

Yong Jae Lee

University of Wisconsin–Madison
Department of Computer Sciences
1210 W Dayton St
Madison, WI 53706 USA

Phone: (608) 262-1804
Email: yongjaelee@cs.wisc.edu
<https://cs.wisc.edu/~yongjaelee/>

RESEARCH INTERESTS

My main research interests are in computer vision, machine learning, and robotics. In particular, I am interested in creating robust AI systems that can learn to understand our multimodal world with minimal human supervision.

APPOINTMENTS

Susan Beth Horwitz Professor University of Wisconsin–Madison, Department of Computer Sciences	August 2025–Present
Principal Research Scientist Adobe Research, San Jose	July 2025–Present
Associate Professor (with tenure) University of Wisconsin–Madison, Department of Computer Sciences	August 2021–July 2025
Cruise AI Visiting Faculty Cruise, San Francisco	July 2020–August 2021
Associate Professor (with tenure) Assistant Professor University of California, Davis, Department of Computer Science	July 2020–June 2021 July 2014–June 2020
Postdoctoral Fellow University of California, Berkeley, Department of EECS Carnegie Mellon University, Robotics Institute Supervised by Prof. Alexei A. Efros	Sept 2013–June 2014 Aug 2012–Aug 2013
Graduate Research Assistant University of Texas at Austin, Department of Computer Science Supervised by Prof. Kristen Grauman	June 2007–May 2012
Research Intern Microsoft Research, Interactive Visual Media Group Supervised by Dr. Larry Zitnick and Dr. Michael Cohen	June 2010–Sept 2010

EDUCATION

University of Texas at Austin, Austin, TX
Ph.D. in Electrical and Computer Engineering, May 2012
Thesis: “Visual Object Category Discovery in Images and Videos”
Supervised by Prof. Kristen Grauman

University of Texas at Austin, Austin, TX
M.S.E. in Electrical and Computer Engineering, August 2008
Thesis: “Foreground Focus: Finding Meaningful Features in Unlabeled Images”
Supervised by Prof. Kristen Grauman

University of Illinois at Urbana-Champaign, Champaign, IL
B.S. in Electrical Engineering (Honors), May 2006
Minor in Computer Science, Business

SELECTED HONORS AND AWARDS

- H. I. Romnes Faculty Fellowship, 2025
- Susan Beth Horwitz Professorship, 2025
- Best Paper Award, Workshop on Transformers for Vision, CVPR 2025
For the paper, “X-Fusion: Introducing New Modality to Frozen Large Language Models”, with Sicheng Mo, Thao Nguyen, Xun Huang, Siddharth Srinivasan Iyer, Yijun Li, Yuchen Liu, Abhishek Tandon, Eli Shechtman, Krishna Kumar Singh, Bolei Zhou, and Yuheng Li.
- Sony Focused Research Award, 2023
- SACM Student Choice Professor of the Year Award, 2022
Awarded to one professor in the CS department each year to recognize excellence in teaching, as voted on by the students.
- Adobe Data Science Research Award, 2022
- Best Paper Award, BMVC 2020
For the paper, “Delving Deeper into Anti-aliasing in ConvNets”, with Xueyan Zou, Fanyi Xiao, and Zhiding Yu.
- Sony Focused Research Award, 2020
- Most Innovative Award, COCO Object Detection Challenge, ICCV 2019
For the paper, “YOLACT: Real-time Instance Segmentation”, with Daniel Bolya, Chong Zhou, and Fanyi Xiao.
- UC Davis College of Engineering Outstanding Junior Faculty Award, 2019
Awarded to one Assistant Professor in the College of Engineering each year.
- Adobe Data Science Research Award, 2019
- AWS Machine Learning Research Award 2019
- National Science Foundation (NSF) CAREER Award, 2018
- AWS Machine Learning Research Award 2018
- Army Research Office (ARO) Young Investigator Program (YIP) Award, 2017

- UC Davis Hellman Foundation Fellowship, 2017
Awarded to Assistant Professors who exhibit potential for great distinction in their research. One of 15 awardees university-wide.
- Outstanding Reviewer Awards, ICCV 2013, CVPR 2015
- UT-Austin ECE dept. sole nominee for university-wide Outstanding Dissertation Award, 2012
- Doctoral Consortium Award, CVPR 2011

TEACHING EXPERIENCE

Faculty Instructor at UW–Madison:

- CS 639: Deep Learning for Computer Vision (Undergraduate course). Spring 2025. Enrollment: 114. Instructor rating: 6.19/7.0
- CS 839: Learning based Image Synthesis and Manipulation (Graduate course). Fall 2024. Enrollment: 45. Instructor rating: 6.67/7.0
- CS 639: Deep Learning for Computer Vision (Undergraduate course). Spring 2024. Enrollment: 104. Instructor rating: 6.03/7.0
- CS 839: Learning based Image Synthesis and Manipulation (Graduate course). Fall 2023. Enrollment: 33. Instructor rating: 6.18/7.0
- CS 639: Deep Learning for Computer Vision (Undergraduate course). Spring 2023. Enrollment: 102. Instructor rating: 5.70/7.0
- CS 839: Learning based Image Synthesis and Manipulation (Graduate course). Fall 2022. Enrollment: 42. Instructor rating: 6.38/7.0
- CS 839: Deep Learning for Visual Recognition (Graduate course). Spring 2022. Enrollment: 34. Instructor rating: 6.75/7.0

Faculty Instructor at UC Davis:

- ECS 174: Computer Vision (Undergraduate course). Spring 2020. Enrollment: 163. Instructor rating: 4.5/5.0
- ECS 269: Visual Recognition (Graduate course). Fall 2019. Enrollment: 54. Instructor rating: 4.5/5.0
- ECS 174: Computer Vision (Undergraduate course). Spring 2019. Enrollment: 171. Instructor rating: 4.4/5.0
- ECS 269: Visual Recognition (Graduate course). Fall 2018. Enrollment: 53. Instructor rating: 4.8/5.0
- ECS 174: Computer Vision (Undergraduate course). Spring 2018. Enrollment: 147. Instructor rating: 4.4/5.0
- ECS 289G: Visual Recognition (Graduate course). Winter 2018. Enrollment: 43. Instructor rating: 4.5/5.0
- ECS 174: Computer Vision (Undergraduate course). Spring 2017. Enrollment: 137. Instructor rating: 4.1/5.0
- ECS 289G: Visual Recognition (Graduate course). Fall 2016. Enrollment: 20. Instructor rating: 4.7/5.0

- ECS 289G: Visual Recognition (Graduate course). Fall 2015. Enrollment: 14. Instructor rating: 4.9/5.0
- ECS 189G: Intro to Computer Vision (Undergraduate course). Spring 2015. Enrollment: 141. Instructor rating: 4.2/5.0
- ECS 289H: Visual Recognition (Graduate course). Fall 2014. Enrollment: 15. Instructor rating: 4.9/5.0

Teaching Assistant at UT Austin:

- CS 376: Computer Vision (Undergraduate/Graduate course). Fall 2009, Spring 2011.

PUBLICATIONS

Books and Book Chapters

1. Fanyi Xiao and Yong Jae Lee. Localizing and Visualizing Relative Attributes. *Chapter in Visual Attributes*. R. Feris, C. Lampert, and D. Parikh, Editors. Springer, 2016.
2. Yong Jae Lee, Alexei A. Efros, and Martial Hebert. Discovering Mid-level Visual Connections in Space and Time. *Chapter in Visual Analysis and Geo-Localization of Large Scale Imagery*. A. Zamir, A. Hakeem, L. Van Gool, M. Shah, and R. Szeliski, Editors. Springer, 2016
3. Hyun Soo Park, Albert Ali Salah, Yong Jae Lee, Louis-Philippe Morency, Yaser Sheikh, Rita Cucchiara (Eds.). *Human Behavior Understanding*. Lecture Notes in Computer Science, Springer, 2014.

Journal Articles

1. Jaden Park, Valentin Deschaintre, Jason Kuen, Kangning Liu, Iliyan Georgiev, Krishna Kumar Singh, Yong Jae Lee, and Michael Fischer. MAOAM: Unified Object and Material Selection with Vision-Language Models. *ACM Transactions on Graphics* (Proceedings of SIGGRAPH), July 2026.
2. Kwanyong Park, Sojung An, Yong Jae Lee, and Donghyun Kim. Learning Compositionality from Multifaceted Synthetic Data for Language-based Object Detection. *International Journal of Computer Vision (IJCV)*, August 2025.
3. Bo Li, Haotian Liu, Liangyu Chen, Yong Jae Lee, Chunyuan Li, and Ziwei Liu. Benchmarking and Analyzing Generative Data for Visual Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (TPAMI), September 2025.
4. Zhuoran Yu, Chenchen Zhu, Sean Culatana, Raghuraman Krishnamoorthi, Fanyi Xiao*, and Yong Jae Lee*. Diversify, Don't Fine-Tune: Scaling Up Visual Recognition Training with Synthetic Images. *Transactions on Machine Learning Research (TMLR)*, 2025. (*equal advising)
5. Rafael E. P. Ferreira, Yong Jae Lee, and Joao Reboucas Dorea. Using Pseudo-Labeling to Improve Performance of Deep Neural Networks for Animal Identification. *Nature Scientific Reports*, 2023.
6. Xueyan Zou, Fanyi Xiao, Zhiding Yu, Yuheng Li, and Yong Jae Lee. Delving Deeper into Anti-aliasing in ConvNets. *International Journal of Computer Vision (IJCV)*, 2022. [[Invited article for best papers of BMVC 2020](#)]
7. Daniel Bolya*, Chong Zhou*, Fanyi Xiao, and Yong Jae Lee. YOLACT++: Better Real-time Instance Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (TPAMI), August 2020. (*equal contribution)

8. Farzad Fereidouni, Austin Todd, Yuheng Li, Che-Wei Chang, Keith Luong, Avi Rosenberg, Yong Jae Lee, James W. Chan, Alexander Borowsky, Karen Matsukuma, Kuang-Yu Jen, and Richard Levenson. Dual-mode Emission and Transmission Microscopy for Virtual Histochemistry using Hematoxylin- and Eosin-Stained Tissue Sections. *Biomedical Optics Express*, Vol. 10, No. 12, November 2019.
9. M. Rezaur Rahman, Jinyoung Han, Yong Jae Lee, and Chen-Nee Chuah. Analyzing the Adoption and Cascading Process of OSN-Based Gifting Applications: An Empirical Study. *ACM Transactions on the Web (TWEB)*, Vol. 11, No. 2, April 2017.
10. Yong Jae Lee and Kristen Grauman. Predicting Important Objects for Egocentric Video Summarization. *International Journal of Computer Vision (IJCV)*, Vol. 114, No. 1, pp. 38-55, January 2015.
11. Jun-Yan Zhu, Yong Jae Lee, and Alexei A. Efros. AverageExplorer: Interactive Exploration and Alignment of Visual Data Collections. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, Vol. 33, No. 4, July 2014.
12. Yong Jae Lee and Kristen Grauman. Object-Graphs for Context-Aware Visual Category Discovery. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 34, No. 2, pp. 346-358, February 2012.
13. Yong Jae Lee, C. Lawrence Zitnick, and Michael Cohen. ShadowDraw: Real-Time User Guidance for Freehand Drawing. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, Vol. 30, No. 4, July 2011.
14. Yong Jae Lee and Kristen Grauman. Foreground Focus: Unsupervised Learning from Partially Matching Images. *International Journal of Computer Vision (IJCV)*, Vol. 85, No. 2, pp. 143-166, May 2009.

Peer Reviewed Conference Papers (acceptance rates typically ~ 2%-25%)

1. Zhuoran Yu, Le Thien Phuc Nguyen, Jaden Park, Xinyi Gu, Zexue He, Soochahn Lee, Rogerio Feris, and Yong Jae Lee. DocHop: Benchmarking Out-of-domain Multi-hop Reasoning in Information-Dense Documents. In *Proceedings of the International Conference on Machine Learning (ICML)*, Seoul, South Korea, July 2026.
2. Thomas Liang*, Zhuoran Yu*, and Yong Jae Lee. Large Language Model Teaches Visual Students: Cross-Modality Transfer of Fine-Grained Conceptual Knowledge. In *Proceedings of the International Conference on Machine Learning (ICML)*, Seoul, South Korea, July 2026. (*equal contribution)
3. Aniket Rege, Arka Sadhu, Yuliang Li, Kejie Li, Ramya Korlakai Vinayak, Yuning Chai, Yong Jae Lee, and Hyo Jin Kim. Agentic Very Long Video Understanding. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, San Diego, CA, July 2026.
4. Zeyi Huang, Yuyang Ji, Anirudh Sundara Rajan, Zefan Cai, Wen Xiao, Haohan Wang, Junjie Hu, and Yong Jae Lee. VisualToolAgent (VisTA): A Reinforcement Learning Framework for Visual Tool Selection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Denver, CO, June 2026.
5. Thao Nguyen, Sicheng Mo, Krishna Kumar Singh, Yilin Wang, Jing Shi, Nicholas Kolkin, Eli Shechtman, Yong Jae Lee*, and Yuheng Li*. Relational Visual Similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Denver, CO, June 2026. (*equal advising)

6. Junsung Lee, Hyunsoo Lee, Yong Jae Lee, and Bohyung Han. Low-Resolution Editing is All You Need for High-Resolution Editing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Denver, CO, June 2026.
7. Sicheng Mo, Thao Nguyen, Richard Zhang, Nicholas Kolkin, Siddharth Srinivasan Iyer, Eli Shechtman, Krishna Kumar Singh, Yong Jae Lee, Bolei Zhou, and Yuheng Li. Group Diffusion: Enhancing Image Generation by Unlocking Cross-Sample Collaboration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Denver, CO, June 2026.
8. Le Thien Phuc Nguyen*, Zhuoran Yu*, Samuel Low Yu Hang, Subin An, Jeongik Lee, Yohan Ban, SeungEun Chung, Thanh-Huy Nguyen, JuWan Maeng, Soochahn Lee, and Yong Jae Lee. See, Hear, and Understand: Benchmarking Audiovisual Human Speech Understanding in Multimodal Large Language Models. In *Findings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR Findings)*, Denver, CO, June 2026. (*equal contribution)
9. Jaden Park, Mu Cai, Feng Yao, Jingbo Shang, Soochahn Lee, and Yong Jae Lee. Contamination Detection for VLMs using Multi-Modal Semantic Perturbation. In *International Conference on Learning Representations (ICLR)*, Rio de Janeiro, Brazil, April 2026.
10. Zhuoran Yu and Yong Jae Lee. How Multimodal LLMs Solve Image Tasks: A Lens on Visual Grounding, Task Reasoning, and Answer Decoding. In *Conference on Language Modeling (COLM)*, Montreal, CA, October 2025.
11. Sicheng Mo, Thao Nguyen, Xun Huang, Siddharth Srinivasan Iyer, Yijun Li, Yuchen Liu, Abhishek Tandon, Eli Shechtman, Krishna Kumar Singh, Yong Jae Lee, Bolei Zhou, and Yuheng Li. X-Fusion: Introducing New Modality to Frozen Large Language Models. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Honolulu, Hawaii, October 2025.
Best Paper at CVPR 2025 Transformers for Vision (T4V) Workshop
12. Aniket Rege, Zinnia Nie, Mahesh Ramesh, Unmesh Raskar, Zhuoran Yu, Aditya Kusupati*, Yong Jae Lee*, and Ramya Korlakai Vinayak*. CuRe: Cultural Gaps in the Long Tail of Text-to-Image Models. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Honolulu, Hawaii, October 2025. (*equal advising)
13. Yuyang Ji, Zeyi Huang, Haohan Wang, and Yong Jae Lee. Customizing Domain Adapters for Domain Generalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Honolulu, Hawaii, October 2025.
14. Yuzhang Shang*, Mu Cai*, Bingxin Xu, Yong Jae Lee‡, and Yan Yan‡. LLaVA-PruMerge: Adaptive Token Reduction for Efficient Large Multimodal Models. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Honolulu, Hawaii, October 2025. (*equal contribution, ‡equal advising)
15. Anirudh Sundara Rajan and Yong Jae Lee. Stay-Positive: A Case for Ignoring Real Image Features in Fake Image Detection. In *Proceedings of the International Conference on Machine Learning (ICML)*, Vancouver, CA, July 2025.
16. Thao Nguyen, Krishna Kumar Singh, Jing Shi, Trung Bui, Yong Jae Lee*, and Yuheng Li*. Yo'Chameleon: Personalized Vision and Language Generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, June 2025. (*equal advising)
17. Zeyi Huang*, Yuyang Ji*, Xiaofang Wang, Nikhil Mehta, Tong Xiao, Donghyun Lee, Sigmund VanValkenburgh, Shengxin Zha, Bolin Lai, Licheng Yu, Ning Zhang, Yong Jae Lee‡, and Miao Liu‡. Building a Mind Palace: Structuring Environment-Grounded Semantic Graphs or Effective Long Video Analysis with LLMs. In *Proceedings of the IEEE Conference on Computer*

- Vision and Pattern Recognition (CVPR)*, Nashville, TN, June 2025. (*equal contribution, ‡equal advising)
18. Yichen Xie, Hongge Chen, Gregory P. Meyer, Yong Jae Lee, Eric Wolff, Masayoshi Tomizuka, Wei Zhan, Yuning Chai, and Xin Huang. Cohere3D: Exploiting Temporal Coherence for Unsupervised Representation Learning of Vision-based Autonomous Driving. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Atlanta, GA, June 2025.
 19. Mu Cai, Jianwei Yang, Jianfeng Gao, and Yong Jae Lee. Matryoshka Multimodal Models. In *International Conference on Learning Representations (ICLR)*, Singapore, April 2025.
 20. Anirudh Sundara Rajan, Utkarsh Ojha, Jedidiah Schloesser, and Yong Jae Lee. On the Effectiveness of Dataset Alignment for Fake Image Detection. In *International Conference on Learning Representations (ICLR)*, Singapore, April 2025.
 21. Xiang Li, Cristina Mata, Jongwoo Park, Kumara Kahatapitiya, Yoo Sung Jang, Jinghuan Shang, Kanchana Ranasinghe, Ryan Burgert, Mu Cai, Yong Jae Lee, and Michael S. Ryoo. LLaRA: Supercharging Robot Learning Data for Vision-Language Policy. In *International Conference on Learning Representations (ICLR)*, Singapore, April 2025.
 22. Mu Cai*, Zeyi Huang*, Yuheng Li, Haohan Wang, and Yong Jae Lee. Leveraging Large Language Models for Scalable Vector Graphics-Driven Image Understanding. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, Tucson, AZ, March 2025. (*equal contribution)
 23. Thao Nguyen, Haotian Liu, Mu Cai, Yuheng Li, Utkarsh Ojha, and Yong Jae Lee. Yo’LLaVA: Your Personalized Language and Vision Assistant. In *Neural Information Processing Systems (NeurIPS)*, Vancouver, CA, December 2024.
 24. Xueyan Zou, Linjie Li, Jianfeng Wang, Jianwei Yang, Mingyu Ding, Junyi Wei, Zhengyuan Yang, Feng Li, Hao Zhang, Shilong Liu, Arul Aravinthan, Yong Jae Lee*, and Lijuan Wang*. What Can Foundation Models’ Embeddings Do? In *Neural Information Processing Systems (NeurIPS)*, Vancouver, CA, December 2024. (*equal advising)
 25. Bocheng Zou*, Mu Cai*, Jianrui Zhang, and Yong Jae Lee. VGBench: Evaluating Large Language Models on Vector Graphics Understanding and Generation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Miami, FL, November 2024. (*equal contribution)
 26. Young Kyun Jang, Junmo Kang, Yong Jae Lee, and Donghyun Kim. MATE: Meet At The Embedding – Connecting Images with Long Texts. In *Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP Findings)*, Miami, FL, November 2024.
 27. Yuheng Li, Haotian Liu, Mu Cai, Yijun Li, Eli Shechtman, Zhe Lin, Yong Jae Lee, and Krishna Kumar Singh. Removing Distributional Discrepancies in Captions Improves Image-Text Alignment. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Milan, Italy, October 2024.
 28. Jianrui Zhang*, Mu Cai*, Tengyang Xie, and Yong Jae Lee. CounterCurate: Enhancing Physical and Semantic Visio-Linguistic Compositional Reasoning via Counterfactual Examples. In *Findings of the Association for Computational Linguistics (ACL Findings)*, Bangkok, Thailand, August 2024. (*equal contribution)
 29. Mu Cai, Chenxu Luo, Yong Jae Lee, and Xiaodong Yang. Cross-Modal Self-Supervised Learning with Effective Contrastive Units for Point Clouds. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Abu Dhabi, UAE, October 2024.

30. Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved Baselines with Visual Instruction Tuning (LLaVA-1.5). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 2024. ([highlight](#), 2.8% acceptance rate)
31. Mu Cai, Haotian Liu, Siva Karthik Mustikovela, Gregory P. Meyer, Yuning Chai, Dennis Park, and Yong Jae Lee. ViP-LLaVA: Making Large Multimodal Models Understand Arbitrary Visual Prompts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 2024.
32. Thao Nguyen, Utkarsh Ojha, Yuheng Li, Haotian Liu, and Yong Jae Lee. Edit One for All: Interactive Batch Image Editing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 2024.
33. Moniek Smink, Haotian Liu, Doerte Doepfer, and Yong Jae Lee. Computer Vision on the Edge: Individual Cattle Identification in Real-time with ReadMyCow System. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, January 2024.
34. Yuexiang Zhai, Shengbang Tong, Xiao Li, Mu Cai, Qing Qu, Yong Jae Lee, and Yi Ma. Investigating the Catastrophic Forgetting in Multimodal Large Language Model Fine-Tuning. In *Conference on Parsimony and Learning (CPAL)*, Hong Kong, January 2024.
35. Kazuki Shin, Sankalp Yamsani, Roman Mineyev, Hongyu Chen, Nitish Gandhi, Yong Jae Lee, and Joohyung Kim. Exploring the Capabilities of a General-Purpose Robotic Arm in Chess Gameplay. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Austin, TX, December 2023.
36. Haotian Liu*, Chunyuan Li*, Qingyang Wu, and Yong Jae Lee. Visual Instruction Tuning (LLaVA). In *Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, December 2023. (*equal contribution) ([oral presentation](#), 0.5% acceptance rate)
37. Utkarsh Ojha*, Yuheng Li*, Anirudh Sundara Rajan*, Yingyu Liang, and Yong Jae Lee. Dissecting Knowledge Distillation: An Exploration of its Inner Workings and Applications. In *Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, December 2023. (*equal contribution)
38. Thao Nguyen, Yuheng Li, Utkarsh Ojha, and Yong Jae Lee. Visual Instruction Inversion: Image Editing via Image Prompting. In *Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, December 2023.
39. Xueyan Zou, Jianwei Yang, Hao Zhang, Feng Li, Linjie Li, Jianfeng Wang, Lijuan Wang, Jianfeng Gao*, and Yong Jae Lee*. Segment Everything Everywhere All at Once. In *Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, December 2023. (*equal advising)
40. Zeyi Huang, Andy Zhou, Zijian Ling, Mu Cai, Haohan Wang, and Yong Jae Lee. A Sentence Speaks a Thousand Images: Domain Generalization through Distilling CLIP with Language Guidance. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Paris, France, October 2023.
41. Yuheng Li, Haotian Liu, Qingyang Wu, Fangzhou Mu, Jianwei Yang, Jianfeng Gao, Chunyuan Li*, and Yong Jae Lee*. GLIGEN: Open-Set Grounded Text-to-Image Generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, CA, June 2023. (*equal advising)
42. Xueyan Zou*, Zi-Yi Dou*, Jianwei Yang*, Zhe Gan, Linjie Li, Chunyuan Li, Xiyang Dai, Jianfeng Wang, Lu Yuan, Nanyun Peng, Lijuan Wang, Harkirat Behl, Yong Jae Lee‡, and Jianfeng Gao‡. Generalized Decoding for Pixel, Image, and Language. In *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, CA, June 2023. (*equal contribution, ‡equal advising)

43. Utkarsh Ojha*, Yuheng Li*, and Yong Jae Lee. Towards Universal Fake Image Detectors that Generalize Across Generative Models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, CA, June 2023. (*equal contribution)
44. Haotian Liu, Kilho Son, Jianwei Yang, Ce Liu, Jianfeng Gao, Yong Jae Lee*, and Chunyuan Li*. REACT: Learning Customized Visual Models with Retrieval-Augmented Knowledge. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, CA, June 2023. (*equal advising) ([highlight](#), 2.5% acceptance rate)
45. Zhuoran Yu, Yin Li, and Yong Jae Lee. InPL: Pseudo-labeling the Inliers First for Imbalanced Semi-supervised Learning. In *International Conference on Learning Representations (ICLR)*, May 2023.
46. Chunyuan Li*, Haotian Liu*, Liunian Harold Li, Pengchuan Zhang, Jyoti Aneja, Jianwei Yang, Ping Jin, Houdong Hu, Zicheng Liu, Yong Jae Lee, and Jianfeng Gao. ELEVATER: A Benchmark and Toolkit for Evaluating Language-Augmented Visual Models. In *Neural Information Processing Systems (NeurIPS)*, Datasets and Benchmarks Track, New Orleans, LA, December 2022. (*equal contribution)
47. Haotian Liu, Mu Cai, and Yong Jae Lee. Masked Discrimination for Self-Supervised Learning on Point Clouds. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Tel Aviv, Israel, October 2022.
48. Yuheng Li, Yijun Li, Jingwan Lu, Eli Shechtman, Yong Jae Lee, and Krishna Kumar Singh. Contrastive Learning for Diverse Disentangled Foreground Generation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Tel Aviv, Israel, October 2022.
49. Haohan Wang, Zeyi Huang, Hanlin Zhang, Yong Jae Lee, and Eric Xing. Toward Learning Human-aligned Cross-domain Robust Models by Countering Misaligned Features. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, Eindhoven, Netherlands, August 2022.
50. Yang Xue, Yuheng Li, Krishna Kumar Singh, and Yong Jae Lee. GIRAFFE HD: A High-Resolution 3D-aware Generative Model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, June 2022.
51. Zeyi Huang*, Haohan Wang*, Dong Huang, Yong Jae Lee[†], and Eric Xing[†]. The Two Dimensions of Worst-case Training and the Integrated Effect for Out-of-domain Generalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, June 2022. (*equal contribution, [†]equal advising)
52. Maheen Rashid, Sofia Broome, Katrina Ask, Elin Hernlund, Pia Haubro Andersen, Hedvig Kjellstrom, and Yong Jae Lee. Equine Pain Behaviour Classification via Self-supervised Disentangled Pose Representation. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, January 2022.
53. Yuheng Li, Krishna Kumar Singh, Yang Xue, and Yong Jae Lee. PartGAN: Weakly-supervised Part Decomposition for Image Generation and Segmentation. In *Proceedings of the British Machine Vision Conference (BMVC)* 2021.
54. Yuheng Li, Yijun Li, Jingwan Lu, Eli Shechtman, Yong Jae Lee, and Krishna Kumar Singh. Generate Globally, Enhance Locally: High Resolution Conditional Image Generation with Class-specific Models. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, October 2021.

55. Haotian Liu*, Rafael A. Rivera-Soto*, Fanyi Xiao, and Yong Jae Lee. YolactEdge: Real-time Instance Segmentation on the Edge. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, June 2021. (*equal contribution)
56. Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A. Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot Image Generation via Cross-domain Correspondence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021.
57. Xueyan Zou, Linjie Yang, Ding Liu, and Yong Jae Lee. Progressive Temporal Feature Alignment Network for Video Inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021.
58. Utkarsh Ojha, Krishna Kumar Singh, and Yong Jae Lee. Generating Furry Cars: Disentangling Object Shape and Appearance across Multiple Domains. In *International Conference on Learning Representations (ICLR)*, May 2021.
59. Rajat Arora and Yong Jae Lee. SinGAN-GIF: Learning a Generative Video Model from a Single GIF. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, January 2021.
60. Utkarsh Ojha, Krishna Kumar Singh, Cho-Jui Hsieh, and Yong Jae Lee. Elastic-InfoGAN: Generative Modeling of Disentangled Representations in Class-Imbalanced Data. In *Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, December 2020.
61. Xueyan Zou, Fanyi Xiao, Zhiding Yu, and Yong Jae Lee. Delving Deeper into Anti-aliasing in ConvNets. In *Proceedings of the British Machine Vision Conference (BMVC)*, Manchester, United Kingdom, September 2020. ([oral presentation](#), 5.1% acceptance rate)
Best Paper Award (1 of 670 submissions)
62. Xiuye Gu, Weixin Luo, Michael Ryoo, and Yong Jae Lee. Password-conditioned Anonymization and Deanonimization with Face Identity Transformers. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Glasgow, United Kingdom, August 2020.
63. Zainul Abi Din, Hari Venugopalan, Jaime Park, Andy Li, Weisu Yin, Haohui Mai, Yong Jae Lee, Steven Liu, Samuel T. King. Boxer: Preventing Fraud by Scanning Credit Cards. In *Proceedings of the USENIX Security Symposium (USENIX Security)*, August 2020.
64. Yuheng Li, Krishna Kumar Singh, Utkarsh Ojha, and Yong Jae Lee. MixNMatch: Multifactor Disentanglement and Encoding for Conditional Image Generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 2020.
65. Krishna Kumar Singh, Dhruv Mahajan, Kristen Grauman, Yong Jae Lee, Matt Feiszli, and Deepti Ghadiyaram. Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 2020. ([oral presentation](#), 5.7% acceptance rate)
66. Zhongzheng Ren, Zhiding Yu, Xiaodong Yang, Ming-Yu Liu, Yong Jae Lee, Alexander Schwing, and Jan Kautz. Instance-aware, Context-focused, and Memory-efficient Weakly-supervised Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 2020.
67. Maheen Rashid, Hedvig Kjellstrom, and Yong Jae Lee. Action Graphs: Weakly-supervised Action Localization with Graph Convolution Networks. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Aspen, CO, March 2020.
68. Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. YOLACT: Real-time Instance Segmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea, November 2019. ([oral presentation](#), 4.3% acceptance rate)
Most Innovative Award, COCO Object Detection Challenge, ICCV 2019

69. Fanyi Xiao, Haotian Liu, and Yong Jae Lee. Identity from here, Pose from there: Self-supervised Disentanglement and Generation of Objects using Unlabeled Videos. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea, November 2019.
70. Krishna Kumar Singh*, Utkarsh Ojha*, and Yong Jae Lee. FineGAN: Unsupervised Hierarchical Disentanglement for Fine-Grained Object Generation and Discovery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, June 2019. (*equal contribution) (oral presentation, 5.5% acceptance rate)
71. Krishna Kumar Singh and Yong Jae Lee. You reap what you sow: Using Videos to Generate High Precision Object Proposals for Weakly-supervised Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, June 2019.
72. Xiuye Gu, Yijie Wang, Chongruo Wu, Yong Jae Lee, and Panqu Wang. HPLFlowNet: Hierarchical Permutohedral Lattice FlowNet for Scene Flow Estimation on Large-scale Point Clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, June 2019.
73. Krishna Kumar Singh, Santosh Divvala, Ali Farhadi, and Yong Jae Lee. DOCK: Detecting Objects by transferring Common-sense Knowledge. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, September 2018.
74. Fanyi Xiao and Yong Jae Lee. Video Object Detection with an Aligned Spatial-Temporal Memory. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, September 2018.
75. Zhongzheng Ren, Yong Jae Lee, and Michael Ryoo. Learning to Anonymize Faces for Privacy Preserving Action Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, September 2018.
76. Mingyang Zhou, Runxiang Cheng, Yong Jae Lee, and Zhou Yu. A Visual Attention Grounding Neural Model for Multimodal Machine Translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium, November 2018.
77. Zhongzheng Ren and Yong Jae Lee. Cross-Domain Self-supervised Multi-task Feature Learning using Synthetic Imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, June 2018.
78. Wenjian Hu, Krishna Kumar Singh*, Fanyi Xiao*, Jinyoung Han, Chen-Nee Chuah, and Yong Jae Lee. Who Will Share My Image? Predicting the Content Diffusion Path in Online Social Networks. In *ACM International Conference on Web Search and Data Mining (WSDM)*, Los Angeles, CA, February 2018. (*equal contribution)
79. Maheen Rashid, Sofia Broome, Pia Andersen, Karina Glerup, and Yong Jae Lee. What should I annotate? An automatic tool for finding video segments for EquiFACS annotation. In *International Conference on Methods and Techniques in Behavioral Research (Measuring Behavior)*, Manchester, UK, June 2018.
80. Pia Andersen, Karina Glerup, Jennifer Wathan, Britt Coles, Hedvig Kjellstrom, Sofia Broome, Yong Jae Lee, Maheen Rashid, Claudia Sonder, Erika Rosenberger, and Deborah Forster. Can a Machine Learn to See Horse Pain? An Interdisciplinary Approach Towards Automated Decoding of Facial Expressions of Pain in the Horse. In *International Conference on Methods and Techniques in Behavioral Research (Measuring Behavior)*, Manchester, UK, June 2018.

81. Krishna Kumar Singh and Yong Jae Lee. Hide-and-Seek: Forcing a Network to be Meticulous for Weakly-supervised Object and Action Localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 2017.
82. Fanyi Xiao, Leonid Sigal, and Yong Jae Lee. Weakly-supervised Visual Grounding of Phrases with Linguistic Structures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, June 2017.
83. Maheen Rashid, Xiuye Gu, and Yong Jae Lee. Interspecies Knowledge Transfer for Facial Keypoint Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, June 2017.
84. Chenyou Fan, Jangwon Lee, Mingze Xu, Krishna Kumar Singh, Yong Jae Lee, David Crandall and Michael Ryoo. Identifying First-Person Camera Wearers in Third-Person Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, June 2017.
85. Zhongzheng Ren, Adriana Noronha, Annie Vogel Ciernia, and Yong Jae Lee. Who Moved My Cheese? Automatic Annotation of Rodent Behaviors with Convolutional Neural Networks. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Santa Rosa, CA, March 2017.
86. Krishna Kumar Singh and Yong Jae Lee. End-to-End Localization and Ranking for Relative Attributes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Amsterdam, Netherlands, October 2016.
87. Fanyi Xiao and Yong Jae Lee. Track and Segment: An Iterative Unsupervised Approach for Video Object Proposals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, June 2016. ([spotlight presentation](#), 9.7% acceptance rate)
88. Krishna Kumar Singh, Fanyi Xiao, and Yong Jae Lee. Track and Transfer: Watching Videos to Simulate Strong Human Supervision for Weakly-Supervised Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, June 2016.
89. Fanyi Xiao and Yong Jae Lee. Discovering the Spatial Extent of Relative Attributes. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, December 2015. ([oral presentation](#), 3.3% acceptance rate)
90. Tinghui Zhou, Yong Jae Lee, Stella Yu, and Alexei A. Efros. FlowWeb: Joint Image Set Alignment by Weaving Consistent, Pixel-wise Correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, June 2015. ([oral presentation](#), 3.3% acceptance rate)
91. Hyun Oh Song, Yong Jae Lee, Stefanie Jegelka, and Trevor Darrell. Weakly-supervised Discovery of Visual Pattern Configurations. In *Neural Information Processing Systems (NIPS)*, Montreal, Canada, December 2014.
92. Yong Jae Lee, Alexei A. Efros, and Martial Hebert. Style-aware Mid-level Representation for Discovering Visual Connections in Space and Time. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Sydney, Australia, December 2013. ([oral presentation](#), 2.5% acceptance rate)
93. Yong Jae Lee, Joydeep Ghosh, and Kristen Grauman. Discovering Important People and Objects for Egocentric Video Summarization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, June 2012.

94. Yong Jae Lee, Jaechul Kim, and Kristen Grauman. Key-Segments for Video Object Segmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, November 2011.
95. Yong Jae Lee and Kristen Grauman. Face Discovery with Social Context. In *Proceedings of the British Machine Vision Conference (BMVC)*, Dundee, Scotland, August 2011.
96. Yong Jae Lee and Kristen Grauman. Learning the Easy Things First: Self-Paced Visual Category Discovery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, June 2011.
97. Yong Jae Lee and Kristen Grauman. Object-Graphs for Context-Aware Category Discovery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, June 2010. ([oral presentation](#), 4.5% acceptance rate)
98. Yong Jae Lee and Kristen Grauman. Collect-Cut: Segmentation with Top-Down Cues Discovered in Multi-Object Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, June 2010.
99. Yong Jae Lee and Kristen Grauman. Shape Discovery from Unlabeled Image Collections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, June 2009.
100. Yong Jae Lee and Kristen Grauman. Foreground Focus: Finding Meaningful Features in Unlabeled Images. In *Proceedings of the British Machine Vision Conference (BMVC)*, Leeds, United Kingdom, September 2008. ([oral presentation](#), 12.5% acceptance rate)
101. Changhai Xu, Yong Jae Lee, and Benjamin Kuipers. Ray-based Color Image Segmentation. In *Proceedings of the Canadian Conference on Computer and Robot Vision (CRV)*, Windsor, Ontario, May 2008.

INVITED TALKS

- Invited Talk, Embedded Vision Summit, Santa Clara, CA, May 2026.
“Multimodal AI Agents for Content Editing.”
- Invited Talk, ICCV Workshop on Closing the Loop Between Vision and Language (Decade Mark), Honolulu, HI, October 2025.
“Understanding and Advancing Multimodal Models.”
- Invited Talk, ICCV Workshop on What is Next in Multimodal Foundation Models?, Honolulu, HI, October 2025.
“Understanding and Advancing Multimodal Models.”
- Invited Talk, NetApp, September 2025.
“Toward Agentic Multimodal Models.”
- Invited Talk, Edge AI and Vision Innovation Forum, September 2025.
“Multimodal AI Agents.”
- Invited Talk, UC Berkeley, August 2025.
“Beyond Understanding: Toward Controllable and Agentic Multimodal Models.”
- Plenary Talk, Korean Artificial Intelligence Association Conference, August 2025.
“Beyond Understanding: Toward Controllable and Agentic Multimodal Models.”
- Invited Talk, Sogang University, June 2025.
“Toward Agentic Generalist Multimodal Models.”

- Keynote Talk, 3rd MSU-ND Computer Vision and Biometrics Workshop, April 2025.
“Generalist Multimodal Models.”
- Invited Talk, GenTech Seminar, Adobe, March 2025.
“Generalist Multimodal Models.”
- Invited Talk, AI Insight, LG, November 2024.
“Generalist Vision-Language Models.”
- Computer Science Colloquium, Washington University in St. Louis, October 2024.
“Generalist Vision-Language Models.”
- Invited Talk, ML seminar series, KLA, October 2024.
“Next Steps in Generalist Multimodal Models.”
- Invited Talk, Midwest Computer Vision Workshop, September 2024.
“Generalist Vision-Language Models.”
- Invited Talk, London Machine Learning Meetup, September 2024.
“Next Steps in Generalist Multimodal Models.”
- Distinguished AI Speaker Seminar, Salesforce AI, September 2024.
“Next Steps in Generalist Multimodal Models.”
- Keynote Talk, Korean Conference on Computer Vision, Busan, South Korea, August 2024.
“Learning to Understand Our Multimodal World.”
- Distinguished Speaker, National Yang Ming Chiao Tung University H&J Global Chair Colloquium, Hsinchu, Taiwan, August 2024.
“Learning to Understand Our Multimodal World.”
- Distinguished Speaker, National Yang Ming Chiao Tung University H&J Global Chair Colloquium, Hsinchu, Taiwan, August 2024.
“Generating Images and Detecting Them.”
- Invited Talk, Samsung, Seoul, South Korea, July 2024.
“Learning to Understand Our Multimodal World.”
- Invited Talk, Nota AI, Seoul, South Korea, July 2024.
“Next Steps in Generalist Multimodal Models.”
- Invited Talk, Korea University, Seoul, South Korea, July 2024.
“Next Steps in Generalist Multimodal Models.”
- Invited Talk, Seoul National University, Seoul, South Korea, July 2024.
“Next Steps in Generalist Multimodal Models.”
- Invited Talk, Meta Reality Labs, Redmond, WA, June 2024.
“Learning to Understand Our Multimodal World with Minimal Supervision.”
- Invited Talk, CVPR Workshop on Computer Vision in the Wild, Seattle, WA, June 2024.
“Building Steerable Generalist Multimodal Models.”
- Invited Talk, CVPR Workshop on What is Next in Multimodal Foundation Models?, Seattle, WA, June 2024.
“Learning to Understand Our Multimodal World.”
- Invited Talk, CVPR Workshop on Prompting in Vision, Seattle, WA, June 2024.
“Visual Prompting in Large Multimodal Models.”

- Invited Talk, CVPR Tutorial on Generalist Agent AI, Seattle, WA, June 2024.
“Generalist Multimodal Models.”
- Keynote Talk, Embedded Vision Summit, Santa Clara, CA, May 2024.
“Learning to Understand Our Multimodal World with Minimal Supervision.”
- AI Speaker Series, Upwork, May 2024.
“Learning to Understand Our Multimodal World with Minimal Supervision.”
- Computer Science Colloquium, University of Pittsburgh, April 2024.
“Learning to Understand Our Multimodal World with Minimal Supervision.”
- Microsoft Research, April 2024.
“Next Steps in Large Multimodal Models.”
- HP Labs, March 2024.
“Large Multimodal Vision-Language Models for Image Understanding.”
- Adobe, San Jose, CA, December 2023.
“Large Multimodal Vision-Language Models for Image Understanding.”
- Carnegie Mellon University, Robotics Institute VASC Seminar, October 2023.
“Large Multimodal Vision-Language Models for Image Generation and Understanding.”
- MLLM Talk, University of Wisconsin–Madison, October 2023.
“Large Multimodal Vision-Language Models for Image Generation and Understanding.”
- Yonsei University, LG Energy Solution, Krafton, MakinaRocks, August 2023.
“Large Multimodal Vision-Language Models for Image Generation and Understanding.”
- MLLM Talk, University of Wisconsin–Madison, February 2023.
“On Detecting Fake (AI-generated) Images.”
- Keynote Talk, Hyundai & Seoul National University AI Consortium, April 2022.
“Towards 3D Visual Understanding with Minimal Human Supervision.”
- SILO Talk, University of Wisconsin–Madison, March 2022.
“Towards 3D Visual Understanding with Minimal Human Supervision.”
- Keynote Talk, KAIST AI Graduate School, October 2021.
“Controllable Visual Content Generation without Expensive Supervision.”
- Keynote Talk, OpenMMLab Tutorial, CVPR 2021.
“Real-time Instance Segmentation with the YOLACT Family.”
- AI Summer School, Seoul National University, August 2020.
“Learning Disentangled Visual Representations with Minimal Human Supervision.”
- School of Interactive Computing Seminar, Georgia Tech, April 2020.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Computer Sciences Seminar, University of Wisconsin–Madison, April 2020.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Computer Science Colloquium, Purdue University, March 2020.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Computer Science Seminar, Brown University, March 2020.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Computer Vision Seminar, UC Berkeley, February 2020.
“Learning to Understand Visual Data with Minimal Human Supervision.”

- Cruise, San Francisco, CA, February 2020.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Google Research, Mountain View, CA, October 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Computer Science Seminar, University of British Columbia, September 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Computer Science Seminar, University of California, Irvine, September 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Snap Research, Santa Monica, CA, September 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- ECE Seminar, University of Southern California, September 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Seminar, LG Research, Seoul, Korea, July 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Computer Science Colloquium, University of Maryland, College Park, April 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- Seminar, NVIDIA Research, Santa Clara, CA, March 2019.
“Learning to Understand Visual Data with Minimal Human Supervision.”
- CVPR Area Chair Workshop, San Diego, February 2019.
“FineGAN: Unsupervised Hierarchical Disentanglement for Fine-Grained Object Generation and Discovery.”
- Center for Mind and Brain Seminar, University of California, Davis, February 2019.
“Learning to Localize, Generate, and Anonymize Objects with Indirect Supervision.”
- Carnegie Mellon University, Robotics Institute VASC Seminar, October 2018.
“Learning to Localize and Anonymize Objects with Indirect Supervision.”
- Data Science Institute Seminar, Lawrence Livermore National Laboratory, September 2018.
“Learning to Localize and Anonymize Objects with Indirect Supervision.”
- Computer Vision Seminar, KAIST University, Seoul, Korea, August 2018.
“Learning to Localize and Anonymize Objects with Indirect Supervision.”
- Korean Conference on Computer Vision, Seoul, Korea, July 2018.
“Cross-Domain Self-supervised Multi-task Feature Learning using Synthetic Imagery.”
- Computer Vision Seminar, UC Berkeley, July 2018.
“Learning to Localize and Anonymize Objects with Indirect Supervision.”
- Center for Vision, Cognition, Learning, and Autonomy Seminar, UCLA, February 2018.
“Learning to See with Weak Human Supervision.”
- Computational Neuroscience Seminar, University of California, Davis, November 2017.
“Computer Vision for Animal Behavior Recognition.”
- Seminar, Intel Labs, Santa Clara, CA, October 2017.
“Learning to See with Minimal Human Supervision.”
- Seoul National University, UNIST, Samsung, LG, Naver, Kakao, Korea, August 2017.
“Learning Where and How to Look for Weakly-Supervised Visual Recognition.”

- Keynote Talk, VALSE Seminar, China, April 2017.
“Learning Where and How to Look for Weakly-Supervised Visual Recognition.”
- Keynote Talk, 1st International Workshop on Human Activity Analysis with Highly Diverse Cameras, Santa Rosa, CA, March 2017.
“Hide-and-Seek: Forcing a Network to be Meticulous for Weakly-Supervised Object and Action Localization.”
- Computer Science Colloquium, Sonoma State University, September 2016.
“Attending to the Right Image Regions for Weakly-Supervised Visual Recognition.”
- Seminar, Lawrence Livermore National Laboratory, September 2016.
“Discovering the Relevant Image Regions for Weakly-Supervised Visual Recognition.”
- Machine Learning Seminar, University of California, Davis, December 2015.
“Watching Videos to Simulate Strong Human Supervision for Visual Recognition.”
- Korean Conference on Computer Vision, Seoul, Korea, August 2015.
“Interactive Exploration and Alignment of Visual Data Collections.”
- Computational Neuroscience Seminar, University of California, Davis, May 2015.
“Visual Data Mining in Image Collections.”
- Computer Science Departmental Colloquium, University of California, Davis, April 2014.
“Visual Data Mining in Image Collections.”
- Computer Science Departmental Seminar, University of Massachusetts, Amherst, March 2014.
“Visual Data Mining in Image Collections.”
- Creative Technologies Lab, Adobe Research, San Francisco, CA, February 2014.
“Visual Data Mining in Image Collections.”
- IEEE International Conf. on Computer Vision, Sydney, Australia, December 2013.
“Style-aware Mid-level Representation for Discovering Visual Connections in Space and Time.”
- IEEE CVPR Workshop on Egocentric Vision, Providence, RI, June 2012.
“Discovering Important People and Objects for Egocentric Video Summarization.”
- Carnegie Mellon University, Robotics Institute VASC Seminar, Pittsburgh, December 2011.
“Visual Category Discovery in Images and Videos.”
- Microsoft Research, MSRNE Vision Seminar, Boston, MA, December 2011.
“Visual Category Discovery in Images and Videos.”
- ACM SIGGRAPH conference, Vancouver, Canada, August 2011.
“ShadowDraw: Real-Time User Guidance for Freehand Drawing.”
- Microsoft Research, Vision Technology (VisTech) Seminar, Redmond, WA, June 2010.
“Collect-Cut: Segmentation with Top-Down Cues Discovered in Multi-Object Images.”
- IEEE Conf. on Computer Vision and Pattern Recognition, San Francisco, CA, June 2010.
“Object-Graphs for Context-Aware Category Discovery.”
- British Machine Vision Conference, Leeds, U.K., September 2008.
“Foreground Focus: Finding Meaningful Features in Unlabeled Images.”

INVITED PANELS

- ICCV Workshop on Closing the Loop Between Vision and Language (Decade Mark), Honolulu, HI, October 2025.
- CVPR Workshop on What is Next in Multimodal Foundation Models?, Nashville, TN, June 2025.
- Embedded Vision Summit, Santa Clara, CA, May 2024.
“Multimodal LLMs at the Edge: Are We There Yet?”
- CVPR Workshop on Prompting in Vision, Seattle, WA, June 2024.

PATENTS

- C. Lawrence Zitnick, Yong Jae Lee, and Michael Cohen. Realtime User Guidance for Freehand Drawing. US Patent US8827710B2. Granted Sep 2014.
- Yong Jae Lee and Haotian Liu. Method and System of using a Global Transformer for Efficient Modeling of Global Context in Point Clouds. US Patent US11908202B2. Granted Feb 2024.

LIVE DEMOS

- Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. YOLACT: Real-time Instance Segmentation. ICCV 2019.
- Zhongzheng Ren, Yong Jae Lee, Hyun Jong Yang, and Michael Ryoo. Activity-Preserving Face Anonymization for Privacy Protection. ECCV 2018.

PROFESSIONAL SERVICE ACTIVITY

- **Program Chair:** WACV 2023
- **Area Chair:** NeurIPS 2026, ACL 2026, ICLR 2026, CVPR 2026 (Senior AC), NeurIPS 2025, ICCV 2025 (Senior AC), ICLR 2025, CVPR 2025, NeurIPS 2024, ECCV 2024, ICLR 2024, CVPR 2024, NeurIPS 2023, ICCV 2023, CVPR 2023, AAAI 2023, ACCV 2023, CVPR 2021, ACCV 2020, ECCV 2020, ICCV 2019, CVPR 2019, ICVGIP 2016
- **Associate Editor:** IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020-2026
- **Associate Editor:** Journal of Artificial Intelligence Research (JAIR), 2019-2022
- **Associate Editor:** Journal of Visual Communication and Image Representation (JVCI), 2016-2021
- **Best Paper Awards Committee:** ICLR 2026, AAAI 2023
- **Senior Program Committee:** IJCAI 2016, AAAI 2020, IJCAI 2020
- **Conference Program Committee:**
 - CVPR 2011-2018, 2020, 2022 ([Outstanding Reviewer Award CVPR 2015](#))
 - ICCV 2009, 2011, 2013, 2015, 2017 ([Outstanding Reviewer Award ICCV 2013](#))
 - ECCV 2012, 2014, 2018, 2022, 2026
 - NeurIPS 2015, 2018, 2019, 2020
 - ICML 2018, 2019, 2026

- ACCV 2014
- WACV 2021, 2025
- BMVC 2021
- **Journal Reviewer:** International Journal of Computer Vision (IJCV), IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), ACM SIGGRAPH, ACM SIGGRAPH ASIA, ACM Multimedia, Pacific Graphics, Computer Graphics Forum, IEEE Transactions on Human-Machine Systems
- **Workshop Program Committee:** International Workshop on Attention/Intention Understanding (AIU), ACCV 2018; 1st International Workshop on Egocentric Perception, Interaction and Computing (EPIC), ECCV 2016; 4th Workshop on Web-scale Vision and Social Media (VSM), ECCV 2016; 4th International Workshop on Large Scale Visual Recognition and Retrieval (BigVision), CVPR 2016; 1st Workshop on Storytelling with Images and Videos, ECCV 2014; Workshop on Learning From Unlabeled Videos (LUV) CVPR 2020; NewInML NeurIPS 2021; Transformers for Vision (T4V) CVPR 2022
- **Panelist:** NSF 2025, 2024, 2022, 2018, 2017, 2016; DOE 2016
- **Chair / Co-Chair:** Demo & Exhibition, CVPR 2019
- **Organizer:** 4th Workshop on Computer Vision in the Wild (Challenge Track), CVPR 2025
- **Advisory Committee:** 1st Workshop on Computer Vision in the Wild, ECCV 2022
- **Organizer:** 4th Workshop on Egocentric (First-Person) Vision, CVPR 2016
- **Organizer:** 3rd Workshop on Egocentric (First-Person) Vision, CVPR 2014
- **Organizer:** 5th Workshop on Human Behavior Understanding, ECCV 2014
- **Faculty Advisor:** AggieMentors Outreach Program, UC Davis, 2018-2020
- **Mentor:** AggieMentors Science Fair Mentor, UC Davis, 2017-2019
- **Mentor:** Doctoral Consortium Mentor, ICCV 2021, CVPR 2024
- **Judge:** AggieMentors Science Fair Judge, UC Davis, 2016; RoboPlay C-STEM Video Judge, UC Davis, 2016-2020
- **Organizer:** UT-Austin Vision and Learning Reading Group, 2008-2012
- **UC Davis Committees:**
 - Research and Library Committee (2014-2020)
 - Graduate Admissions Committee (2014-2016, 2017-2018)
 - Faculty Recruitment Committee (NLP 2016-2017, Comp. Neuroscience 2017-2018, ML 2019-2020)
 - Faculty Representative Committee (2018-2020)
 - Data Science Committee (2019-2020)
- **UW–Madison Committees:**
 - Professional Programs Committee (2021-Present)
 - Office of the Vice Chancellor for Research and Graduate Education Committee (2022-2025)
 - Department Vision Committee (2023-2024)
 - Faculty Recruitment Committee (NLP 2024-2025, VetMed AI 2024-2025)

ADVISING ACTIVITY

- Fanyi Xiao (PhD student), 2015-2020
 - Thesis: “Multimodal Learning with Minimal Human Supervision from Videos and Natural Language”
 - [UC Davis Best Graduate Researcher in Computer Science Award 2018](#)
 - Next position: Research Scientist at Amazon Research
- Krishna Kumar Singh (PhD student), 2015-2020
 - Thesis: “Understanding and Generating Visual Data with Minimal Human Supervision”
 - [UC Davis Best Graduate Researcher in Computer Science Award \(Honorable Mention\) 2019](#)
 - Next position: Research Scientist at Adobe Research
- Maheen Rashid (PhD student), 2015-2021
 - Thesis: “Towards Automatic Visual Recognition of Horse Pain”
 - Next position: Computer Vision Scientist at Univrse AB
- Xueyan Zou (PhD student), 2018-2024
 - Thesis: “Towards Generalized Multimodal Foundation Models”
 - Next position: Postdoc at UCSD
- Utkarsh Ojha (PhD student), 2019-2024
 - Thesis: “Understanding Generative Models and Detecting their Images”
 - Next position: Postdoc at Carnegie Mellon University
- Haotian Liu (PhD student), 2019-2024
 - Thesis: “Steerable Visual Intelligence”
 - Next position: Member of Technical Staff at xAI
- Yuheng Li (PhD student), 2019-2024
 - Thesis: “Controllable Image Generative Models”
 - Next position: Research Scientist at Adobe Research
- Mu Cai (PhD student), 2021-2025
 - Thesis: “Toward Versatile and Efficient Multimodal Models”
 - Next position: Research Scientist at Google DeepMind
- Zhuoran Yu (PhD student), 2021-2026
 - Thesis: “Data Centric Visual Intelligence: from Unimodal to Multimodal”
 - Next position: Research Scientist at Autodesk
- Zeyi Huang (PhD student), 2022-Present
- Thao Nguyen (PhD student), 2022-Present
- Anirudh Sundara Rajan (PhD student), 2023-Present
- Jaden Park (PhD student), 2024-Present

- Jianrui (Harris) Zhang (PhD student), 2024-Present
- Aniket Rege (PhD student, co-advised with Ramya Vinayak), 2024-Present
- Zhongzheng (Jason) Ren (MS student), 2016-2018
 - Next position: PhD student at UIUC
- Wenjian Hu (MS student / Physics PhD), 2017-2018
 - Next position: Research Scientist at Facebook
- Leonardo Ferrer (MS student), 2017-2018
 - Next position: Software Engineer at Google
- Chong Zhou (MS student), 2019-2020
 - Next position: PhD student at Nanyang Technological University
- Yuheng Li (MS student), 2019-2020
 - Next position: PhD student at UC Davis
- Wei-Pang (Tyler) Jan (MS student), 2019-2020
 - Next position: Software Engineer at Amazon
- Rafael A. Rivera-Soto (MS student), 2020-2021
 - Next position: PhD student at Johns Hopkins University
- Antonia Creswell (undergraduate student), 2014-2015
 - Next position: PhD student at Imperial College London
- Xie Zhou (undergraduate student), 2017-2018
 - Next position: MS student at UC Berkeley
- Yi Mang (Terry) Yang (undergraduate student), 2017-2018
 - Next position: Software Engineer at Amazon
- Daniel Bolya (undergraduate student), 2018-2019
 - [Chancellor's Award for Excellence in Undergraduate Research \(Honorable Mention\) 2019](#)
 - Next position: PhD student at Georgia Tech
- Aron Sarmasi (undergraduate student), 2018-2019
 - Next position: MS student at UC Davis
- Waiyu Lam (undergraduate student), 2019-2020
 - Next position: MS student at Cornell
- Moniek Smink (Undergraduate student), 2022-2024
 - [Hilldale Research Fellowship 2023-2024](#)
 - Next position: MS student at TU Delft
- Jedidiah Schloesser (Undergraduate student), 2023-2024

- Bocheng Zou (Undergraduate student), 2024-2025
 - [Trewartha Senior Thesis Research Award 2024](#)
 - Next position: MS student at UW-Madison
- Sigmund Vanvalkenburgh (Undergraduate student), 2024-2025
- Donghyun Lee (Undergraduate student), 2024-2025
- Le Thien Phuc Nguyen (Undergraduate student), 2024-2025
- Qi Zhu (undergraduate intern), 2015 summer
 - Next position: MS student at UIUC
- Xiuye Gu (undergraduate intern), 2016 summer
 - Next position: MS student at Stanford
- Haolin Fu (undergraduate intern), 2016 summer
 - Next position: MS student at Yale
- Haotian Liu (undergraduate intern), 2018 summer
 - Next position: PhD student at UC Davis
- Hao Yu (undergraduate intern), 2018 summer
 - Next position: PhD student at Boston University
- Gautam Pradeep (high school student), 2016-2019
 - [1st place, National Environmentors Science Fair 2017](#)
 - [2nd place, National Environmentors Science Fair 2018](#)
 - Next position: undergraduate student at Stanford