MINGHAO YAN

Email: myan@cs.wisc.edu
Personal Website: https://minghaoyan.netlify.app/

EDUCATION

PhD in Computer Science, August 2022 - Now
University of Wisconsin - Madison

Master of Science in Computer Science, May 2022
Bachelor of Science in Computer Science, ***Summa Cum Laude*** May 2021
Rice University, Houston, Texas
GPA: 4.03/4.09 (Only the first year was weighted due to school policy change)

RESEARCH EXPERIENCE

Research Assistant at University of Wisconsin - Madison, August 2022 - Now
Advised by Prof. Shivaram Venkataraman

- Current Project: Understanding the energy implications of algorithmic, software, and hardware optimizations of neural network inference.

Research Assistant at Rice University, May 2019 – May 2022
Advised by Prof. Anshumali Shrivastava

- Project 2: Designed and implemented a communication efficient and scalable MPI based framework for distributed neural network training on CPUs via Locality Sensitive Hashing (LSH)
  - **Workshop tutorial at the 5th Annual Ken Kennedy AI and Data Science Conference**
  - **Deployed at ThirdAI Corp., a VC-backed start-up focusing on democratizing deep learning**
- Project 1: Designed and implemented a fast Bloom Filter based algorithm for Multiple Set Membership Testing (MSMT) that can index 170TB of genomic data in less than 9 hours and query in sublinear time (**Published at SIGMOD 2021**).

INDUSTRY EXPERIENCE

AI Engineer Intern, ThirdAI Corp., August 2021 – May 2022

- Developed and commercialized our framework for training large neural networks on CPUs. Extended the training framework from MLPs to recommendation models and GNNs.
- Developed a 1000x DLRM embedding table compression framework based on hashing techniques.
- Extended support for multiple data formats and hash functions.

PREPRINTS

- Distributed SLIDE: Enabling Training Large Neural Networks on Low Bandwidth and Simple CPU-Clusters via Model Parallelism and Sparsity
  Minghao Yan, Nicholas Meisburger, Tharun Medini, Anshumali Shrivastava
  Arxiv: https://arxiv.org/pdf/2201.12667.pdf
  Deployed at ThirdAI Corp.

- PairConnect: A Compute-Efficient MLP Alternative to Attention
  Zhaozhuo Xu, **Minghao Yan,** Junyan Zhang, Anshumali Shrivastava
  In submission to ACL 2023

PUBLICATION

- Fast Processing and Querying of 170TB of Genomics Data via a Repeated And Merged BloOm Filter (RAMBO)
  Gaurav Gupta*, Minghao Yan*, Benjamin Coleman, Bryce Kille, R. A. Leo Elworth, Tharun Medini, Todd Treangen, Anshumali Shrivastava
  SIGMOD 2021
  (* denotes equal contribution)

TALKS

- Democratizing Deep Learning with Commodity Hardware: How to Train Large Deep Learning Models on CPU Efficiently with Sparsity
  Workshop tutorial at 5th Annual Ken Kennedy AI and Data Science Conference, Oct 27th, 2021

AWARDS

- First-year Graduate Student Scholarship (Awarded to top PhD applicants by the CS department), 2022
- Summa Cum Laude (Top 5% GPA in the School of Engineering), 2021
- President's Honor Roll, 6/7 Eligible semesters
- Research Fellowship for Master of Science in Computer Science, 2020-2022
- Louis J. Walsh Scholarship, 2019-2020, 2020-2021
- Rice Undergraduate Data Science Summer Program (RUDSSP) Fellowship, 2018

TEACHING EXPERIENCE (TA)

- Spring 2023: Introduction to Big Data Systems (CS639)
- Fall 2022: Programming III (CS400)
- Spring 2022: Machine Learning (COMP642)
- Spring 2020: Honors Calculus IV (MATH222)
- Fall 2019, 2020: Reasoning About Algorithms (COMP382)
- Spring 2019: Algorithmic Thinking (COMP182)
- Fall 2018: Honors Calculus III (MATH221)

PROFESSIONAL SERVICE

- Reviewer: NeurIPS 2022, ICML 2022, 2023

SKILLS

- Languages: English (Bilingual); Mandarin (Native)
- Proficient with Python, C/C++, and Java; Basics of MATLAB, R, C#, and JavaScript.