Yuhui Wang

Email: maxtfly@gmail.com
Cell: +65 89410507
Add: 10 Hillview Rise, Singapore (667972)

Education

Aug. 2021 – July. 2024	Master in Computer Science (Transferred from PhD Program), National University of Singapore
Sept. 2015 – June. 2010	Bachelor in Computer Science, Honours Program, Xi'an Jiaotong University
Jan. 2019 – Dec. 2019	Exchange Student, University of California, Berkeley

Interests

Deep Learning, Optimization and Accelerating Neural Networks, High-Performance Computing

Honors & Awards

Dec. 2017	Second Place in Microsoft Hackathon, Xi'an Site
Oct. 2017	**The China Computer Federation (CCF) Elite Collegiate Award (0.2%)**
Aug. 2017	Best Game in 2D Game Engine Class, National University of Singapore
June. 2017	First Prize in The Mathematical Contest in Modeling, Xi'an Jiaotong University
Dec. 2016	**Silver Medal in the ACM-ICPC Asia Regional Contest China-Final 2016**
Oct. 2016	Bronze Medal in the ACM-ICPC Asia Regional Contest Dailan Site 2016
July. 2016	First Prize in Chinese Colleges Computer Competition (CCCC) - Group Programming Ladder Tournament
May. 2016	Golden Medal in the 2016 ACM-ICPC China Shaanxi Provincial Programming Contest
July. 2013	Bronze Medal in the National Olympiad in Informatics (NOI 2013)

Research & Work Experience

Aug. 2021– Jan. 2024 PhD Student in **Computer Vision and Machine Learning Group**, **National University of Singapore**

Mentored by Prof. Angela Yao.
Researching in explainable AI and interpreting deep learning.
– Reporting weakness of CKA, a popular method to measure similarity between networks and between layers. Improving CKA reliability on vision tasks.
– Using improved similarity metrics to select the most diverse models, which reach a high performance in ensembled model.
– Co-designing DropIT, a memory-saving method that drops the intermediate tensor. Proving it has a better theoretical convergence and analyzing the optimal range of the dropping rate parameter. The work is published on ICLR23
– Transferred and graduated as a Master's student.

Dec. 2020– Aug. 2021 Research Assist in **High Performance Computing for Artificial Intelligence Lab**, **National University of Singapore**

Mentored by Prof. Yang You.
Accelerating neural network training by utilizing data parallelism, contributions included:
– Improving network performance under extremely huge data batch while reducing training time. The work is uploaded on Arxiv.
– Investigating popular large batch training strategies and their combinations
– Analyzing hyper-parameters of optimizers specially designed for large batch training. Proving theoretical convergence lower bound of large batch training optimizers, such as LARS and LAMB.

May. 2019– Dec. 2019 Research Assist in **Berkeley Benchmarking and Optimization Group**, **University of California, Berkeley**

Mentored by Grace Dinh and James Demmel.
Reducing the communication cost of convolutional neural network, contributions included:
– Designing the state-of-the-art Communication Avoiding CNN algorithm and testing the practical performance based on simulated caches. Combining theoretical analyses and experiment results, the work is published on MDS20
– Improving existing communication lower bounds for convolutional neural networks by a constant factor.
– Analyzing the way cache line size affects practical communication costs.

Jan. 2018– May. 2018 Research Intern in **Visual Computing Group**, **Microsoft Research Asia**.

Mentored by Houwen Peng and Jifeng Dai.
Researching in computer vision and deep learning, main duties included:
– Updating Faster R-CNN to fit the latest version of mxnet framework and utilize dynamic graphs to effectively debug
– Converting NASNet model from tensorflow to mxnet

Publications

– Joya Chen , Kai Xu, **Yuhui Wang**, Yifei Cheng, Angela Yao. DropIT: Dropping Intermediate Tensors for Memory-Efficient DNN Training, ICLR2023
– Yang You, **Yuhui Wang**, Huan Zhang, Zhao Zhang, James Demmel, Cho-Jui Hsieh. The Limit of the Batch Size, arXiv preprint arXiv:2006.08517 (2020).
– Yang You, Jing Li, Sashank Reddi, **Yuhui Wang**, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, Cho-Jui Hsieh. Reducing BERT Pre-Training Time from 3 Days to 76 Minutes, submitted to JMLR.
– James Demmel, Grace Dinh, Ed Younis, **Yuhui Wang**. Communication-Optimal Convolutional Neural Nets, SIAM MDS2020