

Wenhao Chai

✉ wenhao.chai@princeton.edu
🏠 [homepage](#)
🔍 [google scholar](#)
🐙 github.com/rese1f
🐦 x.com/wenhaocha1
🌐 linkedin.com/in/wenhao-chai

(206) 349-8459
35 Olden St.
Princeton, NJ
United States, 08544
Dept. of Computer Science
Princeton University

Research Overview

My research spans a wide range of topics in computer vision and deep learning.

2023-Present	Large Multi-Modal Models Video Understanding, Embodied Agent
2023-Present	Generative Models Video, Image, 3D, City Layout, Human Motion
2022-2023	Human Pose and Motion 3D Pose Estimation, Motion, Tracking

Education

Ph.D. CS 2025-2029 (exp.)	Princeton University Advisor: Zhuang Liu Incoming
M.S. EE 2023-2025	University of Washington Advisor: Jenq-Neng Hwang Thesis: Large Multi-modal Models for Video Captioning
Visiting Spring/Summer 2022	University of Illinois Urbana-Champaign National Center for Supercomputing Applications
B.S. 2019-2023	Zhejiang University Advisor: Gaoang Wang
High School 2016-2019	Hangzhou No. 2 High School Hangzhou, Zhejiang

Employment

Research Intern Summer 2024	Pika Labs Working on Video Understanding Mentor: Christopher D. Manning
Research Assistant 2023-2024	University of Washington Information Processing Lab PI: Jenq-Neng Hwang
Research Intern Spring/Summer 2023	Microsoft Research Asia Working on Video Generation Mentor: Xun Guo

Selected Publications

The * sign denotes equal contribution. The † sign denotes project lead.

Peer-Reviewed Papers

- C14 Wenhao Chai, Enxin Song, Yilun Du, Chenlin Meng, Vashisht Madhavan, Omer Bar-Tal, Jenq-Neng Hwang, Saining Xie, and Christopher D. Manning. **AuroraCap: Efficient, Performant Video Detailed Captioning and a New Benchmark**. In *International Conference on Learning Representations (ICLR)*. 2025.
- C13 Ruizhe Chen*, Xiaotian Zhang*, Meng Luo*, Wenhao Chai*, and Zuozhu Liu. **PAD: Personalized Alignment at Decoding-time**. In *International Conference on Learning Representations (ICLR)*. 2025.
- C12 Hsiang-Wei Huang, Fu-Chen Chen, Wenhao Chai, Che-Chun Su, Lu Xia, Sanghun Jung, Cheng-Yen Yang, Jenq-Neng Hwang, Min Sun, and Cheng-Hao Kuo. **Zero-shot 3D Question Answering via Voxel-based Dynamic Token Compression**. In *Computer Vision and Pattern Recognition (CVPR)*. 2025.
- C11 Jialuo Li, Wenhao Chai, Xingyu Fu, Haiyang Xu, and Saining Xie. **Science-T2I: Addressing Scientific Illusions in Image Synthesis**. In *Computer Vision and Pattern Recognition (CVPR)*. 2025.
- C10 Hou-I Liu, Christine Wu, Jen-Hao Cheng, Wenhao Chai, Shian-yun Wang, Gaowen Liu, Hugo Latapie, Jih-Ciang Wu, Jenq-Neng Hwang, Hong-Han Shuai, and Wen-Huang Cheng. **MonoTAKD: Teaching Assistant Knowledge Distillation for Monocular 3D Object Detection**. In *Computer Vision and Pattern Recognition (CVPR)*. 2025.
- C9 Tian Ye, Sixiang Chen, Haoyu Chen, Wenhao Chai, Jingjing Ren, Zhaohu Xing, Wenxue Li, and Lei Zhu. **PromptHaze: Prompting Real-world Dehazing via Depth Anything Model**. In *Association for the Advancement of Artificial Intelligence (AAAI)*. 2025.
- C8 Yunlong Lin, Tian Ye, Sixiang Chen, Zhenqi Fu, Yingying Wang, Wenhao Chai, Zhaohu Xing, Lei Zhu, and Xinghao Ding. **AGLLDiff: Guiding Diffusion Models Towards Unsupervised Training-free Real-world Low-light Image Enhancement**. In *Association for the Advancement of Artificial Intelligence (AAAI)*. 2025.
- J3 Zhonghan Zhao, Wenhao Chai, Shengyu Hao, Wenhao Hu, Guan hong Wang, Shidong Cao, Mingli Song, Jenq-Neng Hwang, and Gaoang Wang. **A Survey of Deep Learning in Sports Applications: Perception, Comprehension, and Decision**. In *IEEE Transactions on Visualization and Computer Graphics (TVCG)*. 2025.
- J2 Shidong Cao, Zhonghan Zhao, Shengyu Hao, Wenhao Chai, Jenq-Neng Hwang, Hongwei Wang, and Gaoang Wang. **Efficient Transfer from Image-based Large Multimodal Models to Video Tasks**. In *IEEE Transactions on Multimedia (TMM)*. 2025.
- C7 Zhonghan Zhao*, Wenhao Chai*†, Xuan Wang*, Boyi Li, Shengyu Hao, Shidong Cao, Tian Ye, Jenq-Neng Hwang, and Gaoang Wang. **See and Think: Embodied Agent in Virtual Environment**. In *European Conference on Computer Vision (ECCV)*. 2024.
- C6 Yuan-Hao Ho, Jen-Hao Cheng, Sheng Yao Kuan, Zhongyu Jiang, Wenhao Chai, Hsiang-Wei Huang, Chih-Lung Lin, and Jenq-Neng Hwang. **RT-Pose: A 4D Radar Tensor-based 3D Human Pose Estimation and Localization Benchmark**. In *European Conference on Computer Vision (ECCV)*. 2024.
- C5 Enxin Song*, Wenhao Chai*†, Guan hong Wang, Yucheng Zhang, Haoyang Zhou, Feiyang Wu, Haozhe Chi, Yucheng Zhang, Haoyang Zhou, Feiyang Wu, Haozhe Chi, Xun Guo, Tian Ye, Yanting Zhang, Yan Lu, Jenq-Neng Hwang, and Gaoang Wang. **MovieChat: From Dense Token to Sparse Memory for Long Video Understanding**. In *Computer Vision and Pattern Recognition (CVPR)*. 2024.
- C4 Tian Ye, Sixiang Chen, Wenhao Chai, Zhaohu Xing, Jing Qin, Ge Lin, and Lei Zhu. **Learning Diffusion Texture Priors for Image Restoration**. In *Computer Vision and Pattern Recognition (CVPR) High-light*. 2024.

- C3 Meiqi Sun*, Zhonghan Zhao*, Wenhao Chai^{*†}, Hanjun Luo, Shidong Cao, Yanting Zhang, Jenq-Neng Hwang, and Gaoang Wang. **UniAP: Towards Universal Animal Perception in Vision via Few-shot Learning**. In *Association for the Advancement of Artificial Intelligence (AAAI)*. 2024.
- J1 Shidong Cao*, Wenhao Chai^{*}, Shengyu Hao, Yanting Zhang, Hangyue Chen, and Gaoang Wang. **Diff-fashion: Reference-based Fashion Design with Structure-aware Transfer by Diffusion Models**. In *IEEE Transactions on Multimedia (TMM)*. 2023.
- C2 Wenhao Chai, Zhongyu Jiang, Jenq-Neng Hwang, and Gaoang Wang. **Global Adaptation Meets Local Generalization: Unsupervised Domain Adaptation for 3D Human Pose Estimation**. In *International Conference on Computer Vision (ICCV)*. 2023.
- C1 Wenhao Chai, Xun Guo, Gaoang Wang, and Yan Lu. **StableVideo: Text-driven Consistency-aware Diffusion Video Editing**. In *International Conference on Computer Vision (ICCV)*. 2023.

Workshop and Technical Reports

- W7 Weili Xu, Enxin Song, Wenhao Chai[†], Tian Ye, and Gaoang Wang. **Bringing RNNs Back to Efficient Open-Ended Video Understanding**. In *Computer Vision and Pattern Recognition (CVPR) Workshop @ Efficient Large Vision Models*. 2025.
- W6 Jie Deng, Wenhao Chai, Jianshu Guo, Qixuan Huang, Wenhao Hu, Jenq-Neng Hwang, Gaoang Wang. **City-Gen: Infinite and Controllable 3D City Layout Generation**. In *Computer Vision and Pattern Recognition (CVPR) Workshop @ Urban Scene Modeling*. 2025.
- W5 Zhonghan Zhao, Wenhao Chai[†], Xuan Wang, Ke Ma, Kewei Chen, Dongxu Guo, Tian Ye, Yanting Zhang, Hongwei Wang, and Gaoang Wang. **STEVE Series: Step-by-Step Construction of Agent Systems in Minecraft**. In *Computer Vision and Pattern Recognition (CVPR) Workshop @ Embodied AI*. 2024.
- W4 Zhonghan Zhao*, Kewei Chen*, Dongxu Guo*, Wenhao Chai[†], Tian Ye, Yanting Zhang, and Gaoang Wang. **Hierarchical Auto-Organizing System for Open-Ended Multi-Agent Navigation**. In *International Conference on Learning Representations (ICLR) Workshop @ LLM Agents*. 2024.
- W3 Florin-Alexandru Vasluiianu *et al.*. **NTIRE 2024 Image Shadow Removal Challenge Report**. In *Computer Vision and Pattern Recognition (CVPR) Workshop @ NTIRE*. 2024.
- W2 Yichen Xu, Zihan Xu, Wenhao Chai[†], Zhonghan Zhao, Enxin Song, and Gaoang Wang. **Devil in the Number: Towards Robust Multi-modality Data Filter**. In *International Conference on Computer Vision (ICCV) Workshop @ DataComp*. 2023.
- W1 Shidong Cao*, Wenhao Chai^{*}, Shengyu Hao, and Gaoang Wang. **Image Reference-guided Fashion Design with Structure-aware Transfer by Diffusion Models**. In *Computer Vision and Pattern Recognition (CVPR) Workshop @ Computer Vision for Fashion, Art, and Design*. 2023.

Preprints

- P6 Dongping Li*, Tielong Cai*, Tianci Tang*, Wenhao Chai, Katherine Rose Driggs-Campbell, and Gaoang Wang. **EMMOE: A Comprehensive Benchmark for Embodied Mobile Manipulation in Open Environments**. In *arXiv Preprint*. 2025.
- P5 Ruizhe Chen, Wenhao Chai, Zhifei Yang, Xiaotian Zhang, Joey Tianyi Zhou, Tony Quek, Soujanya Poria, and Zuozhu Liu. **DiffPO: Diffusion-styled Preference Optimization for Efficient Inference-Time Alignment of Large Language Models**. In *arXiv Preprint*. 2025.
- P4 Liang Chen, Shuai Bai, Wenhao Chai, Weichu Xie, Haozhe Zhao, Leon Vinci, Junyang Lin, and Baobao Chang. **Multimodal Representation Alignment for Image Generation: Text-Image Interleaved Control Is Easier Than You Think**. In *arXiv Preprint*. 2025.
- P3 Cheng-Yen Yang, Hsiang-Wei Huang, Wenhao Chai, Zhongyu Jiang, and Jenq-Neng Hwang. **SAMURAI: Adapting Segment Anything Model for Zero-shot Visual Tracking with Motion-aware Memory**. In *arXiv Preprint*. 2024.

P2 Zhonghan Zhao, Ke Ma, [Wenhao Chai](#), Xuan Wang, Kewei Chen, Dongxu Guo, Yanting Zhang, Hongwei Wang, and Gaoang Wang. [Do We Really Need a Complex Agent System? Distill Embodied Agent into a Single Model](#). In *arXiv Preprint*. 2024.

P1 Enxin Song*, [Wenhao Chai](#)*†, Tian Ye, Jenq-Neng Hwang, Xi Li, and Gaoang Wang. [MovieChat+: Question-aware Sparse Memory for Long Video Question Answering](#). In *arXiv Preprint*. 2024.

Invited Talks

Step-by-Step Construction of Agent Systems in Minecraft
CAMEL-AI AgentX Seminar Host: Guohao Li

Virtual
Apr 2024

Towards Universal Animal Perception in Vision
Workshop on Imageomics at AAAI 2024

Vancouver, Canada
Feb 2024

University of Washington Information Processing Lab Seminar
Linear Attention and State Space Models
What is the Intrinsic Dimension of Your Data?
Bridging the Parallel Decoding of LLMs with the Diffusion Process
DPO and RLHF for Large Language Model Post-training
Vision Representation Learning from Synthetic Data
From Large Language Models to Large Multi-Modal Models

Seattle, WA
Apr 2025
Jan 2025
Oct 2024
Apr 2024
Jan 2024
Nov 2023

Research Mentoring

Weili Xu B.S. at University of Illinois Urbana-Champaign & Zhejiang University Topic: Video Understanding Bringing RNNs Back to Efficient Open-Ended Video Understanding	Oct 2024- Present CVPRW 2025
Ruizhe Chen Ph.D. at Zhejiang University ⇒ Intern at ByteDance Topic: LLM Alignment PAD: Personalized Alignment at Decoding-Time DiffPO: Diffusion-styled Preference Optimization for Inference Time Alignment	Jul 2024- Present ICLR 2025 In Submission
Hsiang-Wei Huang Ph.D. at University of Washington ⇒ Intern at Amazon Topic: Spatial Understanding Zero-shot 3D Question Answering via Voxel-based Dynamic Token Compression ToSA: Token Merging with Spatial Awareness 3D Visual Grounding with Reasoning LLM	Jul 2024- Present CVPR 2025 In Submission In Submission
Jie Deng Research Assistant at Zhejiang University Topic: Generative Models in 3D Citygen: Infinite and Controllable 3D City Layout Generation Citycraft: A Real Crafter for 3d City Generation	Jul 2023- Present CVPRW 2025 In Submission
Enxin Song M.S. at Zhejiang University ⇒ Visiting Student at UCSD Topic: Video Understanding MovieChat: From Dense Token to Sparse Memory for Long Video Understanding MovieChat+: Question-aware Sparse Memory for Long Video Question Answering Video-MMLU: A Massive Multi-Discipline Lecture Understanding Benchmark	Jul 2023- Present CVPR 2024 In Submission In Submission
Zhonghan Zhao Ph.D. at Zhejiang University ⇒ Intern at Shanghai AI Lab Topic: Embodied Agent See and Think: Embodied Agent in Virtual Environment Hierarchical Auto-Organizing System for Open-Ended Multi-Agent Navigation Steve Series: Step-by-step Construction of Agent Systems in Minecraft Do We Really Need a Complex Agent System? Distill Agent into a Single Model	Jul 2023- Jul 2024 ECCV 2024 ICLRW 2024 CVPRW 2024 In Submission
Meiqi Sun M.S. at Zhejiang University ⇒ Alibaba Group Topic: Pose Estimation UniAP: Towards Universal Animal Perception in Vision via Few-shot Learning	Jul 2023- Feb 2024 AAAI 2024

Professional Service

Workshop Organization

Multimodal Video Agent Workshop on Computer Vision and Pattern Recognition (CVPR) 2025	Nashville, TN June 2025
Long-form Video Understanding Towards Multimodal AI Assistant and Copilot Workshop on Computer Vision and Pattern Recognition (CVPR) 2024	Seattle, WA June 2024

Conference and Journal Refereeing

Neural Information Processing Systems (NeurIPS)	2024-2025
International Conference in Learning Representations (ICLR)	2025
International Conference in Machine Learning (ICML)	2024-2025
Computer Vision and Pattern Recognition (CVPR)	2024-2025
International Conference on Computer Vision (ICCV)	2025
European Conference on Computer Vision (ECCV)	2024
Annual Meeting of the Association for Computational Linguistics (ACL)	2025
Conference on Language Modeling (COLM)	2025
Association for the Advancement of Artificial Intelligence (AAAI)	2025
International Conference on Artificial Intelligence and Statistics (AISTATS)	2025
Winter Conference on Applications of Computer Vision (WACV)	2025
ACM International Conference on Multimedia (ACM MM)	2024
Pattern Recognition and Computer Vision (PRCV)	2023
International Journal of Computer Vision (IJCV)	
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	

Community Service

Constructive Feedback for alphaXiv and Paper Copilot	2024-2025
Host Discord Server for arXiv Daily Paper Sharing.	2023-Present
Research Experience Sharing Session for Undergraduates at Zhejiang University	2023
Co-Director of the Publicity Dept. of the Student Union at Zhejiang University	2019-2022