

Yuezhou Hu

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

Education

Tsinghua University

(Expected in June 2025) Bachelor of Engineering,
Department of Computer Science and Technology.
GPA 3.80/4.0

Beijing, China

Sep. 2021 - Present

Research Interests

My research interests include **efficient algorithms for machine learning**, including model compression methods, particularly dynamic sparse training and quantization. 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.

Role: Research Intern Advisor: Prof. Jianfei Chen, Prof. Jun Zhu

Beijing, China

2022 – Present

- Conducted a thorough survey on state-of-the-art 2:4 pre-training techniques for large transformers.
- Be the first to report real pre-train acceleration ratio via 2:4 sparsity.

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Tech.

Role: Research Intern Advisor: Prof. Tuo Zhao

Atlanta, US

July 2024 – Present

- Conducted a survey on knowledge distillation (KD) methods for efficient speculative decoders.

Publications

- **Accelerating Transformer Pre-training with 2:4 Sparsity.**
arXiv OpenReview PDF Project page
Yuezhou Hu, Kang Zhao, Weiyu Huang, Jianfei Chen, Jun Zhu
In proc. of International Conference on Machine Learning (ICML), Vienna, Austria, 2024.
- **S-STE: Continuous Pruning Function for Efficient 2:4 Sparse Pre-training.**
arXiv OpenReview Project page
Yuezhou Hu, Jun Zhu, Jianfei Chen
To appear in Neural Information Processing Systems (NeurIPS), Vancouver, Canada, 2024.
- **Pruning Large Language Models with Semi-Structural Adaptive Sparse Training.**
arXiv
Weiyu Huang, **Yuezhou Hu**, Guohao Jian, Jun Zhu, Jianfei Chen
Submitted for AAAI 2025.

Technical Skills

Deep learning programming: Python, Pytorch

GPU Programming: OpenAI Triton, C++, CUDA

Others: Docker, Django, Rust

Language Skills: TOEFL 104 (R:28/L:27/S:23/W:26)