Zirui Zhu

zirui@comp.nus.edu.sg
zirui-zhu.com

EDUCATION

National University of Singapore Singapore
PhD. in *Computer Science* *Aug 2022 – Present*
Advisor: Prof. Yang You

Tsinghua University Beijing, China
B.Eng. in *Electronics Engineering* *Sep 2018 – Jun 2022*
Advisor: Prof. Yong Li

- ✓ Scholarship for Overall Excellence (Top 10%) *Oct 2021*
- ✓ Scholarship for Excellent Academic Performances (Top 10%) *Oct 2018*

RESEARCH INTERESTS

My research interests include:

- LLM Optimization and Post-training
- Video Understanding
- Recommendation Systems

INDUSTRY EXPERIENCE

Cost-aware Keyframe Selection for Long-video Understanding
Research Intern, TikTok Pte. Ltd **Singapore**
Mentor: Dr. Zhenheng Yang and *Dr. Kun Xu* Apr *2025* – Mar *2026*
Highlights: Proposed a cost-aware keyframe selection method that improves long-video VQA accuracy by 11.9% while using <2% of frames.

Large-batch Optimization for LLM Training
Research Intern, Beijing Lingyi Wanwu Information Technology Co., Ltd. (01.AI) **Beijing, China**
Mentor: Dr. Zhenyu Gu *May 2024 – Jul 2024*
Highlights: Developed a large-batch training recipe for Yi-7B, scaling the training batch size from 2M to 16M tokens while maintaining stable convergence.

Multi-behavior Recommendation Algorithms Design
Research Intern, Beijing Kuaishou Technology Co., Ltd **Beijing, China**
Mentor: Dr. Yang Song *Sep 2021 – Feb 2022*
Highlights: Designed a multi-behavior recommender based on hypergraph convolutional networks, achieving up to +23.19% improvement on Kuaishou internal dataset.

ACADEMIC EXPERIENCE

Tactile Sensing Simulation and Sim-to-Real Transfer
Remote Research Assistant, RoboTouch Lab, The Robotics Institute, CMU **Pittsburgh, USA**
Advisor: Prof. Wenzhen Yuan *Jun 2021 – Mar 2022*
Highlights: Built a PyBullet-based grasping simulator with tactile sensing to generate training data and support downstream ML policy training with sim-to-real transfer.

Conscious Deep Reinforcement Learning

Remote Research Assistant, Berkeley AI Research (BAIR) Lab, UC Berkeley **Berkeley, USA**

Advisor: Prof. Yang Gao and Prof. Huazhe Xu *Nov 2020 – Jun 2021*

Highlights: Explored a consciousness-inspired RL framework that dynamically switches between model-based and model-free control on MuJoCo tasks.

Large-scale GNN-based Social Recommendation Systems

Research Assistant, the Future Communications and Internet Lab, Tsinghua University **Beijing, China**

Advisor: Prof. Xu Chen and *Prof. Yong Li* *Apr 2019 – Sep 2020*

Highlights: Proposed a hypergraph-based heterogeneous social recommendation model, achieving state-of-the-art performance on multiple benchmarks.

PUBLICATIONS

Zirui Zhu, Hailun Xu, Yang Luo, Yong Liu, Kanchan Sarkar, Kun Xu, and Yang You. CAMEL: Confidence-Gated Reflection for Reward Modeling. *arXiv preprint*.

Yong Liu, Di Fu, Yang Luo, **Zirui Zhu**, Minhao Cheng, Cho-Jui Hsieh, and Yang You. POME: Post Optimization Model Edit via Muon-style Projection. *arXiv preprint*.

Zirui Zhu, Hailun Xu, Yang Luo, Yong Liu, Kanchan Sarkar, Zhenheng Yang, and Yang You. FOCUS: Efficient Keyframe Selection for Long Video Understanding. ***ICLR 2026***.

Yong Liu, Di Fu, Shenggan Cheng, **Zirui Zhu**, Yang Luo, Minhao Cheng, Cho-Jui Hsieh, and Yang You. SeedLoRA: A Fusion Approach to Efficient LLM Fine-Tuning. ***ICML 2025***.

Yang Luo, Zangwei Zheng, Ziheng Qin, **Zirui Zhu**, Yong Liu, and Yang You. MERIT: Maximum-normalized Element-wise Ratio for Language Model Large-batch Training. ***ICML 2025***.

Yong Liu, **Zirui Zhu**, Chaoyu Gong, Minhao Cheng, Cho-Jui Hsieh, and Yang You. Sparse MeZO: Less Parameters for Better Performance in Zeroth-Order LLM Fine-Tuning. ***NeurIPS 2025***.

Zirui Zhu, Yong Liu, Zangwei Zheng, Huifeng Guo, and Yang You. Helen: Optimizing CTR Prediction Models with Frequency-wise Hessian Eigenvalue Regularization. ***WWW 2024***.

Yang Luo, Zangwei Zheng, **Zirui Zhu**, and Yang You. How Does the Textual Information Affect the Retrieval of Multimodal In-Context Learning? ***EMNLP 2024***.

Zirui Zhu, Chen Gao, Xu Chen, Nian Li, Depeng Jin, and Yong Li. Inhomogeneous Social Recommendation with Hyper-graph Convolutional Networks. ***ICDE 2022***.

Zilin Si, **Zirui Zhu**, Arpit Agarwal, Stuart Anderson, and Wenzhen Yuan. Predicting Grasp Stability with Sim2Real Transfer from Tactile Sensing. ***IROS 2022***.

SERVICE & TEACHING

Conference Reviewer: ICML, ICLR, CVPR, ECCV, ACL

Journal Reviewer: IEEE Transactions on Knowledge and Data Engineering (TKDE)

Teaching Assistant: NUS CS3244 Machine Learning, Fall 2025

SKILLS

Programming: Python; Bash; Git

Frameworks: PyTorch; Transformers; vLLM; TRL; DeepSpeed; Accelerate; Hydra; W&B

Systems: Linux; Docker; Slurm