BEICHEN HUANG

huangb21@mcmaster.ca +1 647-896-3986
https://github.com/BeichenHuang https://beichenhuang.github.io/

EDUCATION

McMaster University, Canada *Sept. 2019 - now*

- Bachelor of Engineering: Mechatronics Engineering Cumulated GPA: *3.9 / 4.0*
- Relevant Coursework: Statistics, Algebra, Algorithm and Software Design, Machine Learning, Scientific Computation and Simulation, Analog and Digital Circuits, Dynamic Model and Control, Signals and Systems, Operating System

RESEARCH EXPERIENCE

Research Assistant

Supervisor: Prof. Minjia Zhang *University of Illinois Urbana-Champaign, May 2024 - Present*

- Focused on developing an efficient and cost-effective Large Language Model in the Mixture-of-Expert (MoE) structure, combining the mathematical optimization method with Post Training Quantization to compress the model.
- Conducted an in-depth evaluation of trending quantization methods across six benchmarks, and successfully advancing the quantization from 4-bit to 3-bit.
- Innovatively brought the low rank matrix method to compensate the error of quantized weight for MoE models, and conducted experiments on a variety of rank strategies. Resulted in significant performance improvement with negligible additional memory overhead.

Research Assistant

Supervisor: Prof. Kaiming Shen *The Chinese University of Hong Kong (ShenZhen), Sept 2023 - May 2024*

- Enhanced the Quadratic Transform algorithm to address fractional programming problems, and effectively apply this advanced method in solving complex challenges in machine learning and wireless communication.
- Addressed the clustering problem with fractional programming methods. Employed the Quadratic Transform algorithm for direct optimization of the clustering objective function, resulting in optimal clustering outcomes.
- Formulated an innovative wireless communication model incorporating Aerial Intellectual Reflective Surface (AIRS). Focused on optimizing and analyzing load balancing within this model, enhancing overall system efficiency.

PUBLICATION

- **Multidimensional Fractional Programming for Normalized Cuts** under review at *NeurIPS 2024*
  Yannan Chen*, **Beichen Huang***, Licheng Zhao, Kaiming Shen
- **Aerial-IRS-Assisted Load Balancing In Downlink Networks** *ICASSP 2024*
  Shuyi Ren, **Beichen Huang**, Xiaoyang Li, Kaiming Shen

WORKING EXPERIENCE

Software Engineer Intern *Magna Electronics, May 2022 - May 2023*

- Designed, developed, and debugged for image processing algorithm with ground truth and debugging information visualization function for the autonomous driving system. Diligently managed the project repository on GitHub.
- Effectively maintained the C++ Advanced Driver-Assistance System program, mainly focused on solving the defects of the Human Machine Interface and the data pipeline in response to customer feedback.

Teaching Assisstant *McMaster University, Dec. 2021 - May 2022*

- Actively engaged in 10 lab and tutorial sessions related to embedded programming, designed and taught material to inspire students to have a clear understanding of the software for embedded systems.
- Taught and solved questions and requests from over 60 students, and received a high rating at the end of the term.

SKILLS

Language:	English (Fluent), Chinese (Native)
Programming:	Python and Pytorch, MATLAB, C, C++, ARM Assembly, Simulink, Keil, Git, LaTeX, R
Software & Tool:	PyChram, MATLAB, Colab, VS Code, Autodesk Inventor, Altium Designer, NI Multisim