MINGHAO YAN

Email: Minghao.Yan@rice.edu
Personal Website: https://minghaoyan.netlify.app/

EDUCATION

Master of Science in Computer Science, Expected: 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 Rice University Hashing and Sketching Lab, May 2019 – Now
Advised by Prof. Anshumali Shrivastava

- Current Project: Designed and implemented of an MPI based distributed framework for fast Neural Network training on CPUs by leveraging hashing techniques **(In submission to MLSys 2022).**
  - **Workshop tutorial at 5th Annual Ken Kennedy AI and Data Science Conference**
- Previous Project: Designed and implemented a fast Bloom Filter based algorithm for Multiple Set Membership Testing (MSMT) problem that can index 170TB of genomic data in less than 9 hours and query in sublinear time (**Published at SIGMOD 2021**).

Research Assistant at Rice Undergraduate Data Science Summer Program (RUDSSP), May 2018 – July 2018
Advised by Prof. Luay Nakhleh and Prof. Hamim Zafar
Project: Simulating, inferring, and scaling up single nucleotide variants in single-cell genomic data

- Built a genome data simulator for ground truth generation, a genome amplifier, and optimized a genome reads generator.

INDUSTRY EXPERIENCE

AI Engineer Intern, Third AI Corp., August 2021 – Now

- Developing and commercializing our framework for training large neural networks on CPUs. Extending the training framework from MLP to recommendation models and GNNs.
- Extending support for multiple data format and hash functions.

PREPRINTS

- PairConnect: A Compute-Efficient MLP Alternative to Attention
  Zhaozhuo Xu, **Minghao Yan,** Junyan Zhang, Anshumali Shrivastava
  In submission to ACL 2022.

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

- Summa Cum Laude (Top 5% in the School of Engineering), 2021
- President's Honor Roll, 6/8 Semesters
- Research Fellowship for Master of Science in Computer Science, 2020
- Louis J. Walsh Scholarship, 2019-2020, 2020-2021
- Rice Undergraduate Data Science Summer Program (RUDSSP) Fellowship, 2018

TEACHING EXPERIENCE (TA)

- Fall 2018: Honors Calculus III (MATH221)
- Spring 2019: Algorithmic Thinking (COMP182)
- Fall 2019, 2020: Reasoning About Algorithms (COMP382)
- Spring 2020: Honors Calculus IV (MATH222)

SKILLS

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