

Long Zhao

Google LAX
340 Main Street, Los Angeles, CA 90291

Email: gary.zhao9012@gmail.com
Homepage: <https://garyzhao.github.io>
LinkedIn: <https://www.linkedin.com/in/garyzhao9012>

RESEARCH INTERESTS

- **Large Vision Foundation Models** (e.g., Video-Language Models, Multimodal Models, World Models)
- **Generative Models** (e.g., Diffusion Models, Visual Autoregressive Models)
- **Self-Supervised Representation Learning** (e.g., Contrastive Learning, Mask Modeling)
- **Contextualized Machine Perception** (e.g., Recognition, Detection, Segmentation, Localization)

EDUCATION

- **Rutgers, The State University of New Jersey – New Brunswick** **Piscataway, NJ, USA**
Ph.D. in Computer Science | GPA: 4.0/4.0 *09/2016 – 01/2022*
 - **Advisor:** Prof. Dimitris N. Metaxas (Distinguished Professor)
 - **Ph.D. Thesis:** “Towards Efficient and Reliable Skeleton-Based Human Pose Modeling”
 - 3D Human/Hand Pose Estimation from RGB Images [CVPR’19, CVPR’20]
 - Face/Pose/Video Generation with GANs [IJCAI’18, ECCV’18, IJCV’20]
 - Domain Generalization via Adversarial Training & Meta-Learning [CVPR’20, NeurIPS’20]
 - Representation Learning on Graphs with Graph Convolutional Networks [NeurIPS’19]
- **Tongji University** **Shanghai, China**
M.S. in Software Engineering | GPA: 87.89/100.0 *09/2012 – 06/2015*
 - **Advisor:** Prof. Jinyuan Jia & Prof. Shuang Liang
 - **Master Thesis:** “Sketch-Based 3D Model Retrieval”
B.Eng. in Software Engineering | GPA: 4.56/5.0 *09/2008 – 07/2012*
 - **Key Courses:** Programming Language, Data Structure and Algorithm, Operation Systems, Computer Network, Computer Architecture, Software Engineering, Computer Graphics

EXPERIENCE

- **Foundational Research, Google DeepMind** **Los Angeles, CA, USA**
Research Scientist. *05/2024 – Present*
 - Video Understanding (e.g., Multimodal Models, World Models)
 - Video Generation (e.g., Diffusion Models, Visual Autoregressive Models)
 - Machine Perception (e.g., Recognition, Detection, Localization)
- **Perception Team, Google Research** **Los Angeles, CA, USA**
Research Scientist. *11/2021 – 05/2024*
 - Video Foundation Models and Benchmarks [TMLR’24, ICML’24]
 - Multimodal Learning with Vision-Language Models [ICLR’24, CVPR’24]
 - Contextualized Machine Perception [ICCV’23, CVPR’24, CVPR’24]
- **Brain Team, Google Research** **Mountain View, CA, USA**
Student Researcher. Host: Dr. Han Zhang *12/2020 – 05/2021*
 - Boosting Transformers for High-Resolution Image Generation [NeurIPS’21]
 - Improving Efficiency and Interpretability for Vision Transformers [AAAI’22 (Oral)]

• **Mobile Vision Team, Google Research**

Los Angeles, CA, USA

• *Research Intern & Student Researcher. Host: Dr. Ting Liu*

05/2020 – 12/2020

- View-Disentangled Human Pose Representation Learning [[CVPR'21 \(Oral\)](#)]
- View-Invariant, Occlusion-Robust Probabilistic Pose Embedding [[IJCV'21](#)]

• **Visual Computing Group, Microsoft Research Asia (MSRA)**

Beijing, China

• *Research Intern. Mentor: Dr. Yichen Wei*

12/2013 – 11/2014

- Generic Object Proposal Generation [[CVPR'15](#)]
- Salient Object Detection [[ACCV'14](#)]
- Won the *award of excellence* in the “MSRA Stars of Tomorrow Internship Program”

SELECTED PUBLICATIONS

(* indicates equal contributions. Please check [Google Scholar](#) for the full list of my publications.)

Technical Reports:

- [1] Yihong Sun, Hao Zhou, Liangzhe Yuan, Jennifer J. Sun, Yandong Li, Xuhui Jia, Hartwig Adam, Bharath Hariharan, **Long Zhao**, and Ting Liu, “[Video Creation by Demonstration](#)”. *Technical Report*, arXiv:2412.09551, 2024.
- [2] Jennifer J. Sun*, Hao Zhou, **Long Zhao**, Liangzhe Yuan, Bryan Seybold, David Hendon, Florian Schroff, David A. Ross, Hartwig Adam, Bo Hu, and Ting Liu*, “[Video Foundation Models for Animal Behavior Analysis](#)”. *Technical Report*, bioRxiv:2024.07.30.605655, 2024.

Book Chapters:

- [3] Dimitris N. Metaxas, **Long Zhao**, and Xi Peng, “[Disentangled Representation Learning and Its Application to Face Analytics](#)”. In: *Deep Learning-Based Face Analytics*, Pages 45-72, Springer, 2021.

Journals:

- [4] Liangzhe Yuan*, Nitesh B. Gundavarapu*, **Long Zhao***, Hao Zhou*, Yin Cui, Lu Jiang, Xuan Yang, Menglin Jia, Tobias Weyand, Luke Friedman, Mikhail Sirotenko, Huisheng Wang, Florian Schroff, Hartwig Adam, Ming-Hsuan Yang, Ting Liu, and Boqing Gong, “[VideoGLUE: Video General Understanding Evaluation of Foundation Models](#)”. *Transactions on Machine Learning Research (TMLR)*, 2024.
- [5] Xi Peng, Fengchun Qiao, and **Long Zhao**, “[Out-of-Domain Generalization from a Single Source: An Uncertainty Quantification Approach](#)”. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Volume 46, Issue 3, Pages 1775-1787, 2024.
- [6] Ting Liu*, Jennifer J. Sun*, **Long Zhao**, Jiaping Zhao, Liangzhe Yuan, Yuxiao Wang, Liang-Chieh Chen, Florian Schroff, and Hartwig Adam, “[View-Invariant, Occlusion-Robust Probabilistic Embedding for Human Pose](#)”. *International Journal of Computer Vision (IJCV)*, Volume 130, Issue 1, Pages 111-135, 2022.
- [7] **Long Zhao**, Xi Peng, Yu Tian, Mubbasir Kapadia, and Dimitris Metaxas, “[Towards Image-to-Video Translation: A Structure-Aware Approach via Multi-Stage Generative Adversarial Networks](#)”. *International Journal of Computer Vision (IJCV)*, Volume 128, Issue 10, Pages 2514-2533, 2020.

Conference Proceedings:

- [8] **Long Zhao**, Sanghyun Woo, Ziyu Wan, Yandong Li, Han Zhang, Boqing Gong, Hartwig Adam, Xuhui Jia, and Ting Liu, “[Epsilon-VAE: Denoising as Visual Decoding](#)”. In: Proceedings of the International Conference on Machine Learning (ICML), 2025.

- [9] Zhuowei Li, Haizhou Shi, Yunhe Gao, Di Liu, Zhenting Wang, Yuxiao Chen, Ting Liu, **Long Zhao**, Hao Wang, Dimitris N. Metaxas, “[The Hidden Life of Tokens: Reducing Hallucination of Large Vision-Language Models Via Visual Information Steering](#)”. In: Proceedings of the International Conference on Machine Learning (ICML), 2025.
- [10] Xiaoyu Zhu, Hao Zhou, Pengfei Xing, **Long Zhao**, Hao Xu, Junwei Liang, Alexander Hauptmann, Ting Liu, and Andrew Gallagher, “[Open-Vocabulary 3D Semantic Segmentation with Text-to-Image Diffusion Models](#)”. In: *Proceedings of the 18th European Conference on Computer Vision (ECCV)*, Pages 357-375, 2024. **[Strong Double Blind]**
- [11] **Long Zhao**^{*}, Nitesh B. Gundavarapu^{*}, Liangzhe Yuan^{*}, Hao Zhou^{*}, Shen Yan[†], Jennifer J. Sun[†], Luke Friedman[†], Rui Qian[†], Tobias Weyand, Yue Zhao, Rachel Hornung, Florian Schroff, Ming-Hsuan Yang, David A. Ross, Huisheng Wang, Hartwig Adam, Mikhail Sirotenko[‡], Ting Liu[‡], and Boqing Gong[‡], “[VideoPrism: A Foundational Visual Encoder for Video Understanding](#)”. In: *Proceedings of the International Conference on Machine Learning (ICML)*, Pages 60785-60811, 2024. (*equal primary contributions; [†]equal core technical contributions; [‡]equal senior contributions.)
- [12] Yue Zhao, **Long Zhao**, Xingyi Zhou, Jialin Wu, Chun-Te Chu, Hui Miao, Florian Schroff, Hartwig Adam, Ting Liu, Boqing Gong, Philipp Krähenbühl, and Liangzhe Yuan, “[Distilling Vision-Language Models on Millions of Videos](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 13106-13116, 2024.
- [13] Shiyu Zhao, **Long Zhao**, Vijay Kumar B.G, Yumin Suh, Dimitris N. Metaxas, Manmohan Chandraker, and Samuel Schulter, “[Generating Enhanced Negatives for Training Language-Based Object Detectors](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 13592-13602, 2024.
- [14] Shiyu Zhao, Samuel Schulter, **Long Zhao**, Zhixing Zhang, Vijay Kumar B.G, Yumin Suh, Manmohan Chandraker, and Dimitris N. Metaxas, “[Taming Self-Training for Open-Vocabulary Object Detection](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 13938-13947, 2024.
- [15] Yuanhao Xiong, **Long Zhao**, Boqing Gong, Ming-Hsuan Yang, Florian Schroff, Ting Liu, Cho-Jui Hsieh, and Liangzhe Yuan, “[Structured Video-Language Modeling with Temporal Grouping and Spatial Grounding](#)”. In: *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024.
- [16] Qitong Wang, **Long Zhao**, Liangzhe Yuan, Ting Liu, and Xi Peng, “[Learning from Semantic Alignment between Unpaired Multiviews for Egocentric Video Recognition](#)”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Pages 3307-3317, 2023.
- [17] **Long Zhao**, Liangzhe Yuan, Boqing Gong, Yin Cui, Florian Schroff, Ming-Hsuan Yang, Hartwig Adam, and Ting Liu, “[Unified Visual Relationship Detection with Vision and Language Models](#)”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Pages 6962-6973, 2023.
- [18] Shiyu Zhao, Zhixing Zhang, Samuel Schulter, **Long Zhao**, Vijay Kumar B.G, Anastasis Sathopoulos, Manmohan Chandraker, Dimitris N. Metaxas, “[Exploiting Unlabeled Data with Vision and Language Models for Object Detection](#)”. In: *Proceedings of the 17th European Conference on Computer Vision (ECCV)*, Pages 159-175, 2022.
- [19] Yuxiao Chen, **Long Zhao**, Jianbo Yuan, Yu Tian, Zhaoyang Xia, Shijie Geng, Ligong Han, Dimitris N. Metaxas, “[Hierarchically Self-supervised Transformer for Human Skeleton Representation Learning](#)”. In: *Proceedings of the 17th European Conference on Computer Vision (ECCV)*, Pages 185-202, 2022.
- [20] Honglu Zhou, Asim Kadav, Aviv Shamsian, Shijie Geng, Farley Lai, **Long Zhao**, Ting Liu, Mubbasir Kapadia, and Hans Peter Graf, “[COMPOSER: Compositional Reasoning of Group Activity in Videos with Keypoint-Only Modality](#)”. In: *Proceedings of the 17th European Conference on Computer Vision (ECCV)*, Pages 249-266, 2022.

- [21] Shiyu Zhao, **Long Zhao**, Zhixing Zhang, Enyu Zhou, and Dimitris Metaxas, “[Global Matching with Overlapping Attention for Optical Flow Estimation](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 17592-17601, 2022.
- [22] Mengmeng Ma, Jian Ren, **Long Zhao**, Davide Testuggine, and Xi Peng, “[Are Multimodal Transformers Robust to Missing Modality?](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 18177-18186, 2022.
- [23] Zizhao Zhang, Han Zhang, **Long Zhao**, Ting Chen, Sercan Arik, and Tomas Pfister, “[Nested Hierarchical Transformer: Towards Accurate, Data-Efficient and Interpretable Visual Understanding](#)”. In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Pages 3417-3425, 2022. **[Oral Presentation]**
- [24] **Long Zhao**, Zizhao Zhang, Ting Chen, Dimitris N. Metaxas, and Han Zhang, “[Improved Transformer for High-Resolution GANs](#)”. In: *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, Pages 18367-18380, 2021.
- [25] **Long Zhao**, Yuxiao Wang, Jiaping Zhao, Liangzhe Yuan, Jennifer J. Sun, Florian Schroff, Hartwig Adam, Xi Peng, Dimitris Metaxas, and Ting Liu, “[Learning View-Disentangled Human Pose Representation by Contrastive Cross-View Mutual Information Maximization](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 12793-12802, 2021. **[Oral Presentation]**
- [26] Mengmeng Ma, Jian Ren, **Long Zhao**, Sergey Tulyakov, Cathy Wu, and Xi Peng, “[SMIL: Multimodal Learning with Severely Missing Modality](#)”. In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Pages 2302-2310, 2021.
- [27] **Long Zhao**, Ting Liu, Xi Peng, and Dimitris Metaxas, “[Maximum-Entropy Adversarial Data Augmentation for Improved Generalization and Robustness](#)”. In: *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, Pages 14435-14447, 2020.
- [28] **Long Zhao**, Xi Peng, Yuxiao Chen, Mubbasir Kapadia, and Dimitris N. Metaxas, “[Knowledge as Priors: Cross-Modal Knowledge Generalization for Datasets without Superior Knowledge](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 6528-6537, 2020.
- [29] Fengchun Qiao, **Long Zhao**, and Xi Peng, “[Learning to Learn Single Domain Generalization](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 12556-12565, 2020.
- [30] Yu Tian*, **Long Zhao***, Xi Peng, and Dimitris N. Metaxas, “[Rethinking Kernel Methods for Node Representation Learning on Graphs](#)”. In: *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, Pages 11681-11692, 2019.
- [31] Yuxiao Chen, **Long Zhao**, Xi Peng, Jianbo Yuan, and Dimitris N. Metaxas, “[Construct Dynamic Graphs for Hand Gesture Recognition via Spatial-Temporal Attention](#)”. In: *Proceedings of the British Machine Vision Conference (BMVC)*, 2019.
- [32] **Long Zhao**, Xi Peng, Yu Tian, Mubbasir Kapadia, and Dimitris N. Metaxas, “[Semantic Graph Convolutional Networks for 3D Human Pose Regression](#)”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 3425-3435, 2019.
- [33] **Long Zhao**, Xi Peng, Yu Tian, Mubbasir Kapadia, and Dimitris N. Metaxas, “[Learning to Forecast and Refine Residual Motion for Image-to-Video Generation](#)”. In: *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, Pages 387-403, 2018.
- [34] Yu Tian, Xi Peng, **Long Zhao**, Shaoting Zhang, and Dimitris N. Metaxas, “[CR-GAN: Learning Complete Representations for Multi-view Generation](#)”. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, Pages 942-948, 2018.
- [35] Chaoyang Wang, **Long Zhao**, Shuang Liang, Liqing Zhang, Jinyuan Jia, and Yichen Wei, “[Object Proposal by Multi-branch Hierarchical Segmentation](#)”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Pages 3873-3881, 2015.

MISCELLANEOUS

Professional Activities:

- **Area Chair:** ICML 2025, ICCV 2025, NeurIPS 2025
- **Organizing Committee** of [2nd OmniLabel Workshop at ECCV 2024](#)
- **Organizing Committee** of [1st OmniLabel Workshop at CVPR 2023](#)
- Conference Reviewer: CVPR, ICCV, ECCV, NeurIPS, ICLR, AAAI, WACV
- Journal Reviewer: TPAMI, TIP, CVIU, GMOD

Honors & Activities:

- **NeurIPS Top Reviewer Award.** Neural Information Processing Systems, 2024
- **NeurIPS Outstanding Reviewer Award.** Neural Information Processing Systems, 2021
- **Outstanding Graduate Student Fellowship.** Rutgers University, 2016 – 2018
- **Excellent Award of Stars of Tomorrow Internship Program.** Microsoft Research Asia (MSRA), 2015
- **Excellent Learning Scholarship.** Tongji University, 2009 – 2012