

# YUJUN CAI

◊ Email: vanora.caiyj@gmail.com

[Homepage](#) || [Google Scholar](#) || [Linkedin](#)

## WORK

<b>Lecturer at University of Queensland (QS Word Ranking #42)</b>	Nov. 2024 - present
EECS School. Research and Teaching on Multi-modal Large Language Models.	
<b>Senior Research Scientist at Meta Reality Lab (Seattle)</b>	Dec. 2023 - Nov. 2024
Research and System Developments on Embodied AI and Multi-modal Large Language Models.	
<b>Research Scientist at Meta Reality Lab (Seattle)</b>	Mar. 2022 - Dec. 2023
Research and System Developments on Embodied AI, Virtual Reality, Model Acceleration.	

## RESEARCH INTERESTS

Computer Vision, Multi-modal Large Language Models.

## EDUCATION

<b>Doctor of Philosophy in Computer Science</b>	Aug. 2017 - Dec. 2021
Nanyang Technological University, Singapore	
<b>Bachelor of Science in Information Technology</b>	Aug. 2013 - Jun. 2017
Southeast University, China, GPA: 4.341/4.8, Rank: 1/229	

## RESEARCH PUBLICATION

- [1] Jianing Chen, Zehao Li, Yujun Cai, Hao Jiang, Chengxuan Qian, Juyuan Kang, Shuqin Gao, Honglong Zhao, Tianlu Mao, Yucheng Zhang. HAIF-GS: Hierarchical and Induced Flow-Guided Gaussian Splatting for Dynamic Scene, *NeurIPS 2025*.
- [2] Shuyang Hao, Yiwei Wang, Bryan Hooi, Jun Liu, Muhan Chen, Zi Huang, Yujun Cai. Making every step effective: Jailbreaking large vision-language models through hierarchical kv equalization, *EMNLP 2025*.
- [3] Haonan Ge, Yiwei Wang, Ming-Hsuan Yang, Yujun Cai. MRFD: Multi-Region Fusion Decoding with Self-Consistency for Mitigating Hallucinations in LLMs, *EMNLP 2025*.
- [4] Zhen Xiong, Yujun Cai, Zhecheng Li, Yiwei Wang. Mapping the Minds of LLMs: A Graph-Based Analysis of Reasoning LLMs, *EMNLP 2025*.
- [5] Hang Wu, Hongkai Chen, Yujun Cai, Chang Liu, Qingwen Ye, Ming-Hsuan Yang, Yiwei Wang. DiMo-GUI: Advancing Test-time Scaling in GUI Grounding via Modality-Aware Visual Reasoning, *EMNLP 2025*.
- [6] Sifan Li, Yujun Cai, Yiwei Wang. SemVink: Advancing VLMs' Semantic Understanding of Optical Illusions via Visual Global Thinking, *EMNLP 2025*.
- [7] Xingjian Tao, Yiwei Wang, Yujun Cai, Zhicheng Yang, Jing Tang. Understanding GUI Agent Localization Biases through Logit Sharpness, *EMNLP 2025*.
- [8] Honghao Fu, Junlong Ren, Qi Chai, Deheng Ye, Yujun Cai, Hao Wang. VistaWise: Building Cost-Effective Agent with Cross-Modal Knowledge Graph for Minecraft, *EMNLP 2025*.
- [9] Shuyang Hao, Bryan Hooi, Jun Liu, Kai-Wei Chang, Zi Huang, Yujun Cai. Exploring Visual Vulnerabilities via Multi-Loss Adversarial Search for Jailbreaking Vision-Language Models, *CVPR 2025*.

- [10] Muchen Li, Sammy Christen, Chengde Wan, Yujun Cai, Renjie Liao, Leonid Sigal, Shugao Ma. LatentHOI: On the Generalizable Hand Object Motion Generation with Latent Hand Diffusion, *CVPR 2025*.
- [11] Yihong Luo, Tianyang Hu, Jiacheng Sun, Yujun Cai, Jing Tang. Learning Few-Step Diffusion Models by Trajectory Distribution Matching, *ICCV 2025*.
- [12] Zhecheng Li, Yiwei Wang, Bryan Hooi, Yujun Cai, Zhen Xiong, Nanyun Peng, Kai-Wei Chang. Vulnerability of LLMs to Vertically Aligned Text Manipulations, *ACL 2025*.
- [13] Zhecheng Li, Yiwei Wang, Bryan Hooi, Yujun Cai, Nanyun Peng, Kai-Wei Chang. DRS: Deep Question Reformulation With Structured Output, *ACL 2025*.
- [14] Zhaochen Wang, Bryan Hooi, Yiwei Wang, Ming-Hsuan Yang, Zi Huang, Yujun Cai. Text Speaks Louder than Vision: ASCII Art Reveals Textual Biases in Vision-Language Models, *COLM 2025*.
- [15] Zhecheng Li, Guoxian Song, Yujun Cai, Zhen Xiong, Junsong Yuan, Yiwei Wang. Texture or Semantics? Vision-Language Models Get Lost in Font Recognition, *COLM 2025*.
- [16] Chunxue Xu, Yiwei Wang, Bryan Hooi, Yujun Cai, Songze Li. How does Watermarking Affect Visual Language Models in Document Understanding?, *COLM 2025*.
- [17] Cheng Wang, Yiwei Wang, Yujun Cai, Bryan Hooi. Tricking Retrievers with Influential Tokens: An Efficient Black-Box Corpus Poisoning Attack. *NAACL 2025*.
- [18] Cheng Wang, Yiwei Wang, Bryan Hooi, Yujun Cai, Nanyun Peng, Kai-Wei Chang. Con-ReCall: Detecting Pre-training Data in LLMs via Contrastive Decoding, *COLING 2025*.
- [19] Haoxuan Qu, Zhuoling Li, Hossein Rahmani, Yujun Cai, Jun Liu. DisC-GS: Discontinuity-aware Gaussian Splatting, *NeurIPS 2024*.
- [20] Sasha Salter, Richard Warren, Collin Schlager, Adrian Spurr, Shangchen Han, Rohin Bhasin, Yujun Cai, Peter Walkington, Anuoluwapo Bolarinwa, Robert Wang, others. emg2pose: A Large and Diverse Benchmark for Surface Electromyographic Hand Pose Estimation, *NeurIPS Datasets and Benchmarks 2024*.
- [21] Yihong Luo, Siya Qiu, Xingjian Tao, Yujun Cai, Jing Tang. Energy-Calibrated VAE with Test Time Free Lunch, *ECCV 2024*.
- [22] Haoxuan Qu, Yujun Cai, Jun Liu. Llms are good action recognizers, *CVPR 2024*.
- [23] Li Xu, Haoxuan Qu, Yujun Cai, Jun Liu. 6d-diff: A keypoint diffusion framework for 6d object pose estimation, *CVPR 2024*.
- [24] Kenrick Kin, Chengde Wan, Ken Koh, Andrei Marin, Necati Cihan Camgoz, Yubo Zhang, Yujun Cai, Fedor Kovalev, Moshe Ben-Zacharia, Shannon Hoople, others. STMG: A Machine Learning Microgesture Recognition System for Supporting Thumb-Based VR/AR Input, *CHI 2024*.
- [25] Yiwei Wang\*, Yujun Cai\*, Muhan Chen, Yuxuan Liang, Bryan Hooi. Primacy effect of chatgpt, *EMNLP 2023*.
- [26] Julian Tanke, Linguang Zhang, Amy Zhao, Chengcheng Tang, Yujun Cai, Lezi Wang, Po-Chen Wu, Juergen Gall, Cem Keskin. Social diffusion: Long-term multiple human motion anticipation, *ICCV 2023*.
- [27] Haoxuan Qu, Yujun Cai, Lin Geng Foo, Ajay Kumar, Jun Liu. A Characteristic Function-Based Method for Bottom-Up Human Pose Estimation, *CVPR 2023*.
- [28] Yiwei Wang, Bryan Hooi, Fei Wang, Yujun Cai, Yuxuan Liang, Wenxuan Zhou, Jing Tang, Manjuan Duan, Muhan Chen. How Fragile is Relation Extraction under Entity Replacements?, *CONLL 2023*.

- [29] Chi Zhang, Yujun Cai, Guosheng Lin, Chunhua Shen. Deepemd: Differentiable earth mover’s distance for few-shot learning, *TPAMI* 2022.
- [30] Haoxuan Qu, Li Xu, Yujun Cai, Lin Geng Foo, Jun Liu. Heatmap distribution matching for human pose estimation, *NeurIPS* 2022.
- [31] Mingfei Chen, Jianfeng Zhang, Xiangyu Xu, Lijuan Liu, Yujun Cai, Jiashi Feng, Shuicheng Yan. Geometry-guided progressive nerf for generalizable and efficient neural human rendering, *ECCV* 2022.
- [32] Yiwei Wang, Muhan Chen, Wenxuan Zhou, Yujun Cai, Yuxuan Liang, Bryan Hooi. Graphcache: Message passing as caching for sentence-level relation extraction, *NAACL* 2022.
- [33] Yiwei Wang, Muhan Chen, Wenxuan Zhou, Yujun Cai, Yuxuan Liang, Dayiheng Liu, Baosong Yang, Juncheng Liu, Bryan Hooi. Should we rely on entity mentions for relation extraction? debiasing relation extraction with counterfactual analysis, *NAACL* 2022.
- [34] Yiwei Wang, Yujun Cai, Yuxuan Liang, Henghui Ding, Changhu Wang, Bryan Hooi. Time-Aware Neighbor Sampling on Temporal Graphs, *2022 International Joint Conference on Neural Networks (IJCNN)* 2022.
- [35] Ziyin Xin, Yujun Cai, Louis T Dang, Hannah MS Burke, Jerico Revote, Natalie Charitakis, Denis Bienroth, Hieu T Nim, Yuan-Fang Li, Mirana Ramialison. MonaGO: a novel gene ontology enrichment analysis visualisation system, *BMC bioinformatics* 2022.
- [36] Shangchen Han, Po-chen Wu, Yubo Zhang, Beibei Liu, Linguang Zhang, Zheng Wang, Weiguang Si, Peizhao Zhang, Yujun Cai, Tomas Hodan, others. UmeTrack: Unified multi-view end-to-end hand tracking for VR, *SIGGRAPH Asia 2022 conference papers* 2022.
- [37] Yiwei Wang, Yujun Cai, Yuxuan Liang, Henghui Ding, Changhu Wang, Siddharth Bhatia, Bryan Hooi. Adaptive data augmentation on temporal graphs, *NeurIPS* 2021.
- [38] Jianfeng Zhang, Yujun Cai, Shuicheng Yan, Jiashi Feng, others. Direct multi-view multi-person 3d pose estimation, *NeurIPS* 2021.
- [39] Yujun Cai, Yiwei Wang, Yiheng Zhu, Tat-Jen Cham, Jianfei Cai, Junsong Yuan, Jun Liu, Chuanxia Zheng, Sijie Yan, Henghui Ding, others. A unified 3d human motion synthesis model via conditional variational auto-encoder, *ICCV* 2021.
- [40] Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Bryan Hooi. Progressive supervision for node classification, *ECML* 2021.
- [41] Yiwei Wang, Yujun Cai, Yuxuan Liang, Henghui Ding, Changhu Wang, Bryan Hooi. Time-aware neighbor sampling for temporal graph networks, *The 2022 International Joint Conference on Neural Networks (IJCNN)* 2021.
- [42] Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Bryan Hooi. Curgraph: Curriculum learning for graph classification, *WWW* 2021.
- [43] Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Bryan Hooi. Mixup for node and graph classification, *WWW* 2021.
- [44] Yujun Cai, Liuqiao Ge, Jianfei Cai, Nadia Magnenat Thalmann, Junsong Yuan. 3D hand pose estimation using synthetic data and weakly labeled RGB images, *TPAMI* 2020.
- [45] Chi Zhang, Yujun Cai, Guosheng Lin, Chunhua Shen. Deepemd: Few-shot image classification with differentiable earth mover’s distance and structured classifiers, *CVPR* 2020.
- [46] Yiwei Wang, Wei Wang, Yujun Cai, Bryan Hooi, Beng Chin Ooi. Detecting implementation bugs in graph convolutional network based node classifiers, *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)* 2020.

- [47] Yujun Cai, Lin Huang, Yiwei Wang, Tat-Jen Cham, Jianfei Cai, Junsong Yuan, Jun Liu, Xu Yang, Yiheng Zhu, Xiaohui Shen, others. Learning progressive joint propagation for human motion prediction, *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16 2020*.
- [48] Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Juncheng Liu, Bryan Hooi. Nodeaug: Semi-supervised node classification with data augmentation, *KDD 2020*.
- [49] Yujun Cai, Liuhan Ge, Jun Liu, Jianfei Cai, Tat-Jen Cham, Junsong Yuan, Nadia Magnenat Thalmann. Exploiting spatial-temporal relationships for 3d pose estimation via graph convolutional networks, *ICCV 2019*.
- [50] Yujun Cai, Liuhan Ge, Jianfei Cai, Junsong Yuan. Weakly-supervised 3d hand pose estimation from monocular rgb images, *ECCV 2018*.
- [51] Liuhan Ge, Yujun Cai, Junwu Weng, Junsong Yuan. Hand pointnet: 3d hand pose estimation using point sets, *CVPR 2018*.

## TEACHING

---

Instructor, ELEC4630: Computer Vision and Deep Learning	UQ, 2025
Instructor, COMP7110: Introduction to Software Innovation	UQ, 2025
Teaching Assistant, CE 7491: Digital Image Processing	NTU, 2018-2019
Teaching Assistant, DM6190: Recent Advance on Image Segmentation and Its Applications	NTU, 2019
Teaching Assistant, Engineering Mathematics II	NTU, 2018

## SERVING AS ASSOCIATE EDITOR

---

Neurocomputing
Pattern Recognition

## SERVING AS AREA CHAIR

---

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	2026
Association for the Advancement of Artificial Intelligence (AAAI)	2025
Conference on Neural Information Processing Systems (NeurIPS)	2025
International Conference on Computer Vision (ICCV)	2025
International Conference on Learning Representations (ICLR)	2025
International Conference on Machine Learning (ICML)	2025
ACM Multimedia	2025
The Web Conference (WWW)	2025
The Conference on Information and Knowledge Management (CIKM)	2025
IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)	2024, 2025
Conference on Language Modeling (COLM)	2025
International Joint Conference on Neural Networks (IJCNN)	2025

## SERVING AS REVIEWER

---

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	2018-2025
International Conference on Computer Vision (ICCV)	2019-2024
European Computer Vision Association (ECCV)	2018-2024
Conference on Empirical Methods in Natural Language Processing (EMNLP)	2024, 2025
Annual Meeting of the Association for Computational Linguistics (ACL)	2024, 2025
North American Chapter of the Association for Computational Linguistics (NAACL)	2024, 2025
International Conference on Learning Representations (ICLR)	2023, 2024
Conference on Neural Information Processing Systems (NeurIPS)	2022-2024
The AAAI Conference on Artificial Intelligence	2021-2024
International Joint Conference on Artificial Intelligence	2019-2024
IEEE International Conference on Multimedia and Expo	2022, 2023, 2024
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)	
International Journal of Computer Vision (IJCV)	
IEEE Transactions on Image Processing (TIP)	
IEEE Transactions on Multimedia (TMM)	
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	
ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)	