TIANLE ZHONG

+01 434-833-0770 | zhluosuu@outlook.com | luosuu.github.io

EDUCATION

University of Virginia (UVA) — **Charlottesville, USA**
PhD in Computer Science, Dept. Computer Science, School of Applied Science and Engineering — *Sept 2022 to Now*

University of Electro-Communications (UEC) — **Tokyo, Japan**
Exchange Student, School of Informatics — *Oct 2020 to Aug 2021*

University of Electronics Science and Technology of China (UESTC) — **Chengdu, China**
Bachelor of Computer Science and Applied Mathematics, School of Computer Science and Engineering — *Sept 2018 to July 2022*

RESEARCH EXPERIENCE

Network Systems Science and Advanced Computing, UVA Biocomplexity Institute — **Charlottesville, USA**
Research Interests: Systems for Machine Learning and HPC — *Sept 2022 to Now*
- Advisor: Prof. Geoffrey C. Fox

Networking Research Group, Microsoft Research Asia — **Beijing, China**
Research Interests: Large-scale Distributed Machine Learning System & Algorithms — *Oct 2021 to May 2022*
- Mentor: Lei Qu (Senior Research Engineer)

UEC Haneda Sound Media Lab — **Tokyo, Japan**
Research Interests: Speech Processing, Machine Learning, Learning Representation, Acoustics — *Nov 2020 to Sept 2021*
- Advisor: Prof. Yoichi Haneda

SELECTED PUBLICATIONS

REIMU: Optimizing Data I/O for LLM Datasets on Remote Storage
Accepted by Cloud Intelligence/AIOps 2024 Workshop (Co-located with ASPLOS 2024)
Tianle Zhong, Jiechen Zhao, Xindi Guo, Qiang Su, Geoffrey Fox
- Core design: a granularity sweet point balancing data I/O efficiency and shuffle quality for LLM datasets on remote storage.

RINAS: Training with Dataset Shuffling Can Be General and Fast
Prepint arXiv:2312.02368
Tianle Zhong, Jiechen Zhao, Xindi Guo, Qiang Su, Geoffrey Fox
- Core design: explore the potential of in-batch asynchronous I/O and processing for ML data pipeline.

RTP: Rethinking Tensor Parallelism with Memory Deduplication
Preprint arXiv: 2311.01635.
Cheng Luo, Tianle Zhong, Geoffrey Fox
- Core design: a partitioned all-to-all operation to delay partial computation, hence achieving a lazy tensor materialization.

Spherical Convolutional Recurrent Neural Network For Real-time Robust Sound Source Tracking
Accepted by 2022 IEEE International Conference on Acoustics, Speech, & Signal Processing (ICASSP)
Tianle Zhong, Israel Mendoza Velazquez, Ren Yi, Hector Manuel Perez Meana, Yoichi Haneda
- Core design: a Laplacian graph-based spherical convolution to learn spatial stereoscopic features with $SO(3)$ equivariance

Involution Based Speech Autoencoder: Investigating the Advanced Vision Operator Performance on Speech Feature Extraction
Accepted by Oral Session of 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)
Tianle Zhong, Israel Mendoza Velazquez, Yoichi Haneda
- Nominated for *Outstanding Student Paper Award*

NOTABLE PROJECTS

Cylon: Distributed Pandas-like DataFrame for HPC — **UVA**
Open Source Research Project — *2022.09 - 2022.12*
- A fast, scalable, distributed memory, parallel runtime with a Pandas like DataFrame

PerfSim: Distributed Machine Learning System Performance Simulator — **MSRA**
A research project at Networking Research Group, Microsoft Research Asia — *2021.10 - 2022.05*
- Based on a graph-based computation operator flow profiler
- Extend the single node performance simulation to multi-node cases by a NCCL performance predictor

SKILLS, AWARDS & OTHERS

- **Programming:** Python (PyTorch), C & C++, Matlab, Java, Rust
- **Awards:** 2020 Most Valuable Member of Microsoft Student Club (issued by MSRA); Excellent Member of 2020 Tencent Cloud Development Summer Camp (10 of 300); UETSC College Pacesetter Scholarship; UEC Outstanding Student Certificate
- **Languages:** English: IELTS 7.0, GRE 321+3.5, have written 2 conference papers; Mandarin: native; Japanese: conversational.
- **Contests:** 2018 & 2019 UESTC Mathematical Modeling Contest (Second Award)