Zhizhou Sha

☎ (+86) 18210341740 | ✉ shazhizhou0@gmail.com | https://jamessand.github.io

Education

Tsinghua University September 2020 - June 2025

B.Eng. in **Computer Science and Technology** Beijing, China

- **GPA**: 3.83 / 4.00
- **Honors**: Comprehensive Scholarship (Top 30%) 2023, 2024
- **Selected Courses of A or A+**:
  - Research on Trending Topics in Natural Language Processing; Fundamentals of Computer Graphics
  - Introduction to High Performance Computing; Computer Architecture; Cybersecurity Fundamentals; Database Special Topic Training; Principles and Practice of Compiler Construction

Research Experience

Remote Research Intern March 2024 - Present

Advised by **Yingyu Liang**, **Zhenmei Shi**, and **Zhao Song**.

- Multi-Layer Transformers Gradient Can be Approximated in Almost Linear Time(Submit to ICLR 2025)[**Paper**]. We prove from a theoretical perspective, the training time of multi-layer Transformers can be accelerated from $O(n^2)$ to $n^{1+o(1)}$ through low rank approximation.
- Differential Privacy Mechanisms in Neural Tangent Kernel Regression (Submit to WACV 2025). [**Paper**]. We provide DP guarantees for NTK Regression, while ensuring the PSD property of the NTK kernel matrix.
- Looped ReLU MLPs May Be All You Need as Programmable Computers (Submit to AISTATS 2025). [**Paper**]. We explore the capability upper bound of the Looped ReLU MLP and prove its capability is equivalent to a programmable computer.
- HSR-Enhanced Sparse Attention Acceleration (Submit to ICLR 2025). [**Paper**]. We leverage the sparsity in ReLU attention and combine it with the HSR data structure to improve the computation time of cross attention from $O(mnd)$ to $O(mn^{4/5}d)$.

mlPC Lab @ University of California San Diego June 2023 - December 2023

Advised by **Zhuowen Tu**

- TokenCompose: Text-to-Image Diffusion with Token-level Supervision (**CVPR 2024**). [**Paper**], [**Project Page**]. Incorporate a token-level loss into the fine-tuning objective of Stable Diffusion Models, thereby augmenting the capability for compositional generation of instances across multiple categories.
- OmniControlNet: Dual-stage Integration for Conditional Image Generation (**CVPR 2024 Workshop**). [**Paper**]. Unify the external condition control for image generation in a single dense model.
- Dolfin: Diffusion Layout Transformers without Autoencoder (**ECCV 2024**). [**Paper**]. Enhance layout generation capability of Diffusion models by removing the autoencoder.

Internship Experience

Research Intern @ Bytedance December 2023 - January 2024

- Finetune LLMs to classify text content according to given rules, reducing the need for manual labeling efforts.

Academic Services

Conference Reviewer: AAAI 2025, WACV 2025, ICLR 2025.

Skills

Language: Mandarin Chinese (Native), English (Proficient)

Programming Skill: Python (PyTorch), LaTeX, Linux, C++, JavaScript, Verilog, VHDL, CUDA.