

Chenliang Xu

Prepared on July 9, 2020

CONTACT	3005 Wegmans Hall 250 Hutchison Road University of Rochester Rochester, NY 14627	PHONE: (585) 275-5426 EMAIL: chenliang.xu@rochester.edu WEB: www.cs.rochester.edu/~cxu22
CURRENT APPOINTMENT	Assistant Professor Department of Computer Science University of Rochester	ROCHESTER, NY 9/2016 - PRESENT
RESEARCH INTERESTS	My research interests range from solving core computer vision problems to pushing the limits of their understanding in the broader AI contexts, including multisensory perception, cognitive robotics, and data science.	
RESEARCH AREA	Computer Vision Deep Learning Multimodal Modeling	
EDUCATION	University of Michigan Ph.D. in Computer Science and Engineering ADVISOR: Jason J. Corso DISSERTATION TITLE: <i>"Scale-Adaptive Video Understanding"</i>	ANN ARBOR, MI 8/2016
	State University of New York at Buffalo M.S. in Computer Science and Engineering	BUFFALO, NY 2/2012
	Nanjing University of Aeronautics and Astronautics B.S. in Information and Computing Science	NANJING, CHINA 6/2010
DISTINCTIONS	Best Paper Award at 17th ACM SIGGRAPH VRCAI <i>for the paper "3D Human Avatar Digitization from a Single Image"</i>	2019
	National Science Foundation BIGDATA Award <i>Audio-Visual Scene Understanding</i>	2017
	Best Paper Award at 14th Sound and Music Computing Conference <i>for the paper "Audio-visual Source Association for String Ensembles through Multi-modal Vibrato Analysis"</i>	2017
	University of Rochester AR/VR Pilot Grant Award <i>Real-Time Synthesis of a Virtual Talking Face from Acoustic Speech</i>	2017
	Rackham Conference Travel Grant – University of Michigan	2015, 2016
	Travel Award of IEEE ICCV 2013	2013
	Best Open Source Code Third Prize at IEEE CVPR 2012	2012
	Best Demo Prize at the 2nd Greater New York Multimedia and Vision Meeting	2012
	First Tier Jiangsu Province Award – Mathematics Modeling Competition in China	2007
	Outstanding Student Scholarship – Nanjing Univ. of Aero. and Astro.	2007, 2008, and 2009

Qualifiers added where known: Acceptance Rate (AR), Impact Factor (IF), h5-Index (h5) provided by Google Scholar metrics.

2020

- P52. Yapeng Tian, Dingzeyu Li, and Chenliang Xu. Unified multisensory perception: Weakly-supervised audio-visual video parsing. In *European Conference on Computer Vision*, 2020. **Spotlight Presentation: 5% (ECCV)**
- P51. 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*, 2020. **(ECCV)**
- P50. 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: 5.7% (CVPR)**
- P49. 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)**
- P48. 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)**
- P47. 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)**
- P46. 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*, 2020. **(CVPRW)**
- P45. 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*, 2020. **(CVPRW)**
- P44. 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*, 2020. **(CVPRW)**
- P43. 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*, 2020. **(CVPRW)**
- P42. Jing Bi, Vikas Dhiman, Tianyou Xiao, and Chenliang Xu. Learning from interventions with hierarchical policies for safe learning. In *AAAI Conference on Artificial Intelligence*, pages 10352–10360, 2020. **Oral Presentation: under 5% (AAAI)**
- P41. 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)**
- P40. 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)**

¹CVPR, ICCV and ECCV are premier conferences in computer vision. For each, typical number of submissions is several thousand and the overall acceptance rate is under 30%. CVPR is the highest rated publication venue for computer vision and eighth-highest across all engineering and computer science, according to Google Scholar metrics.

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*, number 12, 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. AR: 25.2%; h5: 188 (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. AR: 25.2%; h5: 188 (CVPR)
- P36. 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; h5: 47 (CVPRW)**
- P35. 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. h5: 47 (CVPRW)
- P34. 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. h5: 47 (CVPRW)
- P33. 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. AR: 17.9%; h5: 61 (IJCAI)
- P32. 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. AR: 28% (BMVC)
- P31. 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. In Press, IF: 11.5 (IJCV)
- P30. Robert A. Jacobs and Chenliang Xu. Can multisensory training aid visual learning?: A computational investigation. *Journal of Vision*, 19(11):1–12, 2019. IF: 3.8 (JOV)
- P29. 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. In Press, IF: 1.9 (TASLP)
- P28. 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)
- P27. 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)

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. AR: 31.8%; h5: 104 (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.
AR: 31.8%; h5: 104 (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.
AR: 28%; h5: 188 (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; AR: 24%; h5: 59 (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.
AR: 37%; h5: 38 (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)
- 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.
IF: 1.8 (JASA)
- 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.
IF: 2.8 (JCP)
- 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.
AR: 28%; h5: 188 (CVPR)
- P15. Amos Newschwanger 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.
h5: 47 (CVPRW)
- P14. 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.
(ACMMMWS)
- P13. 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.
(ACMMMWS)
- P12. 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.
IF: 4.0; h5: 63 (TMM)

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: 9.7%; h5: 188 (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. **IF: 11.5; h5: 65 (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. **AR: 28%; h5: 188 (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. *Selected as a Best Paper from ICSC.* **h5: 12 (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. **AR: 28%; h5: 124 (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. **AR: 30%; h5: 15 (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. **AR: 26%; h5: 188 (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: 2.8%; h5: 104 (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. **AR: 24%; h5: 188 (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. **(NIST Workshop)**
- 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. **(NIST Workshop)**

TECHNICAL REPORTS AND THESES

- T12. Samuel Lerman, Chenliang Xu, Charles Venuto, and Henry Kautz. Explaining local, global, and higher-order interactions in deep learning. *arXiv preprint arXiv:2006.08601*, 2020.
- T11. 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.
- T10. Burkay 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.

- T9. Yapeng Tian, Chenliang Xu, and Dingzeyu Li. Deep audio prior. *arXiv preprint arXiv:1912.10292*, 2019.
- T8. Guofeng Cui, Ziyi Kou, Shaojie Wang, Wentian Zhao, and Chenliang Xu. Improve cam with auto-adapted segmentation and co-supervised augmentation. *arXiv preprint arXiv:1911.07160*, 2019.
- T7. Haitian Zheng, Lele Chen, Chenliang Xu, and Jiebo Luo. Unsupervised pose flow learning for pose guided synthesis. *arXiv preprint arXiv:1909.13819*, 2019.
- T6. 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.
- T5. Shaojie Wang, Wentian Zhao, Ziyi Kou, and Chenliang Xu. How to make a blt sandwich? learning to reason towards understanding web instructional videos. *arXiv preprint arXiv:1812.00344*, 2018.
- 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

Total Funding: \$2.1 million (\$1.5 million as PI)

Funding is sorted by start date (recent first).

- F11. **PI: III: Small: Collaborative Research: Scalable Deep Bayesian Tensor Decomposition** \$199,134
SOURCE: Information & Intelligent System, NSF 10/2019-9/2022
COLLABORATORS: Shandian Zhe (PI, Univ. of Utah)
OBJECTIVE: The project aims to develop flexible and efficient tensor decomposition approaches that can discover a variety of complicated relationships between the entities in tensors, handle a tremendous amount of data from practical applications, and adapt to rapid data growth.
- F10. **PI: Anomaly Detection for Videos Applied in Manufacturing Settings** \$70,259
SOURCE: Center of Excellence in Data Science & Corning Inc. 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.
- F9. **PI: Research Gift** \$28,500
SOURCE: Adobe Fall 2019
- F8. **PI: Synthesizing Fashion Clothe** \$55,000
SOURCE: Viscovery 9/2018-8/2019
OBJECTIVE: The main objective is to design novel models to synthesize fashion clothe by different design attributes.
- F7. **PI: RI: Small: Learning Dynamics and Evolution towards Cognitive Understanding of Videos** \$465,990
SOURCE: Robust Intelligence, NSF 9/2018-8/2021
COLLABORATORS: Jiebo Luo (Co-PI, CS)
OBJECTIVE: The main objective is to build computational models to study higher-order inference in understanding web instructional videos.
- F6. **Co-PI: CDS&E: D3SC: Applying Video Segmentation to Coarse-grain Mapping Operators in Molecular Simulations** \$488,605
8/2018-7/2021

SOURCE: Division of Chemistry, NSF
 COLLABORATORS: Andrew White (PI, CE)
 OBJECTIVE: The main objective is to apply advances in computer vision, e.g., video segmentation and action recognition, to improve models of multiscale systems in chemistry.

- F5. **PI:** *Deep Reinforcement Learning for Instructional Video Grounding* \$60,000
 SOURCE: Tencent America LLC, AI Lab Rhino-Bird Program 6/2018-5/2019
 OBJECTIVE: The main objective is to develop a novel deep reinforcement learning framework to perform higher-order inference in understanding web instructional videos.
- F4. **PI:** *BIGDATA: F: Audio-Visual Scene Understanding* \$666,000
 SOURCE: BIGDATA Program, NSF 9/2017-8/2021
 COLLABORATORS: Zhiyao Duan (Co-PI, ECE)
 OBJECTIVE: The main objective is to 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.
- F3. **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
 COLLABORATORS: Ross K. Maddox (PI, BOE), Zhiyao Duan (Co-PI, ECE)
 OBJECTIVE: Develop a real-time talking face by analyzing the correlations between auditory and visual signals to assist hearing-impaired people.
- F2. **PI:** *Video Understanding Beyond Human Activities* GPU Donation
 SOURCE: NVIDIA GPU Grant Program 8/2017
 OBJECTIVE: The main objective is to achieve a joint recognition of actors and their actions in videos that overcomes the limitations of separated or cascaded approaches.
- F1. **PI:** *Real-Time Streaming Video Object Detection* \$30,000
 SOURCE: Markable, Inc. 8/2017-1/2018
 OBJECTIVE: Develop an online real-time video object detection system that handles various video challenges, such as motion blur, video defocus, object occlusion and rare pose.

PROFESSIONAL
 SERVICES

ORGANIZING COMMITTEE

- * IEEE CVPR Workshop *Brave New Ideas For Motion and Spatio-Temporal Representations* 2017

PROGRAM COMMITTEE

- * IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016—2020
- * The AAAI Conference on Artificial Intelligence (AAAI) 2017—2020
- * European Conference on Computer Vision (ECCV) 2018, 2020
- * IEEE International Conference on Computer Vision (ICCV) 2015, 2017, 2019
- * British Machine Vision Conference (BMVC) 2015—2019
- * International Conference on Semantic Computing (ICSC) 2017
- * Neural Information Processing Systems (NIPS) 2016
- * Asian Conference on Computer Vision (ACCV) 2016, 2018
- * International Symposium CompIMAGE'16 2016
- * IEEE/ISPRS 3rd Joint Workshop on Multi-Sensor Fusion for Dynamic Scene Understanding 2015
- * ISPRS Geospatial Week – Image Sequence Analysis 2015
- * XXIIIrd ISPRS Congress 2016
- * Indian Conference on Computer Vision, Graphics and Image Processing 2014

JOURNAL REVIEWER

- * IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)
- * IEEE Transactions on Image Processing (IEEE TIP)
- * IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT)
- * IEEE Transactions on Multimedia (IEEE TMM)
- * International Journal of Computer Vision (IJCV)
- * Pattern Recognition (PR)
- * Image and Vision Computing (IVC)
- * Computer Vision and Image Understanding (CVIU)
- * Signal Processing: Image Communication
- * Machine Vision and Applications
- * IET Computer Vision
- * OSA Biomedical Optics Express
- * Electronic Letters on Computer Vision and Image Analysis

* IPSJ Transactions on Computer Vision and Applications

PANELIST

* National Science Foundation (CISE/CNS) 2019
* National Science Foundation (CISE/IIS) 2019

UNIVERSITY SERVICES

Outstanding Dissertation Award Committee (University of Rochester) 2019
PhD Admission Committee (CS, University of Rochester) 2016-20
UG Curriculum Committee (CS, University of Rochester) 2019-20
Lab Committee (CS, University of Rochester) 2018-19
Website Committee (CS, University of Rochester) 2017-18
Faculty Search Committee (CS, University of Rochester) 2018-19
MS Admission Committee (GIDS, University of Rochester) 2017-19

INVITED TALKS

Audio-Visual Scene Understanding
- Corning Inc. Machine Learning Group 11/2019
Audio-Visual Sensing & Efforts toward Building an AI Cognitive Loop
- Laboratory For Laser Energetics, University of Rochester 12/2018
When Computer Vision Meets Audition:
From Cross-Modal Generation to Audio-Visual Scene Understanding
- Western New York Image and Signal Processing Workshop 10/2018
- Data Science Summer Colloquium Series, University of Rochester 6/2018
- Rochester Institute of Technology 11/2017
Applications of Multimodal Deep Learning
- URMC Clinical and Translational Science Institute Analytics Colloquium 4/2018
Towards the What, Who and Where of Video Understanding
- The DAVIS Challenge on Video Object Segmentation, IEEE CVPR Workshop 7/2017
Actor-Action Semantic Segmentation with Grouping Process Models
- Midwest Vision Workshop 4/2016
Scale-Adaptive Video Understanding
- School of Computing, University of Utah 3/2016
- Department of Computer Science, Stevens Institute of Technology 3/2016
- Department of Computer Science, University of Rochester 3/2016
- Department of Computer Science & Engineering, Michigan State University 2/2016
Pulling Information from Scales
- SPEECS Seminar, University of Michigan 1/2016
Action Understanding with Multiple Classes of Actors
- AI Lab Mini-Symposium, University of Michigan 5/2015
Hierarchical Video Segmentation: Methods, Perception and Application
- SPEECS Seminar, University of Michigan 9/2014
LIBSVX and Video Segmentation Evaluation
- IEEE CVPR 2014 Tutorial on Video Segmentation 6/2014
Are Actor and Action Semantics Retained in Video Supervoxel Segmentation?
- IEEE International Conference on Semantic Computing 9/2013

TEACHING

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

University of Rochester, Summer 2017, Summer 2018

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

Pre-College Data Science Program

Hajim Engineering, University of Rochester, Summer 2018

Teach a deep learning tutorial to a group of 17 high school students (including from South Korea, China, Spain and Haiti).

CSC 249/449 Machine Vision

University of Rochester, S'18, S'19, S'20

Instructor for the cross-listed undergraduate- and graduate-level course that covers both basics and advances in modern computer vision.

CSC 298/578 Deep Learning and Graphical Models

University of Rochester, Spring 2017

Instructor for the cross-listed undergraduate- and graduate-level course that covers both essentials in deep learning and probabilistic graphical models. It is a new CS course at UR.

CSC 577 Advanced Topics in Computer Vision

University of Rochester, F'16, F'17, F'18, F'19

Instructor for the graduate-level seminar course that covers special topics in computer vision.

EECS 598 Foundations of Computer Vision

University of Michigan, Fall 2015

Graduate student instructor for this course. Lecture tutorials and reviews, hold office hours, and grade student presentations.

Beyond the M.S. Mentoring Program

University of Michigan, Fall 2015

Encourage and provide guidance for master's students to continue on for a Ph.D.

STUDENT ADVISING

PH.D. ADVISEES

D4. Zhiheng Li (CS, University of Rochester)	UG: SUMMER'17; PHD: FALL 2018 - PRESENT
D3. Lele Chen (CS, University of Rochester)	MS: S'17-S'18; PHD: FALL 2018 - PRESENT
D2. Jing Shi (CS, University of Rochester)	FALL 2017 - PRESENT
D1. Yapeng Tian (CS, University of Rochester)	FALL 2017 - PRESENT

PH.D. COMMITTEE

S23. Lisa Jin (CS, University of Rochester)	
S22. Fangzhou Liu (CS, University of Rochester)	
S21. Hanlin Tang (CS, University of Rochester)	
S20. Yu Chen (CS, University of Rochester)	
S19. Shupeng Gui (CS, University of Rochester)	
S18. Haichuan Yang (CS, University of Rochester)	
S17. Samiha Samrose (CS, University of Rochester)	
S16. Taylan Sen (CS, University of Rochester)	
S15. Weijian Li (CS, University of Rochester)	
S14. Wei Xiong (CS, University of Rochester)	
S13. Zhengyuan Yang (CS, University of Rochester)	
S12. Feng Yang (CS, University of Rochester)	
S11. Samuel Lerman (CS, University of Rochester)	
S10. Nabil Hossain (CS, University of Rochester)	
S9. Sabyasachi Shivkumar (BCS, University of Rochester)	
S8. Bin Yang (BCS, University of Rochester)	
S7. Christopher Bates (BCS, University of Rochester)	
S6. Bochen Li (ECE, University of Rochester)	
S5. Kushal Kafle (Rochester Institute of Technology)	
S4. Haofu Liao (CS, University of Rochester)	Graduated 2019
S3. Jianbo Yuan (CS, University of Rochester)	Graduated 2019
S2. Yichi Zhang (ECE, University of Rochester)	Graduated 2019
S1. Sefik Emre Eskimez (ECE, University of Rochester)	Graduated 2019

MS STUDENTS

M15. Tushar Kumar (CS, University of Rochester)	SUMMER 2019 - PRESENT
M14. Guo Li (CS, University of Rochester)	SUMMER 2019 - PRESENT
M13. Guofeng Cui (CS, University of Rochester)	SPRING 2019 - PRESENT

M12. Ziyi Kou (CS, University of Rochester)	FALL 2018 - PRESENT
M11. Jing Bi (ECE, University of Rochester)	SUMMER 2018 - PRESENT
M10. Jie Chen (CS, University of Rochester)	FALL 2017 - FALL 2019
M9. Justin Tian (CS, University of Rochester)	FALL 2018 - SPRING 2019
M8. Hao Huang (CS, University of Rochester)	SUMMER 2018 - SPRING 2019
M7. Wei Zhang (CS, University of Rochester)	SUMMER 2018 - SPRING 2019
M6. Wentian Zhao (CS, University of Rochester)	SUMMER 2018 - SPRING 2019
M5. Shaojie Wang (CS, University of Rochester)	SUMMER 2018 - SPRING 2019
M4. Li Ding (DS, University of Rochester)	SUMMER 2017
M3. Sudhanshu Srivastava (CS, University of Rochester)	SPRING, SUMMER 2017
M2. Mingyang Zhou (ECE, University of Michigan)	SPRING 2016
M1. Yao Li (CSE, SUNY at Buffalo)	SPRING 2013

UNDERGRADUATES

U17. Tianyou Xiao (CS, University of Rochester)	SUMMER 2018 - PRESENT
U16. Wenxuan Bao (Visiting Student from Tsinghua university)	SUMMER, FALL 2019
U15. Ariel Tello (Xerox Fellow, CS, University of Rochester)	SUMMER 2019
U14. Chenxiao Guan (Xerox Fellow, CE, University of Rochester)	SUMMER 2018
U13. Qiuyue Sun (CS, University of Rochester)	SUMMER 2018
U12. Justin Goodman (REU Student from University of Maryland)	SUMMER 2018
U11. Marc Moore (REU Student from Mississippi State University)	SUMMER 2018
U10. Amos Newswanger (CS, University of Rochester)	SUMMER 2017, FALL 2017
U9. Max Torop (DS, University of Rochester)	SUMMER 2017
U8. Wei Zhao (Visiting Student from USTC)	SUMMER 2017
U7. Wes Smith (REU Student from University of Edinburgh)	SUMMER 2017
U6. Moses Bug (REU Student from Brandeis University)	SUMMER 2017
U5. Shengqi Suizhu (CS, University of Rochester)	FALL 2016
U4. Austin Schaffer (CS, SUNY at Buffalo)	SPRING 2014
U3. Libing Wu (CS, SUNY at Buffalo)	SPRING 2014
U2. Spencer Whitt (CS, SUNY at Buffalo)	SUMMER 2013
U1. Tyler Ganter (CS, SUNY at Buffalo)	SUMMER 2013

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

PAST POSITIONS

Research Assistant ANN ARBOR, MI
Electrical Engineering and Computer Science 9/2014 - 8/2016
University of Michigan
ADVISOR: Jason J. Corso
PRIMARY FOCUS: Investigating unified models of video representation integrating probabilistic methods, machine learning, and dynamic adaptive graphs to advance our ability to solve the video understanding problem.

Research Intern CUPERTINO, CA
NEC Lab American, INC. 5/2014 - 8/2014
MENTOR: Manmohan Chandraker
PROJECT: Video labeling and 3D localization for monocular road scene understanding.

Research Assistant BUFFALO, NY
Computer Science and Engineering 8/2011 - 5/2014
State University of New York at Buffalo
ADVISOR: Jason J. Corso
PRIMARY FOCUS: Improving the representation and indexing of objects and events in large-scale video analysis by efficiently encoding the low-level perceptual entities in the video and grounding them with rich high-level semantics.