Zihang Liu

+1 3106259919 | Homepage | zihang.liu@berkeley.edu | LinkedIn | GitHub | Berkeley, CA

Education

University of California, Berkeley — Berkeley, CA
Master in Electrical Engineering and Computer Science — *Aug 2024 – Present*

Beijing University of Posts and Telecommunications — Beijing, China
B.Eng. in Intelligence Science and Technology ***GPA:** 90%, **Rank:** 4/68* — *Sep 2020 – Jun 2024*

Research Interests: Transparency of Deep Learning, Large-scale Optimization and Efficiency of Neural Networks, Numerical Algorithm Discovery.

Publications

[1] Principal Weights Emerge after Rank Reduction for Sparse Fine-Tuning — ICML 2025
Z. Liu, T. Pang, O. Balabanov, C. Yang, T. Huang, L. Yin, Y. Yang, S. Liu

[2] Model Balancing Helps Low-data Training and Fine-tuning — EMNLP 2024 Oral
Z. Liu, Y. Hu, T. Pang, Y. Zhou, P. Ren, Y. Yang

[3] EnsembleMOT: A Step Towards Ensemble Learning of Multiple Object Tracking — arXiv preprint
Y. Du, **Z. Liu**, F. Su

Research Experience

Numerical Algorithm Discovery with Deep Learning — Berkeley, CA
*Graduate Researcher, Advised by **Prof. Michael Mahoney** at **UC Berkeley*** — *Aug 2024 – Present*

- Designed a neural network training pipeline to generate polynomial preconditioners to accelerate power iteration method for solving eigenproblems with provable guarantees. Introduced a novel training method with random sketching to dramatically improve inference efficiency and generalization in input matrix size and distribution.
- Proposed a transformer-based linear system solver that utilizes to learn preconditioning through in-context learning. Reaches comparable performance to conjugate gradient and steepest descent method.
- Designed an algorithm discovery pipeline with LLMs such as LLaMA-3.1-70B and GPT-4o. Introducing MCTS decoding algorithm to numerical algorithm discovery for the first time. Conducted empirical analysis of LLM's function discovery ability with level-1 and level-2 reasoning.

Weight Matrix Analysis of Foundation Models — Hanover, NH
*Undergraduate Researcher, Advised by **Prof. Yaoqing Yang** at **Dartmouth College*** — *Jul 2023 – Jul 2024*

- Proposed a layer-wise learning rate scheduler based on heavy-tailed self-regularization theory(HT-SR), that balances temperature parameters of neural network models.
- Modeled the heavy-tail behavior of optimizers (SAM) and model architectures. Rescheduling the learning rate to optimize the regularization effects, significantly improving performance on Image Classification and Language Modeling (NeurIPS 2023 Spotlight)
- Diagnosing the limitations of low-data training using Heavy-Tail metrics, and propose layer-wise model balancing to achieve model alignment, achieving up to 10% improvement in LLM fine-tuning. (EMNLP 2024 Oral)

Ensemble Methods in Multiple Object Tracking — Beijing, China
*Undergraduate Research Assistant, Advised by **Prof. Fei Su*** — *Jun 2022 – Jan 2023*

- Proposed a model-independent ensemble method that integrates results from various MOT trackers to achieve higher overall performance, which we named EnsembleMOT.
- Proposed to use both spatial and temporal IoU(Intersection over Union) to merge and prune trajectories, achieving 3% improvement in MOTA and IDF1, alleviating ID-switch and abnormal bounding box problems.
- Co-authored a paper with an MCPRL lab member and submitted our work to the International Conference on Acoustics, Speech and Signal Processing (Submitted to ICASSP 2023).

Awards & Services

Reviwer: NeurIPS 2024 Workshop on Foundation Models for Science, ICLR 2025 Workshop on Scalable Optimization for Efficient and Adaptive Foundation Models, ICLR 2025 Workshop on Sparsity in LLMs

Services: EMNLP 2024 Student Volunteer Coordinator

Grants: Berkeley Conference Travel Grant (2024)

Scholarships: UC Berkeley Fung Excellence Scholarship (**$10000**, 2024), First-class Scholarship (**Ranked 1/68**, 2023), Second-class Scholarship (**Ranked 4/68**, 2021, 2022)

Talks

EMNLP 2024	Model Balancing Helps Low-data Training and Fine-tuning (Recording)	Nov 12, 2024

Skills

English Proficiency: TOEFL IBT **114** (reading 30 listening 29, speaking 27, writing 28), GRE **328** (Quant 170)

Programming Languages: Python, C/C++, SQL, Rust, VHDL

Frameworks: Linux, Pytorch, Git, Slurm

Sports: I was a tennis athlete representing BUPT and have entered quarterfinals in regional and national championships as a doubles player.