Tiancheng (TC) Yuan

 GitHub | LinkedIn | ✉ ty373[AT]cornell.edu

RESEARCH INTERESTS

Deep Learning, Computer Vision, Machine Learning and Systems, Distributed Learning

EDUCATIONS

Cornell University   Jan 2023 -
Ph.D in System Engineering | Artificial Interligence

Cornell University   Jan 2021 - Dec 2022
Master of Science in System Engineering, GPA: 3.7/4.3

University of Liverpool   Aug 2014 - June 2018
B.Eng in Electrical and Computer Engineering, GPA: 3.78/4.0

SKILLS

C/C++, Python, Java, Julia, PyTorch, Transformers, Timm, Megatron, DeepSpeed, Fairseq

PUBLICATIONS AND MANUSCRIPTS

- Khiem Pham, Wentao Guo, Yucheng Lu, <u>Tiancheng Yuan</u>, Fang Ruan, Christopher De Sa. **"CD-GraB: Coordinating Distributed Example Orders for Provably Accelerated Training"** In *Advances in Neural Information Processing Systems 2023* [paper]

- Tao Yu*, Wentao Guo*, Jianan Canal Li*, <u>Tiancheng Yuan</u>*, Christopher De Sa. **"MCTensor: A High-Precision Deep Learning Library with Multi-Component Floating-Point."** In *ICML 2022: Workshop on Hardware Aware Efficient Training.* [paper] [poster] [code]

- Xinwen Wang, Yu-Ju Huang, <u>Tiancheng Yuan</u>, Robbert van Renesse. **"Disaggregated Applications Using Nanoservices"** In *WORDS 2021: The Second Workshop On Resource Disaggregation and Serverless* [paper]

RESEARCH EXPERIENCE

Cornell University   Jan 2021 - present
Research Assistant

- **Distributed Gradient Balancing (advised by Prof. Christopher De Sa)**

  A recent line of research advocates that it is possible to find provably better data permutations for accelerating model convergence rate, and a new method named Gradient Balancing, GraB was proposed by Lu et al. (2022). Following this work, we first propose Pair Balance, a method that minimizes the discrepancy of gradient error without dependencies of stale gradient mean. We also implement D-GraB, an algorithm that involves a data ordering aware variant of Parameter Server named Order Server, which enables the distributed training while determining good data permutations for each worker. In theory and empirical results, we demonstrate that our methods, under the same assumptions, enjoy linear speed up compared to other data permutation methods (GraB and Random Reshuffling).

- **MCTensor: A High-Precision Deep Learning Library with Multi-Component Floating-Point (advised by Prof. Christopher De Sa)**

We developed MCTensor, the first library based on PyTorch for efficient high-precision computations. We implemented operators from basic summation to optimizer completely. MCTensor models in 16-bit float can match or outperform the PyTorch model with float32 or float64 precision. We implemented Hyperbolic Embedding MCTensor models, including Half Space and Poincaré Ball model, with improved performance on reconstructing *WordNet* mammals hierarchy.

- **Dense Transformer (advised by Prof. Kilian Weinberger)**

  We proposed a new architecture for transformers, the densely connected transformers. Previously, ResNet and DenseNet indicated that increasing the connectivity within the network architectures while also making Convolutional Networks deeper will make the model substantially more accurate and efficient to train. We implemented different variants for connecting Q, K, V layers with different weights which are also updated during training. We achieved better accuracies in many famous image classifications datasets, for instance, CIFAR-100, ImageNet-1K, FGVC datasets, etc.

- **Navigator: A Decentralized Scheduler for Latency-Sensitive ML Workflows(advised by Prof. Ken Birman)**

  We implemented the framework for tackling the problem of data placement and task placement for time-critical ML inferencing or training as edge applications. The Time-sensitive Inference and Data Placement Engine, Navigator, is suitable for any DAG-based ML workload. Navigator is capable of minimizing response times while still maintaining high hardware utilization.

WORK EXPERIENCE

- **Hardware Engineer Intern**   June 17 - Aug 17
  FANUC Robotics Co., Ltd., Shanghai, China
  Joined a team of engineers to develop, using MCS-51 microcontrollers, an autonomous restaurant service robot.

- **AI Algorithm Intern**   May 18 - Aug 18
  Xiwei Software Technology Co., Ltd., Hangzhou, China
  Joined a team effort to develop computer vision functionalities in C++ for face recognition, face alignment, age identification, and gender classification

AWARDS

- Student Excellence Award, University of Liverpool
- First Degree with Honors, University of Liverpool
- CIDA Summer Research Stipend

TEACHING AND EXTRACURRICULAR EXPERIENCES

- **CS 5412 Cloud Computing**   Fall 2022
  Teaching assistant, supervised by Prof. Ken Birman.
- **Team Leader, Optimal Selection of Airport Location, ICM/MCM**   Fall 2018
  Built a suite of three mathematical models to guide airport site selection