Zhanpeng Zhou Ph.D. Candidate

(+86) 177-1755-7410
zzp1012@sjtu.edu.cn
https://zzp1012.github.io

Education

School of Computer Science, Shanghai Jiao Tong University — Shanghai, China
Ph.D. in Computer Science and Engineering — 2022 - 2027 *(expected)*
- Advisor: Prof. Junchi Yan
- Research area: Theory of Deep Learning, Science of Large Language Models

UM-SJTU Joint Institute, Shanghai Jiao Tong University — Shanghai, China
B.E. in Electrical and Computer Engineering — 2018 - 2022
- Minor in Data Science.

Experiences

MiniMax. | Shanghai, China — 2025/04 - 2025/10
- Research Intern, focusing on LLM pre-training.

RIKEN AIP. | Tokyo, Japan — 2025/04 - 2025/10
- Student Trainee, advised by Prof. Taiji Suzuki.

National Institute of Informatics. | Tokyo, Japan — 2023/09 - 2024/03
- Visiting Ph.D. Student, advised by Prof. Mahito Sugiyama.

Mila Quebec. | Montreal, Canada — 2021/03 - 2021/06
- Research Intern, advised by Prof. Jian Tang.

Working Papers

(* indicates equal contributions; †indicates correspondence.)

1. **Zhanpeng Zhou***†, Yuhan Sun*, Bingrui Li, Jinbo Wang, Huaijin Wu, Lei Wu, Junchi Yan. How Does Local Landscape Geometry Evolve in Language Model Pre-Training? **In Submission.**
2. **Zhanpeng Zhou**†, Yongyi Yang, Mahito Sugiyama, Junchi Yan. New Evidence of the Two-Phase Learning Dynamics of Neural Networks. **In Submission.**
3. Tongcheng Zhang*, **Zhanpeng Zhou***†, Mingze Wang, Andi Han, Wei Huang, Taiji Suzuki, Junchi Yan. On the Learning Dynamics of Two-Layer Networks with Label Noise SGD. **In Submission.**
4. Tongtian Zhu, Tianyu Zhang, Mingze Wang, **Zhanpeng Zhou**†, Can Wang. On the Surprising Effectiveness of a Single Global Merging in Decentralized Learning. **In Submission.**
5. Xuan Cui, Yunfei Zhao, Bo Liu, Wei Duan, Huiyue Li, Run Zeng, Jinrui Qian, **Zhanpeng Zhou**. IGU-LoRA: Adaptive Rank Allocation via Integrated Gradients and Uncertainty-Aware Scoring. **In Submission.**
6. Jinbo Wang, Binghui Li, **Zhanpeng Zhou**, Mingze Wang, Yuxuan Sun, Jiaqi Zhang, Xunliang Cai, Lei Wu. Towards Revealing the Effect of Batch Size Scheduling on Pre-training. **In Submission.**
7. Bingrui Li, Jiaxin Wen, **Zhanpeng Zhou**, Jun Zhu, Jianfei Chen. Efficient Hyperparameter Tuning via Trajectory Invariance Principle. **In Submission.**
8. Huaijin Wu, Bingrui Li, Yebin Yang, Yi Tu, **Zhanpeng Zhou**, Jianfei Chen, Junchi Yan. Achieving Low-Bit Muon Through Subspace Preservation and Grid Quantization. **In Submission.**
9. Qibing Ren, Xinhao Song, Ke Fan, Lijun Li, **Zhanpeng Zhou**, Gongshen Liu, Junchi Yan, Lizhuang Ma, Jing Shao. Unintended Harmful Knowledge Elicitation Issue in Large Reasoning Models and a RL Solution. **In Submission.**
10. Qiu et al. (**Co-author**). On Path to Multimodal Historical Reasoning: HistBench and HistAgent. **In Submission.**

Publications

(* indicates equal contributions; †indicates correspondence.)

1. Jinbo Wang*, Mingze Wang*, **Zhanpeng Zhou***, Junchi Yan, Weinan E, Lei Wu. The Sharpness Disparity Principle in Transformers for Accelerating Language Model Pre-Training. **ICML 2025**.
2. Andi Han*, Wei Huang*, **Zhanpeng Zhou***†, Gang Niu, Wuyang Chen, Junchi Yan, Akiko Takeda, Taiji Suzuki. On the Role of Label Noise in the Feature Learning Process. **ICML 2025**.
3. Zijun Chen*, **Zhanpeng Zhou***†, Bo Zhang, Weinan Zhang, Xi Sun, Junchi Yan. SE-Merging: A Self-Enhanced Approach for Dynamic Model Merging. **IJCNN 2025**.
4. **Zhanpeng Zhou***†, Mingze Wang*, Yuchen Mao, Bingrui Li, Junchi Yan. Sharpness-Aware Minimization Efficiently Selects Flatter Minima Late in Training. **ICLR 2025 (Spotlight)**.
5. **Zhanpeng Zhou**†, Yongyi Yang, Jie Ren, Mahito Sugiyama, Junchi Yan. On the Cone Effect in the Learning Dynamics. **ICLR 2025 Workshop**.
6. Bingrui Li, Wei Huang, Andi Han, **Zhanpeng Zhou**, Taiji Suzuki, Jun Zhu, Jianfei Chen. On the Optimization and Generalization of Two-layer Transformers with Sign Gradient Descent. **ICLR 2025 (Spotlight)**.
7. **Zhanpeng Zhou***, Zijun Chen*, Yilan Chen, Bo Zhang, Junchi Yan. On the Emergence of Cross-Task Linearity in the Pretraining-Finetuning Paradigm. **ICML 2024**.
8. Yiting Chen, **Zhanpeng Zhou**, Junchi Yan. Going Beyond Neural Network Feature Similarity: The Network Feature Complexity and Its Interpretation Using Category Theory. **ICLR 2024**.
9. **Zhanpeng Zhou***, Wen Shen*, Huixin Chen*, Ling Tang, Quanshi Zhang. Batch Normalization Is Blind to the First and Second Derivatives of the Loss. **AAAI 2024 (Oral)**.
10. **Zhanpeng Zhou**, Yongyi Yang, Xiaojiang Yang, Junchi Yan, Wei Hu. Going Beyond Linear Mode Connectivity: The Layerwise Linear Feature Connectivity. **NeurIPS 2023**.
11. Ling Tang*, Wen Shen*, **Zhanpeng Zhou**, Quanshi Zhang. Defects of Convolutional Decoder Networks in Frequency Representation. **ICML 2023**.
12. Jie Ren, **Zhanpeng Zhou**, Qirui Chen, Quanshi Zhang. Can We Faithfully Represent Absence States to Compute Shapley Values on a DNN? **ICLR 2023**.
13. Jie Ren*, Die Zhang*, Yisen Wang*, Lu Chen, **Zhanpeng Zhou**, Yiting Chen, Xu Cheng, Xin Wang, Meng Zhou, Jie Shi, Quanshi Zhang. A Unified Game-Theoretic Interpretation of Adversarial Robustness. **NeurIPS 2021**.

Awards and Honors

- **National Scholarship (top 0.2%)**, Ministry of Education 2024/11
- **Top Internship Evaluation**, National Institute of Informatics 2024/03
- **Outstanding Graduate Student**, Shanghai Jiao Tong University, 2022/05
- **Yu Liming Scholarship**, Shanghai Jiao Tong University 2021/11
- **John Wu & Jane Sun Scholarship**, Shanghai Jiao Tong University 2020/11
- **Best Technology Award in Summer Expo**, Shanghai Jiao Tong University 2019/08

Academic Services

Conference Reviewer: *ICML ('22-25), NeurIPS ('22-25), ICLR ('24-26)*

Journal Reviewer: *T-PAMI, Intelligent Computing (Science Partner)*