

RUOHAN GAO

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

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

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*
- *Committee: Kristen Grauman, Andrew Zisserman, Raymond Mooney, Qixing Huang*
- *Michael H. Granof Award, University's Top 1 Doctoral Dissertation of 2021*
- *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, multisensory embodied learning, object-centric learning, multi-modal video understanding, learning from unlabeled videos, self-supervised learning

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

Robotics: multisensory robot learning

APPOINTMENTS

Meta Reality Labs, Redmond, WA

Research Scientist

July 2023 - Present

Stanford University, Stanford, CA

SAIL Postdoctoral Research Fellow

Feb. 2021 - June 2023

with Fei-Fei Li, Jiajun Wu, and Silvio Savarese at Stanford Vision and Learning Lab

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 Research (FAIR), Cambridge, MA

Research Intern with Lorenzo Torresani

June 2019 - Sept. 2019

Research Intern with Kristen Grauman

May 2018 - Aug. 2018

HONORS AND AWARDS

- Best Paper Award Runner Up, British Machine Vision Conference (BMVC), 2021
For the paper "Geometry-Aware Multi-Task Learning for Binaural Audio Generation from Video" with R. Garg and K. Grauman
- Stanford Artificial Intelligence Lab (SAIL) Postdoctoral Fellowship, 2021-2023

- University Nominee for ACM Doctoral Dissertation Award, UT Austin, 2021.
- Michael H. Granof Award, UT Austin, 2021
UT Austin's top graduate student award, awarded to top 1 dissertation in all areas from all graduate students 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, Conference on Computer Vision and Pattern Recognition (CVPR), 2020
- Graduate Dean's Prestigious Fellowship Supplement Fellow, UT Austin, 2019 & 2020
- Adobe Research Fellowship, 2019
- Best Paper Award Finalist, Conference on Computer Vision and Pattern Recognition (CVPR) 2019
For the paper, "2.5D Visual Sound", with K. Grauman
- Sir Run Run Shaw Postgraduate Scholarship, CUHK, 2015
- Dean's List, Engineering Faculty, CUHK, 2014 & 2015
- Cheung Wang Ngai Joseph GOAL Programme Scholarship, CUHK, 2014
- Q W Lee Scholarship, Top Academic Excellence Scholarship, CUHK, 2014
- National Scholarship, Ministry of Education of China, 2013

PEER-REVIEWED JOURNAL ARTICLES

(* indicates equal contribution; † indicates equal advising.)

1. Rishabh Garg, **Ruohan Gao**, Kristen Grauman, "Visually-Guided Audio Spatialization in Video with Geometry-Aware Multi-Task Learning", in *International Journal of Computer Vision (IJCV)*, May 2023. (**Invited article for best papers of BMVC 2021**)
2. Hong-Xing Yu*, Michelle Guo*, Alireza Fathi, Yen-Yu Chang, Eric Ryan Chan, **Ruohan Gao**, Thomas Funkhouser, Jiajun Wu, "Learning Object-Centric Neural Scattering Functions for Free-Viewpoint Relighting and Scene Composition", in *Transactions on Machine Learning Research (TMLR)*, April 2023.
3. Simon Le Cleac'h, Hong-Xing Yu, Michelle Guo, Taylor A. Howell, **Ruohan Gao**, Jiajun Wu, Zachary Manchester, Mac Schwager, "Differentiable Physics Simulation of Dynamics-Augmented Neural Objects", in *Robotics and Automation Letters (RA-L)*, March 2023

PEER-REVIEWED CONFERENCE PAPERS

(* indicates equal contribution; † indicates equal advising.)

1. Mason Wang*, Samuel Clarke*, Jui-Hsien Wang, **Ruohan Gao**, and Jiajun Wu, "SoundCam: A Dataset for Tasks in Tracking and Identifying Humans from Real Room Acoustics", in *Proceedings of the Conference on Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS)*, Dec. 2023.
2. Sharon Lee*, Ruohan Zhang*, Minjune Hwang*, Ayano Hiranaka*, Chen Wang, Wensi Ai, Jin Jie Ryan Tan, Shreya Gupta, Yilun Hao, Gabrael Levine, **Ruohan Gao**, Anthony Norcia, Jiajun Wu†, and Li Fei-Fei†, "NOIR: Neural Signal Operated Intelligent Robot for Everyday Activities", in *Proceedings of the Conference on Robot Learning (CoRL)*, Nov. 2023.
3. **Ruohan Gao***, Yiming Dou*, Hao Li*, Tanmay Agarwal, Jeannette Bohg, Yunzhu Li, Li Fei-Fei, Jiajun Wu, "The ObjectFolder Benchmark: Multisensory Object-Centric Learning with Neural and Real Objects", in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023.

4. Samuel Clarke, **Ruohan Gao**, Mason Wang, Mark Rau, Julia Xu, Jui-Hsien Wang, Doug James, Jiajun Wu, “RealImpact: A Dataset of Impact Sound Fields for Real Objects”, in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023. (**Highlight paper**)
5. **Ruohan Gao***, Hao Li*, Gokul Dharan, Zhuzhu Wang, Chengshu Li, Fei Xia, Silvio Savarese, Li Fei-Fei, Jiajun Wu, “Sonicverse: A Multisensory Simulation Platform for Embodied Household Agents that See and Hear”, in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, May 2023.
6. Trevor Scott Standley, **Ruohan Gao**, Dawn Chen, Jiajun Wu, Silvio Savarese, “An Extensible Multi-modal Multi-task Object Dataset with Materials”, in *Proceedings of the International Conference on Learning Representations (ICLR)*, May 2023.
7. Hao Li*, Yizhi Zhang*, Junzhe Zhu, Shaoxiong Wang, Michelle A Lee, Huazhe Xu, Edward Adelson, Li Fei-Fei, **Ruohan Gao†**, Jiajun Wu†, “See, Hear, and Feel: Smart Sensory Fusion for Robotic Manipulation”, in *Proceedings of the Conference on Robot Learning (CoRL)*, Dec. 2022.
8. **Ruohan Gao***, Zilin Si*, Yen-Yu Chang*, Samuel Clarke, Jeannette Bohg, Li Fei-Fei, Wenzhen Yuan, Jiajun Wu, “ObjectFolder 2.0: A Multisensory Object Dataset for Sim2Real Transfer”, in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022.
9. Changan Chen, **Ruohan Gao**, Paul Calamia, Kristen Grauman, “Visual Acoustic Matching”, to appear in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022. (**Oral presentation**, 4.2% acceptance rate)
10. **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.
11. 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)
12. 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**]
13. **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.
14. 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.
15. **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.
16. **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.
17. **Ruohan Gao** and Kristen Grauman, “Co-Separating Sounds of Visual Objects”, in *Proceedings of the International Conference on Computer Vision (ICCV)*, Oct. 2019.
18. **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**]

19. **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)
20. 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.
21. **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)
22. **Ruohan Gao** and Kristen Grauman, “On-demand Learning for Deep Image Restoration”, in *Proceedings of the International Conference on Computer Vision (ICCV)*, Oct. 2017.
23. **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.
24. **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.
25. **Ruohan Gao**, Pili Hu, and Wing Cheong Lau, “Property Preservation under Community-Based Sampling”, in *Proceedings of the Global Communications Conference (GLOBECOM)*, Dec. 2015.
26. **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

(* indicates equal contribution.)

1. Kyle Hsu, Tyler Ga Wei Lum, **Ruohan Gao**, Shixiang Shane Gu, Jiajun Wu, Chelsea Finn, “What Makes Certain Pre-Trained Visual Representations Better for Robotic Learning?” in *Deep Reinforcement Learning Workshop at the Conference on Neural Information Processing Systems (NeurIPS)*, Dec. 2022.
2. Kyle Hsu, Tyler Ga Wei Lum, **Ruohan Gao**, Shixiang Shane Gu, Jiajun Wu, Chelsea Finn, “What Makes Certain Pre-Trained Visual Representations Better for Robotic Learning?” in *Foundation Models for Decision Making Workshop at the Conference on Neural Information Processing Systems (NeurIPS)*, Dec. 2022.
3. **Ruohan Gao***, Zilin Si*, Yen-Yu Chang*, Samuel Clarke, Jeannette Bohg, Li Fei-Fei, Wenzhen Yuan, Jiajun Wu, “ObjectFolder 2.0: A Multisensory Object Dataset for Sim2Real Transfer”, in *Sound for Robots Workshop at the IEEE International Conference on Robotics and Automation (ICRA)*, May 2022.
4. **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.
5. **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.
6. **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.

INVITED AND CONFERENCE TALKS

- University of Southern California
Invited Talk, Los Angeles, CA April 2023
- Cornell University
Invited Talk, Ithaca, NY March 2023
- University of Maryland, College Park
Invited Talk, College Park, MD March 2023
- Georgia Institute of Technology
Invited Talk, Atlanta, GA Feb. 2023
- Gaoling School of Artificial Intelligence
Invited Talk, Renmin University of China, Virtual May 2022
- Intelligent Sensing Winter School
Invited Talk, Queen Mary University of London, Virtual Dec. 2021
- 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

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

Co-Instructor

Spring 2023

CS231N: Deep Learning for Computer Vision

- Co-instructed with Fei-Fei Li and Yunzhu Li
- Taught 400+ students with a course staff of 15 teaching assistants

Stanford University, Stanford, CA

Co-Instructor

Spring 2022

CS231N: Deep Learning for Computer Vision

- Co-instructed with Fei-Fei Li and Jiajun Wu
- Taught 400+ students with a course staff of 16 teaching assistants

Stanford University, Stanford, CA

Guest Lecture

Spring 2021

CS231N: Convolutional Neural Networks for Visual Recognition

- Multimodal Learning

The University of Texas at Austin, Austin, TX

Guest Lecture and Tutorials

Fall 2017

CS381V: Visual Recognition

- Introduction to Deep Learning
- Implementation of CNNs

The University of Texas at Austin, Austin, TX

Teaching Assistant

CS303E: Elements of Computers and Programming

Spring 2016

CS429: Computer Organization and Architecture

Fall 2015

PROFESSIONAL SERVICES**Area Chair**

IEEE International Conference on Computer Vision (ICCV), 2023

Senior Program Committee

AAAI Conference on Artificial Intelligence (AAAI), 2024

AAAI Conference on Artificial Intelligence (AAAI), 2023

Organizer

Co-Organizer, Workshop on AV4D: Visual Learning of Sounds in Spaces, ICCV 2023

Co-Organizer, Workshop on Sight and Sound, CVPR 2023

Co-Organizer, Workshop on Creative AI Across Modalities, AAAI 2023

Co-Organizer, Workshop on AV4D: Visual Learning of Sounds in Spaces, ECCV 2022

Co-Organizer, Workshop on Sight and Sound, CVPR 2022

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)
 Conference on Robot Learning (CoRL)
 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)
 International Symposium on Mixed and Augmented Reality (ISMAR)

Workshop Program Committee Member / Reviewer

Workshop on T4V: Transformers for Vision, CVPR 2022
 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
 IEEE Transactions on Robotics (T-RO)
 Journal of Artificial Intelligence Research (JAIR)
 Computer Vision and Image Understanding (CVIU)
 Journal of the Acoustical Society of America (JASA)
 ACM Computing Surveys

SELECTED MEDIA COVERAGE

TechXplore. , A multisensory simulation platform to train and test home robots.	June 2023
ENGADGET , Auditory AIs promise a more immersive AR/VR experience.	June 2022
Meta AI , Introducing AI-driven acoustic synthesis for AR and VR.	June 2022
The University of Texas at Austin , Looking and Listening in Machine Learning.	Nov. 2021
Facebook AI Blog , New milestones in embodied AI.	Aug. 2020
MIT Technology Review , Deep learning turns mono recordings into immersive sound.	Dec. 2018
Two Minute Papers , This AI produces binaural (2.5D) audio.	Jan. 2019
Facebook AI Blog , Creating 2.5D visual sound for an immersive audio experience.	June 2019