Learning</i>	
• NIST PCSR PI Meeting	7/2020
<i>Learning from Web Instructional Videos</i>	
• UR Warner School LiDA Colloquium Series	2/2020
<i>GAN-EM: GAN Based EM Learning Framework</i>	
• International Joint Conference on Artificial Intelligence	8/2019
<i>Audio-Visual Scene Understanding</i>	
• Corning Inc.	11/2019
• Viscosity 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>	
• URMCI 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**

UR: Spring 2018, Spring 2019, Spring 2020, Fall 2020, Spring 2022

### **CSC 298/578 Deep Learning**

UR: Spring 2017, Fall 2021, Fall 2022

### **CSC 577 Advanced Topics in Computer Vision**

UR: Fall 2016, Fall 2017, Fall 2018, Fall 2019, Spring 2021, Spring 2023

### **CSC 413 Intro to Augmented and Virtual Reality (Co-Instructor)**

UR: Fall 2020, Fall 2021, Fall 2022

### **CSC 414 Selected Topics in Augmented and Virtual Reality**

UR: Spring 2023

### **NSF REU Site: Computational Methods for Understanding Music, Media, and Minds**

UR: Summer 2017, Summer 2018

Taught a multi-day deep learning workshop to NSF REU site students.

### Pre-College Data Science Program

UR: Summer 2018

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)

UMich: Fall 2015

VISITING SCHOLARS HOSTED	Ali Vosoughi <i>Ph.D. Student of ECE, University of Rochester</i>	Fall 2022—Present
	Daiki Shimada <i>Researcher at Sony R&amp;D Center</i>	Fall 2022—Present
	Robert Jacobs <i>Professor of Brain and Cognitive Science, University of Rochester</i>	Fall 2018
	Baojie Fan <i>Associate Professor, Nanjing University of Posts and Telecommunications</i>	Summer 2018—Spring 2019

STUDENT ADVISING	PH.D. ADVISEES	
	D12. Mingqian Feng (UR CS)	2023—Present
	D11. Yunlong Tang (UR CS)	2023—Present
	D10. Zeliang Zhang (UR CS)	2022—Present
	D9. Susan Liang (UR CS)	2022—Present
	D8. Nguyen Manh Nguyen (UR CS)	2022—Present
	D7. Luchuan Song (UR CS)	2022—Present
	D6. Chao Huang (UR CS)	2021—Present
	D5. Jing Bi (UR CS)	MS: 2018—2020; PhD: 2020—Present
	D4. Zhiheng Li (UR CS)	UG: Summer 2017; PhD: Graduated 2023
	DISSERTATION: <i>Discover and Mitigate Biases in Discriminative and Generative Image Models</i>	
	FIRST APPOINTMENT: <i>Research Scientist at Amazon AWS</i>	
	D3. Jing Shi (UR CS)	PhD: Graduated 2022
	DISSERTATION: <i>Vision Based Language to Action Mapping</i>	
	FIRST APPOINTMENT: <i>Research Scientist at Adobe Research</i>	
	D2. Yapeng Tian (UR CS)	PhD: Graduated 2022
	DISSERTATION: <i>Audio-Visual Scene Understanding Towards Unified, Explainable, and Robust Multisensory Perception</i>	
	FIRST APPOINTMENT: <i>Assistant Professor at UT Dallas</i>	
	D1. Lele Chen (UR CS)	PhD: Graduated 2022
	DISSERTATION: <i>High-Fidelity Talking Avatar Video Generation</i>	
	FIRST APPOINTMENT: <i>Staff Research Scientist at Oppo US Research</i>	

### PH.D. COMMITTEE

S32. Nathan Kent (UR CS)	
S31. Max Wasserman (UR CS)	
S30. Meng-Huan Wu (UR BCS/CS)	
S29. Ruoyang Hu (UR BCS/CS)	
S28. Zhexin Xu (UR BCS/CS)	
S27. Yiming Gan (UR CS)	Graduated 2023
S26. Yu Feng (UR CS)	Graduated 2023
S25. Yu Chen (UR CS)	Graduated 2023
S24. Songyang Zhang (UR CS)	Graduated 2023
S23. Haitian Zheng (UR CS)	Graduated 2023
S22. Wei Zhu (UR CS)	Graduated 2023



S21. Ali K. Shargh (UR ME)	Graduated 2022
S20. Sabyasachi Shivkumar (UR BCS/CS)	Graduated 2022
S19. Weijian Li (UR CS)	Graduated 2022
S18. Wei Xiong (UR CS)	Graduated 2022
S17. Joseph German (UR BCS/CS)	Graduated 2022
S16. Lisa Jin (UR CS)	Graduated 2022
S15. Zhengyuan Yang (UR CS)	Graduated 2021
S14. Shupeng Gui (UR CS)	Graduated 2021
S13. Hanlin Tang (UR CS)	Graduated 2021
S12. Samiha Samrose (UR CS)	Graduated 2021
S11. Taylan Sen (UR CS)	Graduated 2021
S10. Nabil Hossain (UR CS)	Graduated 2020
S9. Haichuan Yang (UR CS)	Graduated 2020
S8. Feng Yang (UR CS)	Graduated 2020
S7. Christopher Bates (UR BCS/CS)	Graduated 2020
S6. Bochen Li (UR ECE)	Graduated 2020
S5. Kushal Kafle (Rochester Institute of Technology)	Graduated 2020
S4. Haofu Liao (UR CS)	Graduated 2019
S3. Jianbo Yuan (UR CS)	Graduated 2019
S2. Yichi Zhang (UR ECE)	Graduated 2019
S1. Sefik Emre Eskimez (UR ECE)	Graduated 2019

#### MS STUDENTS

M19. Rongyi Zhu (Fall 2022 - Present, URCS)
M18. Shijian Deng (Fall 2021 - Fall 2022, URCS then Ph.D. student at UT Dallas)
M17. Guangyu Sun (Fall 2020 - Summer 2022, URCS then Ph.D. student at UCF)
M16. Rohan Sharma (Summer - Fall 2020, UR GIDS then Ph.D. student at SUNY Buffalo)
M15. Guo Li (Summer 2019 - Spring 2020, URCS then SDE at Amazon)
M14. Guofeng Cui (Spring 2019 - Spring 2020, URCS then Ph.D. student at Rutgers University)
M13. Ziyi Kou (Fall 2018 - Spring 2020, URCS then Ph.D. student at Univ. of Notre Dame)
M12. Justin Tian (Fall 2018 - Spring 2019, URCS)
M11. Jing Bi (Summer 2018 - Summer 2020, UR ECE then Ph.D. student at URCS)
M10. Wei Zhang (Summer 2018 - Spring 2019, URCS then SDE at Amazon)
M9. Hao Huang (Summer 2018 - Spring 2019, URCS then Ph.D. student at NYU)
M8. Wentian Zhao (Summer 2018 - Spring 2019, URCS then ML Engineer at Adobe)
M7. Shaojie Wang (Summer 2018 - Spring 2019, URCS then Ph.D. student at WUSTL)
M6. Jie Chen (Fall 2017 - Fall 2019 then Microsoft Research)
M5. Li Ding (Summer 2017, UR GIDS then Research Engineer at MIT)
M4. Sudhanshu Srivastava (Spring - Summer 2017, URCS then Ph.D. student at UCSB)
M3. Lele Chen (Spring 2017 - Spring 2018, URCS then Ph.D. student at URCS)
M2. Mingyang Zhou (Spring 2016, Univ. of Michigan then Ph.D. student at UC Davis)
M1. Yao Li (Spring 2013, SUNY Buffalo)

#### UNDERGRADUATES

U29. Pinxin (Andy) Liu (Spring 2023 - Present, URCS)
U28. Ji-Ze Jang (Summer 2021 - Summer 2022, URCS)
U27. Miranda Rublaitus (Summer 2022, NSF REU, Yale University)
U26. Kalsey Colotl (Summer 2022, NSF REU, NYU)

- U25. Sizhe Li (Spring 2020 - Fall 2022, URCS)  
*Finalist for the CRA Outstanding UG Researcher Awards 2023*
- U24. Phuong Vu (Fall 2021, URCS)
- U23. Yiyang Su (Summer 2020 - Spring 2021, URCS then Ph.D. student at Michigan State Univ.)
- U22. Logan Peters (Summer 2020, Carleton College)
- U21. Theodore Chapman (Summer 2020, NSF REU, URCS)
- U20. Patrick Phillips (Summer 2020, NSF REU, URCS)
- U19. Bryce Elizabeth Yahn (Summer 2020, NSF REU, UR BCS)
- U18. Tianyou Xiao (Summer 2018 - Fall 2019, URCS then MS student at Cornell Tech)
- U17. Wenxuan Bao (Summer - Fall 2019, Tsinghua University then Ph.D. student at UIUC)
- U16. Ariel Tello (Summer 2019, Xero Research Fellow, URCS)
- U15. Chenxiao Guan (Summer 2018, Xerox Research Fellow, UR CE then MS student at CMU)
- U14. Qiuyue Sun (Summer 2018, URCS)
- U13. Justin Goodman (Summer 2018, NSF REU, University of Maryland)
- U12. Marc Moore (Summer 2018, NSF REU, Mississippi State University)
- U11. Amos Newswanger (Summer - Fall 2017, URCS)
- U10. Max Torop (Summer 2017, URCS then MS student at WUSTL)
- U9. Zhiheng Li (Summer-Fall 2017, Wuhan University then Ph.D. student at URCS)
- U8. Wei Zhao (Summer 2017, USTC)
- U7. Wes Smith (Summer 2017, NSF REU, University of Edinburgh then Ph.D. student at URCS)
- U6. Moses Bug (Summer 2017, NSF REU, Brandeis University)
- U5. Shengqi Suizhu (Fall 2016 - Spring 2017, URCS)
- U4. Austin Schaffer (Spring 2014, SUNY Buffalo)
- U3. Libing Wu (Spring 2014, SUNY Buffalo)
- U2. Spencer Whitt (Summer 2013, SUNY Buffalo)
- U1. Tyler Ganter (Summer 2013, SUNY Buffalo)

## SOFTWARE & DATA SETS

- 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/>
- 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/>
- 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/>
- 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>
- 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/>

## **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