Renjie LIU

Shenzhen, China

liurj2023@mail.sustech.edu.cn
+86-15818629896 Liu-rj

EDUCATION

Southern University of Science and Technology **Sep. 2023 – Jun. 2026**
M.phil. in ***Electronic Science and Technology*** *ShenZhen, China*

Southern University of Science and Technology **Sep. 2019 – Jun. 2023**
B.Eng. in ***Computer Science and Technology*** *ShenZhen, China*

- **GPA: 3.82/4.00**
- Key Coursework: Linear Algebra (93), Discrete Math (96), Algorithm Design & Analysis (97), Database Principles (91), Computer Organization (94), Computer Network (96), Artificial Intelligence (92)
- Language Proficiency: TOEFL (102), CET-6 (590)

PUBLICATIONS

❋ Haitian Jiang, **Renjie Liu**, Xiao Yan, Zengfeng Huang, Zhenkun Cai, Minjie Wang, and David Wipf.
Forming Scalable, Convergent GNN Layers that Minimize a Sampling-Based Energy.
International Conference on Learning Representations (ICLR'25).

❋ Kaihao Ma*, **Renjie Liu***, Zhenkun Cai, Xiao Yan, Xiang Song, Minjie Wang, and James Cheng.
APT: Adaptive Parallel Training for Graph Neural Networks.
ACM Symposium on Principles and Practice of Parallel Programming (PPoPP'25).
∗Equal Contribution

❋ **Renjie Liu***, Yichuan Wang*, Xiao Yan, Zhenkun Cai, Minjie Wang, Haitian Jiang, Bo Tang and Jinyang Li.
DiskGNN: Bridging I/O Efficiency and Model Accuracy for Out-of-Core GNN Training.
ACM Conference on Management of Data (SIGMOD'25).
∗Equal Contribution

❋ Ping Gong, **Renjie Liu**, Zunyao Mao, Zhenkun Cai, Xiao Yan, Cheng Li, Minjie Wang, and Zhuozhao Li.
GSampler: General and Efficient GPU-based Graph Sampling for Graph Learning.
ACM Symposium on Operating Systems Principles (SOSP'23).

INTERNSHIPS

Amazon Web Services, Shanghai AI Lab **Jun. 2022 – Nov. 2024**
Applied Scientist Intern in ***DGL Team***

- Efficient GNN training on large-scale datasets in single-machine and distributed environments.

PROJECTS

Efficient out-of-core GNN Training on Large-scale Graphs **Sep. 2023 – April 2024**
Research Project Pytorch, C++, CUDA

- Training GNN on large-scale graphs is challenging as typically the graph topology and node features do not fit in the CPU memory of a single machine. Several systems store graph data on disk and conduct out-of-core processing, but they suffer from either read amplification or degraded model accuracy. To close this gap, we develop a system which achieves high I/O efficiency and thus fast training without hurting model accuracy.
  - ✏ Observe that existing disk-based GNN training systems face the tension between I/O efficiency and model accuracy and that they cannot achieve both simultaneously.

- ✏ Design DiskGNN to achieve both I/O efficiency and model accuracy by exploiting offline sampling, which collects the data access beforehand to optimize the data layout for access.
- ✏ Propose a suite of designs tailored for on-disk workloads to make DiskGNN efficient, which include four-level feature store, batched feature packing, and pipelined training.

Adaptive Parallel Training for Graph Neural Networks **April 2023 – Dec. 2023**
Research Project Pytorch, C++, CUDA

- Parallel training of GNN models with distributed environments is required to reduce training time and handle large graphs. Many parallel training systems are proposed, while the trade-off is not clear. In this project, we aim to build a general framework capable of automatically determine the best parallelism by jointly take into account data layout, computation and communication.
  - ✏ Provide a comprehensive survey of existing parallel GNN training schemes and summarize their trade-offs.
  - ✏ Drive all possible parallel training schemes from the perspective of matrix segmentation & multiplication, and implement all strategies with high efficiency.
  - ✏ Design a cost model to guide the system automatically determine the best parallel scheme given a specific environment, dataset and configuration.

General and Efficient GPU-based Graph Sampling for Graph Learning **Aug. 2022 – April 2023**
Research Project Pytorch, C++, CUDA

- Graph sampling prepares training samples for graph learning and can dominate the training time. Due to the increasing algorithm diversity and complexity, existing sampling frameworks are insufficient in the generality of expression and the efficiency of execution.
  - ✏ Survey 15 popular sampling algorithms for graph learning, summarize the unique characteristics, and propose a general 4-step ECSF model to sketch the sampling process.
  - ✏ Propose comprehensive matrix-centric APIs for graph sampling based on the ECSF model, which leads to succinct and intuitive implementations and also enables holistic execution optimizations.
  - ✏ Incorporate a suite of sampling-oriented optimizations for the execution of graph sampling algorithms on GPU and achieves an average speedup of 6.54x compared to SOTA systems.

HONORS AND AWARDS

- China National Scholarship **Nov. 2024**
- University Merit Scholarship for Outstanding Student, **The 1st Class** **Sep. 2023, 2022**
  - ✣ Ranked in the top 5% for overall performance for an academic year.
- Outstanding Graduates of Department of Computer Science and Engineering **Jun. 2023**
- ACM SIGMOD 2022 Programming Contest, **The World's Finalist** **April 2022**
  - ✣ Propose an effective entity resolution approach which achieves good recall in limited time.
  - ✣ Use BERT to encode descriptions to embedding vectors and retrieve top-k related entities by HNSW.

TECHNICAL SKILLS

Languages: Usually use Python, C/C++, CUDA, familiar with Java, SQL, etc.

Frameworks and Tools: Pytorch, DGL, PyG, MapReduce, PostgreSQL, Git, etc.

EXPERIENCES

Teaching Assistant

- **Computer Organization and Architecture** **Mar. 2022 – Jun. 2022**
  - ❍ Designed and graded course assignments and the course project (a Minisys CPU written in Verilog on FPGA).
  - ❍ Answered questions about CPU architecture, MIPS and Verilog programming in laboratory courses.
- **Introduction to Computer Programming** **Sep. 2020 – Jun. 2021**
  - ❍ Designed and graded course assignments and the course project (a Minesweeper implemented by Java).
  - ❍ Answered questions about programming skills and algorithms in laboratory classes.