Xiaonan LUO

xluobd@connect.ust.hk
+852 5574-8799

Education

Bachelor of Science — *The Hong Kong University of Science and Technology*
Major in Computer Science & Data Science — *September 2020 – June 2024*

- *GPA: 3.65/4.3 (Around 10% in CS Dept)*
- *Graduate Courses: Combinatorial Optimization, Computer Vision*
- *Scholarship: Chern Class Scholarship from Department of Mathematics, University's Scholarship Scheme for Continuing Undergraduate Students, Soong Ching Ling Scholarship*

Exchange Student — *Northwestern University*
2022 Fall
Graduate Courses: Operating System, Machine Learning

Research Interests

My research interests are primarily in computer systems and architecture, with more focus on the intersection of systems and machine learning(ML).

Publication

- **Xiaonan Luo***, Yichao Fu,* Cheng Wan, Zhifan Ye, Yingyan Lin. *VR-BNS: Variance Reduction for Boundary Nodes Sampling for full-graph training*. (In preparation)
- Minchen Yu, Ao Wang, Dong Chen, Haoxuan Yu, **Xiaonan Luo**, Zhuohao Li, Wei Wang, Ruichuan Chen, Dapeng Nie, Haoran Yang. *FaaSwap: SLO-Aware, GPU-Efficient Serverless Inference via Model Swapping*. arXiv:2306.03622 (Submitted to EuroSys' 24)

Research Experience

Designing a CXL-GPU Heterogeneous Memory-Tiered System for DLRM — *2023 Summer*
Advised by Prof. Yufei Ding in UCSD — *UCSD, U.S.*

- Conducted research in system and architecture design for memory-intensive Deep Learning Recommendation Model (DLRM), with an aim to alleviate memory pressure and minimize training latency overhead.
- Proposed a CXL-GPU heterogeneous memory-tiered system.
- Designed a CXL-featured cache mechanism, leveraging the granularity of the CXL-enabled system to mitigate inter-device communication.
- Developed a comprehensive memory allocation algorithm to optimize over different memory hierarchies and minimize embedding lookup latency.

VR-BNS: Variance Reduction for Boundary Nodes Sampling GNN training — *2023 Spring, Summer*
Advised by Prof. Yingyan Lin in Georgia Institute of Technology — *Gatech, U.S.*

- Conducted research in Graph Neural Network (GNN) training optimization, closely related to BNS-GCN, a boundary node sampling-based training framework. This aimed to reduce memory footprint and communication volume.
- Optimized and Implemented Graph Convolution Network (GCN) and Graph Attention Network (GAT) computation algorithm, leveraging the insight of history aggregation embedding to approximate feature prediction under full-graph training.
- Included tensor compression technique to further reduce memory footprint on accelerators, in addition to the sampling-based memory reduction.

FaaSwap: SLO-Aware, GPU-Efficient Serverless Inference via Model Swapping *2023 Winter, Spring*
Advised by Prof. Wei Wang in HKUST *HKUST, HK*

- Engaged in a research project focused on optimizing Machine Learning (ML) inference for serverless computing, with the primary objective of enhancing accelerator utilization under latency-aware inference. Submitted to EuroSys 24': FaaSwap
- Designed several critical aspects of FaaSwap, including GPU remoting, model swapping, memory management, asynchronous server-client communication, and a scheduling algorithm technique.
- Conducted extensive experiments to evaluate the performance of the system.

Professional Experience

Software Engineer Intern *2022 Summer*
Meituan *Beijing, China*

- Implemented Meituan Network Automatic Platform(MNAP) for switch operation and maintenance

Course Projects

Course Planning System
HKUST

- Developed a Course Planning System, the source code of which is hosted on Github.
- The project aims to help college students with their course selection, based on pre-requisites, exclusions, credit limitations, and other practical requirements.

Operating System Simulation
Northwestern

- Develop some OS functionalities including paging, schedulers, and device drivers using C, the source code of which is hosted on Github.

Skills

Coding	C/C++, Python, Golang
Framework	PyTorch, DGL
Language	Fluent in English, Native Mandarin Chinese

Activities

Tunatics A-Cappella team, HKUST Since 2020
Player in Soccer Team of Student Society, HKUST Since 2020