Yuezhou Hu

huyz21@mails.tsinghua.edu.cn ● +86 177-4848-6233

Education

Tsinghua University   Beijing, China
(Expected in June 2025) Bachelor of Engineering,
Department of Computer Science and Technology.   Sep. 2021 - Present
GPA 3.82/4.0

Research Interests

My research interests include **efficient algorithms for machine learning**, including model compression methods, particularly structured sparsity, dynamic sparse training, quantization methods, and etc. I want to break the monopoly of unicorns in AI and make deep learning affordable and accessible for every researcher.

Experience

State Key Lab. of Intelligent Teh. & Sys., Tsinghua Uni.   Beijing, China
Group Leader   May 2023 – Present

- Conducted a thorough survey on state-of-the-art 2:4 pre-training techniques for large transformers.
- Devised a set of algorithms adaptable to all phases in transformer pre-training, balancing efficiency and performance at the same time, which succeeded previous 2:4 training methods.
- Implemented a set of tools for efficient 2:4 pre-training, with CUDA, OpenAI Triton, and Pytorch.
- Be the first to report real pre-train acceleration ratio via 2:4 sparsity.

Publications and and working papers

- **Accelerating Transformer Pre-Training with 2:4 Sparsity.**
  Yuezhou Hu, Kang Zhao, Weiyu Huang, Jianfei Chen, Jun Zhu
  To appear in International Conference on Machine Learning (ICML), 2024
- **S-STE: Continuous Pruning Function for Efficient 2:4 Sparse Pre-training.**
  Yuezhou Hu, Jun Zhu, Jianfei Chen
  Submitted for Neural Information Processing Systems (NeurIPS), 2024

Leadership & Activities

Vice President of Tsinghua University Summer School for Undergraduate Applicants.   Aug. 2022
Volunteered in the Undergraduate Admissions Office of Tsinghua University.   July 2022

Technical Skills

Deep learning programming: Python, Pytorch
GPU Programming: OpenAI Triton, C++, CUDA
Others: Docker, Django, Rust