Hexiang (Frank) Hu

hexiang.frank.hu@gmail.com
http://hexiang-hu.github.io

EMPLOYMENT

Research Scientist   March 2023 - Present
Google DeepMind   Seattle, WA

Research Scientist   June 2021 - March 2023
Google Brain   Seattle, WA

EDUCATIONS

University of Southern California   2017 - 2021
Ph.D. in Computer Science
Advisor: Fei Sha

University of California, Los Angeles (UCLA)   2016 - 2017
Ph.D. in Computer Science (Transfer)
Advisor: Fei Sha

Simon Fraser University   2013 - 2016
Bachelor of Science in Computer Science
First class with distinction

Zhejiang University   2011 - 2013
Bachelor of Engineering in Computer Science and Technology

PUBLICATIONS

Peter Shaw, Mandar Joshi, James Cohan, Jonathan Berant, Panupong Pasupat, **Hexiang Hu**, Urvashi Khandelwal, Kenton Lee, and Kristina N Toutanova. "From Pixels to UI Actions: Learning to Follow Instructions via Graphical User Interfaces". In: *NeurIPS*. 2023

Wenhu Chen*, **Hexiang Hu***, Yandong Li, Nataniel Ruiz, Xuhui Jia, Ming-Wei Chang, and William W Cohen. "Subject-driven Text-to-Image Generation via Apprenticeship Learning". In: *NeurIPS*. 2023

Yang Chen, **Hexiang Hu**, Yi Luan, Haitian Sun, Soravit Changpinyo, Alan Ritter, and Ming-Wei Chang. "Can Pre-trained Vision and Language Models Answer Visual Information-Seeking Questions?" In: *EMNLP*. 2023

Hexiang Hu, Yi Luan, Yang Chen, Urvashi Khandelwal, Mandar Joshi, Kenton Lee, Kristina Toutanova, and Ming-Wei Chang. "Open-domain Visual Entity Recognition: Towards Recognizing Millions of Wikipedia Entities". In: *ICCV*. 2023

Jihyung Kil, Soravit Changpinyo, Xi Chen, **Hexiang Hu**, Sebastian Goodman, Wei-Lun Chao, and Radu Soricut. "Prestu: Pre-training for scene-text understanding". In: *ICCV*. 2023

Wenhu Chen, **Hexiang Hu**, Chitwan Saharia, and William W Cohen. "Re-imagen: Retrieval-augmented text-to-image generator". In: *ICLR*. 2023

Wenhu Chen, **Hexiang Hu**, Xi Chen, Pat Verga, and William W Cohen. "Murag: Multimodal retrieval-augmented generator for open question answering over images and text". In: *EMNLP*. 2022

Hexiang Hu, Ozan Sener, Fei Sha, and Vladlen Koltun. "Drinking from a Firehose: Continual Learning with Web-scale Natural Language". In: *T-PAMI* (2023)

---

*: Authors contributed equally

Tai-Yu Pan, Cheng Zhang, Yandong Li, **Hexiang Hu**, Dong Xuan, Soravit Changpinyo, Boqing Gong, and Wei-Lun Chao. "On model calibration for long-tailed object detection and instance segmentation". In: *NeurIPS*. 2021

Linlu Qiu, **Hexiang Hu**, Bowen Zhang, Peter Shaw, and Fei Sha. "Systematic Generalization on gSCAN: What is Nearly Solved and What is Next?" In: *EMNLP*. 2021

Bowen Zhang, **Hexiang Hu**, Linlu Qiu, Peter Shaw, and Fei Sha. "Visually Grounded Concept Composition". In: *Findings of EMNLP*. 2021

Cheng Zhang, Tai-Yu Pan, Yandong Li, **Hexiang Hu**, Dong Xuan, Soravit Changpinyo, Boqing Gong, and Wei-Lun Chao. "MosaicOS: a simple and effective use of object-centric images for long-tailed object detection". In: *ICCV*. 2021

Han-Jia Ye, **Hexiang Hu**, and De-Chuan Zhan. "Learning adaptive classifiers synthesis for generalized few-shot learning". In: *IJCV*. Springer, 2021

Jiacheng Chen*, **Hexiang Hu***, Hao Wu, Yuning Jiang, and Changhu Wang. "Learning the Best Pooling Strategy for Visual Semantic Embedding". In: *CVPR*. 2021

Bowen Zhang*, **Hexiang Hu***, Vihan Jain, Eugene Ie, and Fei Sha. "Learning to Represent Image and Text with Denotation Graph". In: *EMNLP*. 2020

Wang Zhu*, **Hexiang Hu***, Jiacheng Chen, Zhiwei Deng, Vihan Jain, Eugene Ie, and Fei Sha. "BabyWalk: Going Farther in Vision-and-Language Navigation by Taking Baby Steps". In: *ACL*. 2020

Han-Jia Ye, **Hexiang Hu**, De-Chuan Zhan, and Fei Sha. "Few-shot learning via embedding adaptation with set-to-set functions". In: *CVPR*. 2020

Risto Vuorio*, Shao-Hua Sun*, **Hexiang Hu**, and Joseph J Lim. "Multimodal Model-Agnostic Meta-Learning via Task-Aware Modulation". In: *NeurIPS*. 2019

Hexiang Hu, Ishan Misra, and Laurens van der Maaten. "Evaluating text-to-image matching using binary image selection (BISON)". in: *ICCV Workshop*. 2019

Bowen Zhang, Hexiang Hu, and Fei Sha. "Visual Storytelling via Predicting Anchor Word Embeddings in the Stories". In: *ICCV Workshop*. 2019

Nelson Nauata, **Hexiang Hu**, Guang-Tong Zhou, Zhiwei Deng, Zicheng Liao, and Greg Mori. "Structured label inference for visual understanding". In: *T-PAMI* (2019)

Kurt Shuster, Samuel Humeau, **Hexiang Hu**, Antoine Bordes, and Jason Weston. "Engaging image captioning via personality". In: *CVPR*. 2019

Hexiang Hu*, Liyu Chen*, Boqing Gong, and Fei Sha. "Synthesized policies for transfer and adaptation across tasks and environments". In: *NeurIPS*. 2018

Soravit Changpinyo, **Hexiang Hu**, and Fei Sha. "Multi-task learning for sequence tagging: An empirical study". In: *COLING*. 2018

Bowen Zhang*, **Hexiang Hu***, and Fei Sha. "Cross-modal and hierarchical modeling of video and text". In: *ECCV*. 2018

Hexiang Hu*, Wei-Lun Chao, and Fei Sha. "Learning answer embeddings for visual question answering". In: *CVPR*. 2018

Wei-Lun Chao*, **Hexiang Hu***, and Fei Sha. "Cross-dataset adaptation for visual question answering". In: *CVPR*. 2018

Wei-Lun Chao*, **Hexiang Hu***, and Fei Sha. "Being negative but constructively: Lessons learnt from creating better visual question answering datasets". In: *NAACL-HLT*. 2018

Chao-Yuan Wu, Manzil Zaheer, **Hexiang Hu**, R Manmatha, Alexander J Smola, and Philipp Krähenbühl. "Compressed video action recognition". In: *CVPR*. 2018

Hexiang Hu*, **Shiyi Lan***, Yuning Jiang, Zhimin Cao, and Fei Sha. "Fastmask: Segment multi-scale object candidates in one shot". In: *CVPR*. 2017

Hexiang Hu, Guang-Tong Zhou, Zhiwei Deng, Zicheng Liao, and Greg Mori. "Learning structured inference neural networks with label relations". In: *CVPR*. 2016

Zhiwei Deng, Arash Vahdat, **Hexiang Hu**, and Greg Mori. "Structure inference machines: Recurrent neural networks for analyzing relations in group activity recognition". In: *CVPR*. 2016

Chao-Yuan Wu, Manzil Zaheer, **Hexiang Hu**, R Manmatha, Alexander J Smola, and Philipp Krähenbühl. "Compressed video action recognition". In: *CVPR*. 2018

Hexiang Hu*, **Shiyi Lan***, Yuning Jiang, Zhimin Cao, and Fei Sha. "Fastmask: Segment multi-scale object candidates in one shot". In: *CVPR*. 2017

Hexiang Hu, Guang-Tong Zhou, Zhiwei Deng, Zicheng Liao, and Greg Mori. "Learning structured inference neural networks with label relations". In: *CVPR*. 2016

Zhiwei Deng, Arash Vahdat, **Hexiang Hu**, and Greg Mori. "Structure inference machines: Recurrent neural networks for analyzing relations in group activity recognition". In: *CVPR*. 2016

TECHNICAL REPORTS

Gemini Team. "Gemini: a family of highly capable multimodal models". In: *Technical Report*. 2023

Bowen Zhang*, **Hexiang Hu***, Joonseok Lee*, Ming Zhao, Sheide Chammas, Vihan Jain, Eugene Ie, and Fei Sha. "A Hierarchical Multi-Modal Encoder for Moment Localization in Video Corpus". In: *arXiv preprint arXiv:2011.09046* (2020)

Hexiang Hu*, Zhiwei Deng*, Guang-Tong Zhou, Fei Sha, and Greg Mori. "Labelbank: Revisiting global perspectives for semantic segmentation". In: *arXiv preprint arXiv:1703.09891* (2017)

PREPRINTS

Kai Zhang, Yi Luan, **Hexiang Hu**, Kenton Lee, Siyuan Qiao, Wenhu Chen, Yu Su, and Ming-Wei Chang. "MagicLens: Self-Supervised Image Retrieval with Open-Ended Instructions". In: *arXiv preprint arXiv:2403.19651* (2024)

Hexiang Hu*, Kelvin CK Chan*, Yu-Chuan Su*, Wenhu Chen*, Yandong Li, Kihyuk Sohn, Yang Zhao, Xue Ben, Boqing Gong, William Cohen, et al. "Instruct-Imagen: Image generation with multi-modal instruction". In: *arXiv preprint arXiv:2401.01952* (2024)

Xi Chen, Josip Djolonga, Piotr Padlewski, Basil Mustafa, Soravit Changpinyo, Jialin Wu, Carlos Riquelme Ruiz, Sebastian Goodman, Xiao Wang, Yi Tay, et al. "Pali-x: On scaling up a multilingual vision and language model". In: *arXiv preprint arXiv:2305.18565* (2023)

Cong Wei, Yang Chen, Haonan Chen, **Hexiang Hu**, Ge Zhang, Jie Fu, Alan Ritter, and Wenhu Chen. "Uniir: Training and benchmarking universal multimodal information retrievers". In: *arXiv preprint arXiv:2311.17136* (2023)

PATENTS

R Manmatha, **Hexiang Hu**, and Deva Ramanan. *Compressed content object and action detection.* US Patent 10,528,819. Jan. 2020

Jain Vihan, JoonSeok Lee, Ming Zhao, Sheide Chammas, **Hexiang Hu**, Bowen Zhang, Sha Fei, and Tze Way Eugene Ie. *Hierarchical video encoders.* US Patent 11,533,495. Dec. 2022

SERVICES

Conference Reviewer
- Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representation (ICLR)
- Annual Meeting of the Association for Computational Linguistics (ACL)
- The North American Chapter of the ACL (NAACL)
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)
- Asian Conference on Computer Vision (ACCV)
- Winter Conference on Computer Vision (WACV)
- AAAI Conference on Artificial Intelligence (AAAI)
- Internation Conference on Image Processing (ICIP)

Journal Reviewer
- IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)

TECHNICAL TALKS

"Understanding Language in Perception and Embodiment"
- Google, Mountain View, CA — Feb 2021
- Facebook, New York, NY — Feb 2021
- Waymo, Mountain View, CA — March 2021
- University of Washington, Seattle, WA — Feb 2021
- University of California, Berkeley, CA — March 2021
- University of Michigan, Ann Arbor, MI — March 2021
- University of California, Santa Cruz, CA — April 2021

"Learning Visually Grounded Semantics with Denotation Graph"
- Google, Mountain View, CA — Sept 2020
- USC, Los Angeles, CA — March 2021

"Learning Generalizeable Neural Agent across Different Task Horizons"
- Google, Mountain View, CA — Nov 2019

"Synthesized Policies for Transfer and Adaptation across Tasks and Environments"
- NeurIPS, Montreal, QC — Dec 2018

"Being Negative but Constructively: Lessons from Creating Better VQA datasets"
- NAACL, New Orleans, LA — Jun 2018
- USC, Los Angeles, CA — April 2021

TEACHING	*Teaching Assistant*, University of Southern California	
	CSCI-599: Deep Learning and its Applications	Fall 2017 - 2018
	CSCI-567: Machine Learning	Spring 2020 - 2021

AWARDS & HONORS

Awards and Prizes

- NeurIPS Student Travel Award   2018, 2019
- Simon Fraser University President's Honour Roll   2014, 2015
- Zhejiang University First-class Academic Excellence Award   2011, 2012

Scholarship & Fellowship

- USC CS Department Fellowship   2017
- UCLA CS Department Fellowship   2016
- Simon Fraser University Entrance Scholarship   2013

PRESS COVERAGE

"BabyWalk AI breaks complex navigation into simple steps"
by Kyle Wiggers, VentureBeat   May 12, 2020

"Facebook aims to be more engaging to humans in latest AI research"
by Tiernan Ray, ZDNet   Oct 26, 2018

ADVISING & MENTORSHIP

Junior Ph.D Mentor

- Bowen Zhang, Ph.D student at USC
- Liyu Chen, Ph.D student at USC

Visiting Scholar & Intern Mentor

- Kai Zhang, Intern at Google Deepmind
- Yang Chen, Intern at Google Brain
- Jihyung Kil, Intern at Google
- Wang Zhu, SFU, Next: Ph.D student at USC
- Jiacheng Chen, SFU, Next: Ph.D student at USC
- Han-jia Ye, Nanjing University, Next: Assistant professor at Nanjing University
- Risto Vuorio, SK-Brain, Next: Ph.D student at Cambridge University
- Shiyi Lan, Fudan University, Next: Ph.D student at UMD
- Linlu Qiu, AI Resident at Google, Next: Ph.D student at MIT

Undergraduate Student Mentor

- Yuna Lee, USC, Next: Software Engineer at Google