Mingzhe Hu

mingzhe.hu@columbia.edu | (+1) 646-925-0794 | https://superbtum.github.io

EDUCATION

Columbia University **New York, US**
M.S. in Electrical Engineering, GPA: 3.9 / 4.0 *Expected Dec 2022*
Relevant Courses: C++ Design, Big Data Analysis, Machine Learning, High Performance ML, DL, NLP, RL

Southeast University **Nanjing, CN**
B.Eng. in Information Engineering, GPA: 3.6 / 4.0 *Jun 2020*
Exchange @ Computer Science, TUM
Relevant Courses: Computer Vision, Intro to Database, Computer Networks, Computer Graphics Seminar

INDUSTRY EXPERIENCE

AI ModelShare Initiative **New York, NY**
Software Intern, Mentor: Michael D. Parrott *Sep 2022 – Dec 2022*
- Implemented data cleaning and preprocessing pipeline and ONNX transformation in scikit-learn, PyTorch, Keras
- Generated online machine learning models with **Amazon Web Service (AWS) Lambda** microservices through serverless Restful API and invoked by **API gateway**
- Deployed and dockerize tunable AutoML (Dabl & Autogluon) plugins with model ensemble and **Ray** parallelism

NVIDIA Corporation **New York, NY**
Software Intern, Mentor: Thomas Tang *May 2022 – Sep 2022*
- Established multi-camera object tracking through camera calibration and homography mapping with **OpenCV**
- Conducted spatio-temporal-appearance association, and self-adaptive thresholds for intra/inter-sensor DBSCAN clustering, with 80%+ IDF1 and 5~10% improvement of single camera tracking on warehouse and retail stores
- Developed a micro-batch data pipeline for streaming videos with **Apache Kafka** and unit-tested with coverage 85%+

SKILLS & HONORS

Honors: Top 1 in Megvii Workshop in Quantization with Sparsebit | Top 30% (19/77) ACM-ICPC GNY Regional | 1st place in Columbia Climate Change ML Workshop | Top 3% in Tianchi SVHN Detection & Recognition Challenge | Top 5 in Megvii Workshop in Mobile AI Photography of RAW Image Denoising
Programming Languages: C/C++/Cython, Python, CUDA/OpenCL, Matlab, Java, HTML/JavaScript
Platforms and Tools: Google Cloud Platform, AWS, CMake, Airflow, Apache Spark/Kafka, Git, Docker, BigQuery, React
Packages: Scikit-learn, PyTorch, Tensorflow, OpenMP, PostGreSql/SQLite, MongoDB/Postgres, TensorRT, Unittest
Operating Systems: Windows, Linux (CentOS, Ubuntu, Debian)

PROJECTS

Exam Paper and Question Bank Management System **New York, NY**
Project Leader, Attached Course: C++ Design *Sep 2022 – Dec 2022*
- Designed a multi-user single-server system on GCP, with pseudo-concurrency and low latency close to Microsoft
- Built a POSIX-based TCP socket communication with epoll and secured it with **OpenSSL**
- Managed data with **SQLite** & **MongoDB**, encrypted with PRAGMA key and accelerated query with clustered index
- Integrated **C++20** new features and abstract class to optimize code logic, unit-tested and compiled with **CMake**

Faster Pairwise Distance Calculation Library **New York, NY**
Project Leader, Attached Course: C++ Design *Sep 2022 – Nov 2022*
- Accelerated pairwise distance calculation with **OpenMP** and multi-threading and achieved 10X faster than Scipy
- Leveraged C++20 new feature of jthread and openmp **SIMD**/vectorization to make calculation 10X faster
- Wrapped the function with Cython and installed it as a Python library for portable utilization

Real-time New York Traffic Heatmap **New York, NY**
Project Leader, Attached Course: Large Scale Stream Processing *Jan 2022 – May 2022*
- Fetched real-time traffic with **Spark Streaming**, deployed on **Airflow** and displayed with **React** & **JavaScript**
- Reduced RDD data size to 10% with spark filtering, duplication removal and K-Means and stored data in **PostGreSql**
- Achieved 20X data access with asynchronous REST API request & web crawling, keep-alive, and FAIR scheduler

Acceleration of GloVe Representation on Heterogenous Platform **New York, NY**
Project Leader, Attached Course: Heterogenous Computing *Sep 2021 – Dec 2021*
- Worked with handcrafted **CUDA** and **CUBLAS**, designed work-efficient sum and maximum finder, with 5 times faster
- Created Bag of Tricks (BoT) for GPU arrays alignment, asynchronous logic and efficient atomic addition logic
- Achieved ≥ 80X faster than Numpy in naïve version and comparable speed with PyTorch