
EDUCATION

• University of California, San Diego

La Jolla, CA

*PhD in Computer Science; Advisor: Prof. Arun Kumar**Sept. 2017 – June. 2023(expected)*

Courses taken: Advanced Compilers, Principles of Programming Languages, Database Systems, Advanced Algorithms, Machine Learning, Data Mining & Analytics, Advanced Data Analytics, Computer Vision, Introduction to Robotics

PROFESSIONAL EXPERIENCE

• ADALab, University of California, San Diego

La Jolla, CA

*PhD research, deep learning systems**Sept. 2017 - Present***◦ Panorama:**

- * Developed the first data system for tackling the unbounded vocabulary issue (labels that were not present during training) in video querying
- * Designed an end-to-end, domain-agnostic and efficient system, allowing the user to generalize beyond bounded vocabularies without tedious manual re-training of DNNs
- * Implemented the prototype with TensorFlow. Tested with applications including face recognition, pedestrian re-identification, car model recognition, animal species recognition, etc. It can achieve between 2x to 20x higher efficiency and generalizes well to unbounded vocabularies

◦ Cerebro:

- * Developed Cerebro, a resource-efficient deep learning model selection and training system
- * Proposed a novel form of parallelism for deep learning training called model hopper. This approach improves resource-efficiency drastically
- * In terms of resource efficiency, it can be the optimal choice over the TensorFlow parameter server, Horovod, or task parallelism tools like Dask, Vizier, or Ray. Built a prototype to support both TensorFlow and PyTorch
- * Extended the system with various execution backends including Greenplum Database and Apache Spark

• Greenplum R&D at VMware, Inc.

Palo Alto, CA

*Software Engineer Intern, in-database deep learning**Summer 2019*

- Worked on the first in-DBMS deep learning system, allowing training and inference of deep learning models with TensorFlow on database-resident data
- Integrated my research project, Cerebro, into the deep learning training infrastructure of Greenplum, boosting the efficiency by 10x
- Contributed to Apache MADlib project. Lead the development of a major release
- This project has been shipped to VMware's enterprise customers for use

• University of California, San Diego

San Diego, CA

*Teaching Assistant of Course: Systems for Scalable Analytics**Winter 2019*

- TAed the course which is about large-scale data analytics with Big Data tools and machine learning
- Developed the first edition of course assignments and auto-grading programs, involving Dask, Spark, AWS EC2/S3/EBS, and Kubernetes. These assignments have been adopted by the course ever since and used by 500+ students

RECENT PUBLICATIONS

1. A. Kumar, S. Nakandala, **Y. Zhang**, S. Li, A. Gemawat, and K. Nagrecha Cerebro: A Layered Data Platform for Scalable Deep Learning, CIDR (2021)
2. S. Nakandala, **Y. Zhang**, and A. Kumar Cerebro: a data system for optimized deep learning model selection, PVLDB (2020)
3. **Y. Zhang** and A. Kumar, *Panorama: A Data System for Unbounded Vocabulary Querying over Video*, PVLDB (2020)
4. S. Nakandala, **Y. Zhang**, and A. Kumar, *Cerebro: Efficient and Reproducible Model Selection on Deep Learning Systems*, ACM SIGMOD DEEM Workshop (2019)