Wenhao Chai

University of Washington
185 E Stevens Way NE
Seattle WA 98195 USA

wchai@uw.edu
https://rese1f.github.io/

Research Overview

My current research focuses on developing embodied AI agents that interact with the physical world, with a particular emphasis on leveraging video understanding as a core perception tool. A key challenge in video understanding using Large Multi-modal Models (LMMs) is the efficiency during both training and inference. My work addresses this by proposing token merging, where visual tokens are significantly reduced with minimal performance drop. My work has illustrated that by introducing a long-short term memory mechanism, we can extend pre-trained video LMMs to understand videos spanning several hours without further fine-tuning. My work has further shown how to step-by-step build of agent systems in Minecraft.

Research Areas: Computer Vision, Embodied Agent, Generative AI

Education

M.S. EE 2023-2024	University of Washington (UW) Advisors: Jeng-Neng Hwang Thesis: *Large Multi-modal Model for Video Understadning*
B.S. 2019-2023	Zhejiang University (ZJU) GPA: 3.73 / 4.00

Employment

Research Intern Summer 2024	Pika Lab Research Intern Working on Video Captioning
Research Intern Spring/Summer 2023	Microsoft Research Asia Research Intern Working on Video Editing

Selected Publications

The * sign denotes equal contribution.

Peer-Reviewed Papers

C7 Zhao, Zhonghan*, Chai, Wenhao*, Xuan Wang, Li Boyi, Shengyu Hao, Shidong Cao, Tian Ye, Jenq-Neng Hwang, and Gaoang Wang. "See and think: Embodied agent in virtual environment." *European Conference on Computer Vision (ECCV)*, 2024.

C6 Ho, Yuan-Hao, Jen-Hao Cheng, Sheng Yao Kuan, Zhongyu Jiang, Chai, Wenhao, Hsiang-Wei Huang, Chih-Lung Lin, and Jenq-Neng Hwang. "RT-Pose: A 4D Radar Tensor-based 3D Human Pose Estimation and Localization Benchmark." *European Conference on Computer Vision (ECCV)*, 2024.

C5 Song, Enxin*, Chai, Wenhao*, Guanhong Wang, Yucheng Zhang, Haoyang Zhou, Feiyang Wu, Haozhe Chi et al. "Moviechat: From dense token to sparse memory for long video understanding." *Computer Vision and Pattern Recognition (CVPR)*, 2024.

C4 Ye, Tian, Sixiang Chen, Chai, Wenhao, Zhaohu Xing, Jing Qin, Ge Lin, and Lei Zhu. "Learning Diffusion Texture Priors for Image Restoration." *Computer Vision and Pattern Recognition (CVPR)*, 2024.

C3 Sun, Meiqi*, Zhonghan Zhao*, Chai, Wenhao*, Hanjun Luo, Shidong Cao, Yanting Zhang, Jenq-Neng Hwang, and Gaoang Wang. "Uniap: Towards universal animal perception in vision via few-shot learning." *Association for the Advancement of Artificial Intelligence (AAAI)*, 2024.

C2 Chai, Wenhao, Xun Guo, Gaoang Wang, and Yan Lu. "Stablevideo: Text-driven consistency-aware diffusion video editing." *International Conference on Computer Vision (ICCV)*, 2023.

C1 Chai, Wenhao, Zhongyu Jiang, Jenq-Neng Hwang, and Gaoang Wang. "Global adaptation meets local generalization: Unsupervised domain adaptation for 3d human pose estimation." *International Conference on Computer Vision (ICCV)*, 2023.

J1 Cao, Shidong*, Chai, Wenhao*, Shengyu Hao, Yanting Zhang, Hangyue Chen, and Gaoang Wang. "Difffashion: Reference-based fashion design with structure-aware transfer by diffusion models." IEEE Transactions on Multimedia (TMM), 2023.

Preprints

P1 Chai, Wenhao, Enxin Song, Yilun Du, Chenlin Meng, Vashisht Madhavan, Omer Bar-Tal, Jeng-Neng Hwang, Saining Xie, and Christopher D. Manning. "AuroraCap: Efficient, Performant Video Detailed Captioning and a New Benchmark." arXiv, 2024.

Invited Talks

AAAI Workshop on Imageomics Towards Universal Animal Perception in Vision	Feb 2024 Vancouver, Canada

Professional Service

Conference and Journal Refereeing

Neural Information Processing Systems (NeurIPS)	2024
International Conference in Learning Representations (ICLR)	2025
International Conference in Machine Learning (ICML)	2024
Computer Vision and Pattern Recognition (CVPR)	2024
European Conference on Computer Vision (ECCV)	2024
Winter Conference on Applications of Computer Vision (WACV)	2025
ACM Multimedia (ACM MM)	2024
International Conference on Artificial Intelligence and Statistics (AISTATS)	2025
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	2023

Workshop Organization

Workshop on Long-form Video Understanding at CVPR 2024	July 2024