

Rui Pan

ruipan.xyz/ | github.com/ruipeterpan | rpan33@wisc.edu | (+1) 608-960-0303

EDUCATION

University of Wisconsin-Madison

Bachelor of Science, Computer Science and Mathematics

- GPA: 3.96/4.00
- Advisor: Prof. Shivaram Venkataraman

Madison, WI, USA

Sep 2018 – Dec 2021

RESEARCH INTERESTS

I am broadly interested in the system aspects of big data (**Machine Learning Systems, Cloud/Datacenter Systems, Distributed Systems, Networks**). For now, I work on Machine Learning Systems with a focus on optimizing systems for Deep Neural Network training in multi-tenant GPU clusters. I am applying to Ph.D. programs that start in Fall 2022.

PUBLICATIONS

[1] [Conference Paper] Fair and Efficient Cluster Scheduling for Dynamic Adaptation in Machine Learning. Pengfei Zheng, **Rui Pan**, Tarannum Khan, Shivaram Venkataraman, Aditya Akella. 2021. *Under review*.

[2] [Poster] AgDH: A System for Gathering and Distributing Dairy Data. **Rui Pan**, Steven Wangen, Michael Ferris. 2020. *Presented at the 3rd Annual WID Symposium*.

RESEARCH EXPERIENCE

Undergraduate Research Assistant with Prof. Shivaram Venkataraman

Madison, WI, USA

Project 1: Fair and Efficient Resource Allocation for DNN Training in GPU Clusters

Mar 2021 – Present

- Developed a policy to co-optimize long-term fairness and efficiency of the scheduling/resource allocation of resource-adaptive deep learning training workloads in large-scale multi-tenant GPU clusters.
- Implemented and integrated the novel allocation policy into Gavel [OSDI '20], an existing scheduling framework. Implemented the mechanism to support dynamic adaptation (e.g., batch size scaling) of training workloads in Gavel.
- Implemented dynamic optimizations, e.g., Accordion [MLSys '21] & Gradient Noise Scale [arXiv '18], for common DNN training workloads to increase the training efficiency without loss of accuracy.
- Achieved 1.3x efficiency win and 2x fairness win at the same time over state-of-the-art scheduling policies (Themis [NSDI '19], Gavel [OSDI '20], AlloX [EuroSys '20]) on a trace of real-world workloads in both large-scale simulations and physical experiments.

Project 2: How Structured Backpropagation Pruning Improves Deep Learning Clusters

Jun 2020 – Feb 2021

- In this work, we systematically control the amount of backpropagation at individual workers in distributed DNN training. This technique, Structured Backpropagation Pruning (SBP), simultaneously reduces network bandwidth, compute utilization, and memory use while preserving model quality.
- Developed an iteration-level cluster scheduler by extending existing frameworks such as PyTorch Elastic and BytePS [OSDI '20] to capitalize on the resources saved by SBP. The scheduler supports fine-grained iteration-level scheduling, different communication protocols, frequent checkpointing, and worker migration with low overhead.
- Used Microsoft Azure to develop, deploy, and modify existing code bases. Profiled common workloads to identify the communication bottlenecks in distributed DNN training and filed issue reports to open-source frameworks.

Undergraduate Research Assistant @ Wisconsin Institute for Discovery

Jan 2020 – Mar 2021

Advisors: Dr. Steven Wangen and Prof. Michael Ferris

Madison, WI, USA

- Proposed Dairy Brain, an analytics platform for evaluating and predicting the performance of dairy cows by aggregating large quantities of dairy data.

- Developed, deployed and maintained a data warehouse, Agricultural Data Hub (AgDH), for the collection, storage, homogenization, entity matching, and distribution of dairy farm's feeding, milking, and management data in a series of PostgreSQL data marts. Assisted with the implementation of the data pipeline using Apache Airflow.
- Presented our poster at the 3rd Wisconsin Institute for Discovery (WID) Research Symposium and in outreach meetings for the local dairy industry.

Incoming Research Intern @ Max Planck Institute for Informatics

Feb 2022 – May 2022

Advisor: Prof. Yiting Xia (Network-accelerated distributed machine learning)

Saarbrücken, Germany

PROFESSIONAL/VOLUNTEER EXPERIENCE

Tutor in Computer Science and Math

Sep 2019 – May 2020

Undergraduate Learning Center @ College of Engineering, UW-Madison

Madison, WI, USA

- Held weekly drop-in and by-appointment tutoring sessions for ~100 underrepresented students in STEM.
- Tutored 8 core introductory computer science and math courses by summarizing the main lecture takeaways, answering questions on the homework problem sets, and leading group discussions. Offered information and personal advice on course planning and career choices.

Subtitle Translator

May 2019 – May 2020

Coursera (Massive Open Online Course provider)

Online

- Volunteered in translating English subtitles to Simplified Chinese subtitles in multiple online courses (An Introduction to Programming, Building Web Applications in PHP, Neural Networks and Deep Learning) to expand Coursera's coverage for non-English-speaking students.

Student Instructor

Sep 2017 – Aug 2018

High School Affiliated to Shanghai Jiao Tong University

Shanghai, China

- Designed an introductory computer science course based on CS61A/B @ Berkeley and CS50 @ Harvard.
- Gave lectures and held office hours for ~10 high school students with different backgrounds.

RELEVANT COURSES

Graduate CS: Advanced Operating Systems, Big Data Systems (audited), High Performance Computing

Undergraduate CS: Algorithms, Artificial Intelligence, Bioinformatics, Computer Architecture, Computer Vision, Database Systems, Data Analysis, Networks, Operating Systems, Theory of Computing

Undergraduate Math: Calculus, Combinatorics, Discrete Math, Math in Data Science, Numerical Linear Algebra, Probability

Coursera: Cloud Computing (I/P), Deep Learning, Software Engineering, Front-end Web Design, Back-end Web Applications, Cryptography

First Aid: Anaphylaxis/Allergic Reactions, CPR/AED, Epilepsy/Seizure, General First Aid, Heimlich Maneuver/Choking

TECHNICAL SKILLS

Languages: Python, Java/C#, C/C++, SQL, JavaScript, HTML/CSS, R

Frameworks and Tools: PyTorch, Git, Docker, PostgreSQL, CUDA, OpenMP, MPI, Apache Spark, Microsoft Azure

HONORS AND AWARDS

- 2022 Research Fellowship, Max Planck Institute for Informatics (MPI-INF), Internet Architecture Group
- 2021 Participant, Cornell, Maryland, Max Planck Pre-doctoral Research School (CMMRS)
- 2018 Dean's List (6 semesters)