Renyi Tang

Email: tangrenyi99@gmail.com Mobile: 872-216-7537
LinkedIn Website: renyi.live Address: Pittsburgh, PA

Education

- **Carnegie Mellon University** — Pittsburgh, PA
  M.S. in Machine Learning. **GPA**: 4.0/4.0 — Aug 2022 – Dec 2023 (expected)
- **University of Chicago** — Chicago, IL
  M.S. in Computer Science; B.S. in Mathematics. **GPA**: 3.82/4.0 — Sep 2017 – Dec 2021
  - Selected courses: DL Systems, Networks and Distributed Systems, Database Systems, Parallel Computing; Advanced ML, Advanced DL, Advanced NLP, Probabilistic Graphical Models, ML and Text Mining, Computer Vision

Work Experience

- **LinkedIn** — Sunnyvale, CA
  Software Engineering Intern — May 2023 - Aug 2023
  - Restructured and optimized Spark Scala dataframe schemas and transformations, saving HDFS storage by 20%.
  - Developed foundational Python code on negative sampling, supporting all online metric learning tasks at LinkedIn.
  - Enhanced CTR prediction metrics at LinkedIn Feed by 1% through sequence modeling on user interaction data.
  - Adopted agile software development principles by collaborating in a team setting and using online tools such as GitHub, JIRA and Confluence to organize workflow.
- **Huawei** — Shenzhen, China
  Software Engineering Intern — Jan 2022 - July 2022
  - Built scalable visualization tools for large-scale point cloud data in Python and improved render speed with C++ and CUDA by 75%. Enabled features not available in the industry's latest open-source packages.
  - Designed and deployed a 3D video data collection system in Python and C++. Enabled synchronization between different data sources through parallel programming to increase throughput and frame match accuracy by 21%.
  - Enhanced deep learning models based on PointNet(++) and PCPNet to improve multi-view stereo quality by 4%.
- **Squirrel AI** — Pittsburgh, PA
  Software Engineering Intern — Jun 2019 - Aug 2019
  - Implemented pipelines to feed and store one million daily pen stroke data with HDFS in Java.
  - Built real-time communication between front-end and back-end through the WebSocket network protocol, enabling live recognition of handwriting. Designed a smooth and aesthetic interface that raised user satisfaction.
  - Researched and packaged an algorithm that improved the accuracy of recognizing simple math equations to 95%.

Projects

- **Needle, a mini-PyTorch with CPU and CUDA backends (Python, C++, CUDA)**
  - Supported 30% faster high-performance matrix computation through tiling and shared memory.
  - Added advanced features powering the latest research such as subgraph backpropagation and implicit layers.
- **TwitterOps, a lambda architecture big data project (Java, Scala)**
  - Organized 100GB Twitter dataset into tables and analyzed account behavior. Deployed with AWS EC2 and S3.
  - Built batch layer by saving data to HDFS with Hadoop serialization and Hive, serving layer by extracting information with HBase, Spark, Spark SQL, and speed layer with Kafka to send data and maintain queues.
- **SimpleDB, a prototype database (Java)**
  - Implemented a database from scratch, including physical and logical data management, operators, query optimization and transaction management. Wrote comprehensive unit tests to benchmark performance over a variety of contexts.

Research Experience

- **Research Assistant, University of California San Diego, Dept. of ECE** — June 2021 - Sep 2021
  - Improved differential neural architecture search (NAS) methods with class weighting and ensemble learning.
  - Managed distributed large-scale experiments with MLOps tools such as Kubernetes, Docker and MLflow.
- **Research Assistant, University of Chicago, Dept. of CS** — June 2020 - March 2021
  - Conducted visual stream change detection with multi-level variational autoencoder (MLVAE). Proposed a graph-cut based loss function that generalized on long sequence of visual streams and improved accuracy by 12%.
  - Augmented dataset fivefold with different camera and lighting in Blender to improve model's generalizability.

Skills

- Python (advanced), Java, C/C++, CUDA, JavaScript, HTML/CSS, Node.js; PyTorch (advanced), TensorFlow, Spark, Hadoop, Hive, Kafka, Docker, Git, SQL