

Mingzhe Hu

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

EDUCATION

Columbia University

M.S. in Electrical Engineering, GPA: 3.9 / 4.0

Relevant Courses: Machine Learning, Deep Learning, Reinforcement Learning, Big Data Analysis, NLP

New York, US

Expected Dec 2022

Southeast University

B.Eng. in Information Engineering, GPA: 3.6 / 4.0, Exchange @ Computer Science, TUM

Relevant Courses: Computer Vision, Intro to Database, Computer Graphics Seminar

Nanjing, CN

Jun 2020

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: Python, CUDA, C/C++, Cython, Matlab

Platforms and Tools: Google Cloud Platform, Airflow, Twitter Streaming, Apache Spark/Kafka, Git, Shell

Packages: PyCUDA, PyOpenCL, PyTorch, Tensorflow, OpenMP, OpenAI Gym, PyMySQL, TensorRT, Rapids

Operating Systems: Windows, Linux (CentOS, Ubuntu, Debian)

INDUSTRY EXPERIENCE

Software Intern

Metropolis, NVIDIA Corporation, Mentor: Thomas Tang

New York, NY

May 2022 – Present

- Focused on multi-camera multi-target (MTMC) analytics and set up connectivity graph for association
- Implemented Bag of Tricks (BoT) + pairwise homography matching + Hungarian matching for spatio-temporal-appearance association, and self-adaptive thresholds for intra/inter-sensor clustering, with 15%+ IDF1 improvement of across sensor tracking and 5~10% improvement of single camera tracking
- Deployed MTMC framework into micro-batch/streaming scenario with Apache Kafka, lowering memory demand through temporary retention of local trajectories

PROJECTS

Unsupervised Object Reidentification in Smart Intersections

Research Assistant, Kostić Lab in COSMOS Smart Traffic with Prof. Zoran Kostić

New York, NY

Jan 2022 – May 2022

- Collected and annotated YoloV4-detected data in intersection including pedestrian & vehicles in overlapped & sequential, daytime & nighttime scenarios with ≥ 1500 instances and ≥ 100 identities
- Designed ResNet-IBN and ViT backbone under contrastive learning logic with gradient accumulation, and DBSCAN/Infomap for pseudo-label generation
- Evaluated with domain adaption and improved ReID mAP to 89% (pedestrian) & 79% (vehicle) with augmentation, bias diminish, variance reduction, object size tuning, and mixed distance metrics

Transformer Quantization on Text Classification Task

Project Leader, Attached Course: Deep Learning in System Performance

New York, NY

Jan 2022 – May 2022

- Fully quantized 8-bit transformer with exponential moving average observer to compress run-time model
- Achieved close text classification accuracy with AGNews dataset while reducing model size to 30%
- Explored impact of pre-trained weight and quantization latency, with 5% accuracy increase

Real-time New York Traffic Heatmap

Project Leader, Attached Course: Large Scale Stream Processing

New York, NY

Jan 2022 – May 2022

- Fetched real-time update of road condition and traffic density every 15 minutes with Spark Streaming
- Preprocessed RDD data with spark filtering, duplication removal and K-Means clustering
- Accelerated from 10 min to 20 sec with asynchronous REST API request, keep-alive, and FAIR scheduler
- Obtained empirically more reasonable local traffic analysis than Google Map

Acceleration of GloVe Representation on Heterogenous Platform

Project Leader, Attached Course: Heterogenous Computing

New York, NY

Sep 2021 – Dec 2021

- Worked with handcrafted CUDA and designed work-efficient sum and maximum finder, with 5 times faster
- Created Bag of Tricks (BoT) for GPU arrays alignment and efficient atomic addition logic
- Achieved ≥ 80 times faster than Numpy in naïve version and comparable speed with PyTorch