

RUOHAN GAO

Computer Science Department
Stanford University
353 Jane Stanford Way
Stanford, CA 94305

E-mail: rhgao@cs.stanford.edu
<http://www.ai.stanford.edu/~rhgao>

CURRENT APPOINTMENT

Stanford University, Stanford, CA
SAIL Postdoctoral Fellow, Computer Science, Feb. 2021 - Present

EDUCATION

The University of Texas at Austin, Austin, TX
Ph.D. in Computer Science, Jan. 2021
– *Advisor: Kristen Grauman*
– *Thesis: Look and Listen: From Semantic to Spatial Audio-Visual Perception*
– *2021 Michael H. Granof University's Best Doctoral Dissertation Award*
– *2019 Google PhD Fellowship and Adobe Research Fellowship*

The Chinese University of Hong Kong, Hong Kong
B.Eng. in Information Engineering, *First Class Honours*, July 2015

RESEARCH INTERESTS

Computer Vision: audio-visual learning, embodied learning with multiple modalities, multi-modal video understanding, learning from unlabeled videos, self-supervised learning

Machine Learning: multi-modal deep learning, transfer learning, multi-task learning

Robotics: robot learning with multisensory data

APPOINTMENTS

Facebook AI Research (FAIR), Austin, TX
Visiting Researcher March 2020 - Dec. 2020

The University of Texas at Austin, Austin, TX
Graduate Student Researcher Jan. 2016 - Dec. 2020

Facebook AI (Applied) Research (FAIR/FAIAR), Cambridge, MA
Research Intern June 2019 - Sept. 2019

Facebook AI Research (FAIR), Menlo Park, CA
Research Intern May 2018 - Aug. 2018

HONORS AND AWARDS

- Best Paper Award Runner Up, BMVC 2021
- Michael H. Granof Award, UT Austin's Top 1 Dissertation of 2021
- Outstanding Dissertation Award in Mathematics, Engineering, Physical Sciences, and Biological and Life Sciences, UT Austin, 2021
- Google PhD Fellowship, 2019 - 2021

- Outstanding Reviewer Award, CVPR, 2020
- Graduate Dean's Prestigious Fellowship Supplement Fellow, UT Austin, 2019 & 2020
- Adobe Research Fellowship, 2019
- Best Paper Award Finalist, CVPR 2019
- Sir Run Run Shaw Postgraduate Scholarship, CUHK, 2015
- Dean's List, Engineering Faculty, CUHK, 2014 & 2015
- Q W Lee Scholarship, Top Academic Excellence Scholarship, CUHK, 2014
- National Scholarship, Ministry of Education of China, 2013

INVITED AND CONFERENCE TALKS

- Intelligent Sensing Winter School
Invited Talk, Queen Mary University of London, Virtual Dec. 2021
- Mitsubishi Electric Research Labs (MERL)
Invited Talk, MERL Seminar Series, Virtual Sept. 2021
- Special Session on Low-Power Computer Vision
Invited Talk, AICAS, Virtual June 2021
- Stanford University
Interactive Perception and Robot Learning Lab, Stanford, CA April 2021
- Stanford University
Stanford Vision and Learning Lab, Stanford, CA March 2021
- Facebook Reality Lab
Invited Talk, FRL Audio, Seattle, WA Dec. 2020
- The University of California, Berkeley
Invited Talk, Computer Vision Seminar, Berkeley, CA May 2020
- Massachusetts Institute of Technology
Invited Talk, Spoken Language Systems Group, Cambridge, MA April 2020
- The University of Texas at San Antonio
Invited Talk, AI Consortium Seminar Series, San Antonio, TX April 2020
- Massachusetts Institute of Technology
Invited Talk, Vision Seminar, Cambridge, MA Sept. 2019
- Adobe Research
Adobe Research Fellowship Ceremony, San Jose, CA Aug. 2019
- Google
Lightning Talk, PhD Fellowship Summit, Mountain View, CA July 2019
- Sight and Sound
Invited Talk, CVPR Workshop, Long Beach, CA June 2019
- Conference on Computer Vision and Pattern Recognition (CVPR), 2019
Oral Presentation, Long Beach, CA June 2019
- European Conference on Computer Vision (ECCV), 2018
Oral Presentation, Munich, Germany Oct. 2018
- Sight and Sound
Oral Presentation, CVPR Workshop, Salt Lake City, Utah June 2018
- Conference on Computer Vision and Pattern Recognition (CVPR), 2018
Oral Presentation, Salt Lake City, Utah June 2018

PEER-REVIEWED CONFERENCE PUBLICATIONS

1. **Ruohan Gao**, Yen-Yu Chang, Shivani Mall, Li Fei-Fei, Jiajun Wu, “ObjectFolder: A Dataset of Objects with Implicit Visual, Auditory, and Tactile Representations”, in *Proceedings of the Conference on Robot Learning (CoRL)*, Nov. 2021.
2. Samuel Clarke, Negin Heravi, Mark Rau, **Ruohan Gao**, Jiajun Wu, Doug James, Jeannette Bohg, “DiffImpact: Differentiable Rendering and Identification of Impact Sounds”, in *Proceedings of the Conference on Robot Learning (CoRL)*, Nov. 2021. (**oral presentation**, 6.5% acceptance rate)
3. Rishabh Garg, **Ruohan Gao**, Kristen Grauman, “Geometry-Aware Multi-Task Learning for Binaural Audio Generation from Video”, in *Proceedings of the British Machine Vision Conference (BMVC)*, Nov. 2021. (**oral presentation**, 3.3% acceptance rate) [**Best Paper Award Runner Up**]
4. **Ruohan Gao** and Kristen Grauman, “VisualVoice: Audio-Visual Speech Separation with Cross-Modal Consistency”, in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021.
5. Changan Chen, Sagnik Majumder, Ziad Al-Halah, **Ruohan Gao**, Santhosh K. Ramakrishnan and Kristen Grauman, “Learning to Set Waypoints for Audio-Visual Navigation,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, May 2021.
6. **Ruohan Gao**, Changan Chen, Ziad Al-Halah, Carl Schissler, and Kristen Grauman, “VisualEchoes: Spatial Image Representation Learning through Echolocation”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, Aug. 2020.
7. **Ruohan Gao**, Tae-Hyun Oh, Kristen Grauman, and Lorenzo Torresani, “Listen to Look: Action Recognition by Previewing Audio”, in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
8. **Ruohan Gao** and Kristen Grauman, “Co-Separating Sounds of Visual Objects”, in *Proceedings of the International Conference on Computer Vision (ICCV)*, Oct. 2019.
9. **Ruohan Gao** and Kristen Grauman, “2.5D Visual Sound”, in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. (**oral presentation**, 5.6% acceptance rate) [**Best Paper Award Finalist**]
10. **Ruohan Gao**, Rogerio Feris, and Kristen Grauman, “Learning to Separate Object Sounds by Watching Unlabeled Video”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, Sept. 2018. (**oral presentation**, 2.4% acceptance rate)
11. Dinesh Jayaraman, **Ruohan Gao**, and Kristen Grauman, “ShapeCodes: Self-Supervised Feature Learning by Lifting Views to Viewgrids”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, Sept. 2018.
12. **Ruohan Gao**, Bo Xiong, and Kristen Grauman, “Im2Flow: Motion Hallucination from Static Images for Action Recognition”, in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. (**oral presentation**, 2.1% acceptance rate)
13. **Ruohan Gao** and Kristen Grauman, “On-demand Learning for Deep Image Restoration”, in *Proceedings of the International Conference on Computer Vision (ICCV)*, Oct. 2017.
14. **Ruohan Gao**, Dinesh Jayaraman, and Kristen Grauman, “Object-centric Representation Learning from Unlabeled Videos”, in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, Nov. 2016.
15. **Ruohan Gao**, Huanle Xu, Pili Hu, and Wing Cheong Lau, “Accelerating Graph Mining Algorithms via Uniform Random Edge Sampling”, in *Proceedings of IEEE International Conference on Communications (ICC)*, May 2016.
16. **Ruohan Gao**, Pili Hu, and Wing Cheong Lau, “Property Preservation under Community-Based Sampling”, in *Proceedings of the Global Communications Conference (GLOBECOM)*, Dec. 2015.

17. **Ruohan Gao**, Huanle Xu, Pili Hu, and Wing Cheong Lau, “Accelerating Graph Mining Algorithms via Uniform Random Edge Sampling (Poster)”, in *ACM Conference on Online Social Networks (COSN)*, Nov, 2015.

PEER-REVIEWED WORKSHOP PAPERS AND ABSTRACTS

1. **Ruohan Gao** and Kristen Grauman, “2.5D Visual Sound”, in *Learning From Unlabeled Videos Workshop at the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
2. **Ruohan Gao** and Kristen Grauman, “2.5D Visual Sound”, in *Multi-Modal Learning from Videos Workshop at the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
3. **Ruohan Gao**, Rogerio Feris, and Kristen Grauman, “Learning to Separate Object Sounds by Watching Unlabeled Video,” in *Sight and Sound Workshop at the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

THESES

- Ruohan Gao. “Look and Listen: From Semantic to Spatial Audio-Visual Perception”. Ph.D. Thesis, The University of Texas at Austin, 2021.

PATENTS

- US20210174817A1: Systems and Methods for Visually Guided Audio Separation, 2021.

TEACHING EXPERIENCE

Stanford University, Stanford, CA

Guest Lecture

CS231N: Convolutional Neural Networks for Visual Recognition

– Multimodal Learning

Spring 2021

The University of Texas at Austin, Austin, TX

Guest Lecture and Tutorials

CS381V: Visual Recognition

– Introduction to Deep Learning

– Implementation of CNNs

Fall 2017

The University of Texas at Austin, Austin, TX

Teaching Assistant

CS303E: Elements of Computers and Programming

CS429: Computer Organization and Architecture

Spring 2016

Fall 2015

ENTREPRENEURSHIP EXPERIENCE

Open Innovation Lab, Hong Kong

Curator and Co-founder

Open Innovation Lab @ CUHK

– The hub for advancing open source, open data, open culture and entrepreneurship

Sept. 2014 - Feb. 2015

PROFESSIONAL SERVICES

Organizer

Co-Organizer, Workshop on Sight and Sound, CVPR 2021

Lead Organizer, Workshop on Embodied Multimodal Learning, ICLR 2021
 Co-Organizer, Workshop on Sight and Sound, CVPR 2020

Conference Program Committee Member / Reviewer

IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
 IEEE International Conference on Computer Vision (ICCV)
 European Conference on Computer Vision (ECCV)
 ACM International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)
 International Conference on Learning Representations (ICLR)
 Conference on Neural Information Processing Systems (NeurIPS)
 International Conference on Machine Learning (ICML)
 IEEE International Conference on Robotics and Automation (ICRA)
 AAAI Conference on Artificial Intelligence (AAAI)
 British Machine Vision Conference (BMVC)
 Asian Conference on Computer Vision (ACCV)
 IEEE Winter Conference on Computer Vision (WACV)

Workshop Program Committee Member / Reviewer

Workshop on Large Scale Holistic Video Understanding, CVPR 2021
 Workshop on Learning from Unlabeled Videos, CVPR 2021
 Workshop on Self-Supervised Learning: Theory and Practice, NeurIPS 2020
 Workshop on Multi-Modal Video Analysis, ECCV 2020

Journal Reviewer:

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
 IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)
 IEEE Transactions on Image Processing (TIP)
 IEEE Transactions on Multimedia
 IEEE/ACM Transactions on Audio Speech and Language Processing
 Journal of Artificial Intelligence Research (JAIR)
 Computer Vision and Image Understanding (CVIU)
 Journal of the Acoustical Society of America (JASA)

SELECTED MEDIA COVERAGE

The University of Texas at Austin , Looking and Listening in Machine Learning .	11/2021
Facebook AI Blog , New milestones in embodied AI .	08/2020
MIT Technology Review , Deep learning turns mono recordings into immersive sound .	12/2018
Two Minute Papers , This AI produces binaural (2.5D) audio .	01/2019
Facebook AI Blog , Creating 2.5D visual sound for an immersive audio experience .	06/2019