

Yuanhao Zhai

E-mail: yuanhaozhai@gmail.com ◇ Phone: +1 (248)-568-6387 ◇ Homepage: yhzhai.com

RESEARCH INTEREST

Image editing/generation, video understanding/generation, human action analysis/synthesis, reinforcement learning, multi-modal learning, and weakly-supervised learning.

PROFESSIONAL EXPERIENCE

Meta, Menlo Park, CA, USA Jun 2025 - Present
Research scientist
Topic: Image generation and editing, video generation, and reinforcement learning

EDUCATION

State University of New York at Buffalo, Buffalo, NY, USA Feb 2021 - Jun 2025

- PhD student in Computer Science & Engineering.
- Advisors: Prof. David Doermann and Prof. Junsong Yuan.
- Thesis: Video Generation and Understanding with Multimodal Learning

Xi'an Jiaotong University, Xi'an, China Sep 2016 - Jul 2020

- Bachelor of Engineering in Computer Science.
- Advisors: Ziyi Liu and Prof. Le Wang.
- Outstanding Graduate Award

University of California, Berkeley, Berkeley, CA, USA Jan 2019 - May 2019

- Exchange student of the Berkeley International Study Program.

INTERNSHIP EXPERIENCE

Amazon, New York, NY, USA May 2024 - Nov 2024
Applied scientist intern
Topic: Text-to-outfit generation using large language models

Microsoft, Remote Feb 2024 - May 2024
Research intern
Topic: Video diffusion distillation

Microsoft, Remote May 2023 - Aug 2023
Research intern
Topic: Human-centric joint video-depth generation with diffusion models

Wormpex AI Research, Bellevue, WA, USA May 2022 - Aug 2022
Research intern
Topic: Human action analysis and open-set recognition

Institute for Artificial Intelligence and Data Science, University at Buffalo, Buffalo, NY, USA Aug 2021 - May 2024
Research assistant
Topic: Image manipulation detection and Deepfakes detection

Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China Feb 2018 - Jul 2021
Research intern
Project: Weakly-supervised temporal action localization

1. Yuanhao Zhai, Yen-Liang Lin, Minxu Peng, Larry S. Davis, Ashwin Chandramouli, Junsong Yuan, David Doermann, “Text2Outfit: Controllable Outfit Generation with Multimodal Language Models”, in *IEEE International Conference on Computer Vision (ICCV)*, 2025.
2. Mahesh Bhosale, Abdul Wasi, Yuanhao Zhai, Yunjie Tian, Samuel Border, Nan Xi, Pinaki Sarder, Junsong Yuan, David Doermann, Xuan Gong, “PathDiff: Histopathology Image Synthesis with Unpaired Text and Mask Conditions”, in *IEEE International Conference on Computer Vision (ICCV)*, 2025.
3. Tianyu Luan, Yuanhao Zhai, Jingjing Meng, Zhong Li, Zhang Chen, Yi Xu, Junsong Yuan, “Scalable High-Fidelity 3D Hand Shape Reconstruction Via Graph-Image Frequency Mapping and Graph Frequency Decomposition”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2025.
4. Yining Hong, Beide Liu, Maxine Wu, Yuanhao Zhai, Kai-Wei Chang, Linjie Li, Kevin Lin, Chung-Ching Lin, Jianfeng Wang, Zhengyuan Yang, Yingnian Wu, and Lijuan Wang, “SlowFast-VGen: Slow-Fast Learning for Action-Driven Long Video Generation”, in *International Conference on Learning Representations (ICLR)*, 2025.
5. Yuanhao Zhai, Kevin Lin, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Chung-Ching Lin, David Doermann, Junsong Yuan, and Lijuan Wang, “Motion Consistency Model: Accelerating Video Diffusion with Disentangled Motion-Appearance Distillation”, in *Conference on Neural Information Processing Systems (NeurIPS)*, 2024. [[Project](#)] [[Paper](#)] [[Code](#)].
6. Yuanhao Zhai, Kevin Lin, Linjie Li, Chung-Ching Lin, Jianfeng Wang, Zhengyuan Yang, David Doermann, Junsong Yuan, Zicheng Liu, and Lijuan Wang, “IDOL: Unified Dual-Modal Latent Diffusion for Human-Centric Joint Video-Depth Generation”, in *European Conference on Computer Vision (ECCV)*, 2024. [[Project](#)] [[Paper](#)] [[Code](#)].
7. Yangcen Liu, Ziyi Liu, Yuanhao Zhai, Wen Li, David Doerman, and Junsong Yuan, “STAT: Towards Generalizable Temporal Action Localization”, arXiv preprint arXiv:2404.13311, 2024. [[Paper](#)]
8. Fabing Li, Yuanhao Zhai, Shuangyu Cai, and Mingyu Gao, “Seesaw: Compensating for Nonlinear Reduction with Linear Computations for Private Inference”, in *International Conference on Machine Learning (ICML)*, 2024. [[Paper](#)]
9. Tan Wang, Linjie Li, Kevin Lin, Yuanhao Zhai, Chung-Ching Lin, Zhengyuan Yang, Hanwang Zhang, Zicheng Liu, and Lijuan Wang, “DisCo: Disentangled Control for Referring Human Dance Generation in Real World”, in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2024. [[Paper](#)]
10. Tianyu Luan, Yuanhao Zhai, Jingjing Meng, Zhong Li, Zhang Chen, Yi Xu, and Junsong Yuan, “High Fidelity 3D Hand Shape Reconstruction via Scalable Graph Frequency Decomposition”, in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2023. [[Paper](#)]
11. Yuanhao Zhai, Mingzhen Huang, Tianyu Luan, Lu Dong, Ifeoma Nwogu, Siwei Lyu, David Doermann, and Junsong Yuan, “Language-guided Human Motion Synthesis with Atomic Actions”, in *ACM Multimedia (MM)*, 2023. [[Paper](#)] [[Code](#)]
12. Yuanhao Zhai, Ziyi Liu, Zhenyu Wu, Yi Wu, Chunlun Zhou, David Doermann, Junsong Yuan, and Gang Hua, “SOAR: Scene-debiasing Open-set Action Recognition”, in *IEEE International Conference on Computer Vision (ICCV)*, 2023. [[Paper](#)] [[Code](#)]
13. Yuanhao Zhai, Tianyu Luan, David Doermann, and Junsong Yuan, “Towards Generic Image Manipulation Detection with Weakly-Supervised Self-Consistency Learning”, in *IEEE International*

Conference on Computer Vision (ICCV), 2023. [Paper] [Code]

14. Yuanhao Zhai, Le Wang, Wei Tang, Qilin Zhang, Nanning Zheng, David Doermann, Junsong Yuan, and Gang Hua, “Adaptive Two-Stream Consensus Network for Weakly-Supervised Temporal Action Localization”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2022. [Paper]
15. Yuanhao Zhai, Le Wang, Wei Tang, Qilin Zhang, Nanning Zheng, and Gang Hua, “Action Coherence Network for Weakly-Supervised Temporal Action Localization”, in *IEEE Transactions on Multimedia (TMM)*, vol. 24, pp. 1857-1870, 2022. [Paper]
16. Le Wang, Rizhi Ding, Yuanhao Zhai, Qilin Zhang, Wei Tang, Nanning Zheng, and Gang Hua, “Giant Panda Identification”, in *IEEE Transactions on Image Processing (TIP)*, vol. 30, pp. 2837-2849, 2021. [Paper]
17. Yuanhao Zhai, Le Wang, Wei Tang, Qilin Zhang, Junsong Yuan, and Gang Hua, “Two-Stream Consensus Network for Weakly-Supervised Temporal Action Localization”, in *European Conference on Computer Vision (ECCV)*, pp. 37-54, 2020. [Paper]
18. Yuanhao Zhai, Le Wang, Ziyi Liu, Qilin Zhang, Gang Hua, and Nanning Zheng, “Action Coherence Network for Weakly Supervised Temporal Action Localization”, in *IEEE International Conference on Image Processing (ICIP)*, pp. 918-922, 2019. [Paper]

HONORS & AWARDS

Runner Up , CVPR’2021 HACS Weakly-Supervised Temporal Action Localization Challenge	2021
Ji Ke Prize (8-th/763 teams), Alibaba Tianchi Temporal Action Localization Competition	2020
Outstanding Graduate (Top 15%), Xi’an Jiaotong University	2020
Excellent Student (Top 15%), Xi’an Jiaotong University	2017, 2018, 2019
Peng Kang Scholarship (Top 3%), Xi’an Jiaotong University	2017

PROFESSIONAL SERVICES

Conference Reviewer

• IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)	2022, 2024, 2025
• IEEE International Conference on Computer Vision (ICCV)	2023, 2025
• European Conf. on Computer Vision (ECCV)	2022, 2024
• Neural Information Processing Systems (NeurIPS)	2024, 2025
• International Conference on Learning Representations (ICLR)	2025
• Annual AAAI Conference on Artificial Intelligence (AAAI)	2025
• ACM Multimedia (MM)	2024, 2025
• Asian Conference on Computer Vision (ACCV)	2024
• IEEE International Conf. on Automatic Face and Gesture Recognition (FG)	2021, 2023, 2025

Journal Reviewer

- IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)
- International Journal of Computer Vision (IJCV)
- IEEE Trans. on Image Processing (TIP)
- IEEE Trans. on Circuits and Systems for Video Technology (TCSVT)
- IEEE Trans. on Multimedia (TMM)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- IEEE Intelligent Systems
- International Journal of Machine Learning and Cybernetics
- Journal of Visual Communication and Image Representation

- Machine Vision and Applications
- The Journal of Supercomputing