

# Yuhao Zhang

---

CONTACT INFORMATION	EBU3B 3230 CSE Department UC San Diego La Jolla, CA 92093 USA <i>E-mail:</i> yuz870@eng.ucsd.edu <i>Website:</i> <a href="https://yhzhang.info/">https://yhzhang.info/</a>
FULL CV	<a href="https://github.com/makemebitter/yuhao-cv/raw/master/CV-Yuhao-UCSD.pdf">https://github.com/makemebitter/yuhao-cv/raw/master/CV-Yuhao-UCSD.pdf</a>
EDUCATION	<b>University of California, San Diego</b> , La Jolla, California USA Ph.D. Student, Computer Science, September 2017 - Spring 2023 <ul style="list-style-type: none"><li>• Dissertation Topic: “High-throughput Data Systems for Deep Learning Workloads”</li><li>• Advisor: Prof. Arun Kumar</li></ul> <b>Nankai University</b> , Tianjin, China B.S., Theoretical Physics, June 2016 <ul style="list-style-type: none"><li>• Advisor: Prof. Xueqian Li and Prof. Yangang Miao</li></ul>
RESEARCH INTERESTS AND IMPACT	My research is primarily on machine learning systems and large-scale data analytics systems, including systems powered by applied ML that enable novel applications and systems designed for ML to make data science easier and faster. Past and ongoing projects include systems for: distributed deep learning model selection and training, distributed in-database deep learning, distributed large-scale graph neural network training, and video analytics and querying. All of my previous work is released as open source software. My past research on Cerebro and Cerebro-DS have been incorporated into the Apache MADlib open-source project and offered in Greenplum Database by VMware. The same projects have also been integrated with Spark and Databricks is also reviewing the same project to offer to their customers. A prominent graph DBMS vendor is also interested in my work of Lotan.
ACADEMIC EXPERIENCE	<b>University of California, San Diego</b> , La Jolla, California USA <i>PhD Student</i> <b>Sept 2017 - present</b> Ph.D. research, Ph.D. and Masters level coursework and research/consulting projects. Courses taken: Machine Learning, Deep Learning, Data Mining & Analytics, Advanced Data Analytics, Computer Vision, Database Systems, Advanced Algorithms, Advanced Compilers, Principles of Programming Languages, Introduction to Robotics  <i>Collaborator for project DeepPostures: a deep learning library for identifying human postures from wearable devices data</i> <b>Spring 2022 - Spring 2023</b> Built and maintained project website ( <a href="https://adalabucsd.github.io/DeepPostures/">https://adalabucsd.github.io/DeepPostures/</a> ). Wrote documentation and made demos about the project. Provided software engineering support for other public health researchers. Conducted tests and data analysis for the project and ran preliminary experiments for grant applications.  <i>Teaching Assistant for CSE234: Data Systems for Machine Learning</i> <b>Winter 2021 &amp; Winter 2023</b> Helped designing and supervising a research-oriented course on machine learning systems. Mentored 12 master students with their course projects ranging from advanced implementation of cutting-edge research, to evaluation and surveying the state-of-art work, and to open-ended research. Did stand-in lectures.  <i>Teaching Assistant for DSC102: Systems for Scalable Analytics</i> <b>Winter 2020</b> Developed the first edition of course assignments and auto-grading programs with Python and

Bash. The assignments involve Python Dask, Spark, AWS EC2/S3/EBS, and Kubernetes. These assignments have been adopted by the course ever since and used by 500+ students.

**Texas A&M University**, College Station, Texas USA

*Research Intern*

**Summer 2015**

Worked on vision-based object tracking, modeling, and data analytics.

**Institute of Physics, Chinese Academy of Sciences**, Beijing, China

*Research Intern*

**Summer 2014**

Theoretical physics research. Particle physics.

**Nankai University**, Tianjin, China

*Research Assistant*

**2013 - 2016**

Theoretical physics research. Black holes, gravity, and neutrinos. Resulted in two published papers.

PROFESSIONAL  
EXPERIENCE

**Microsoft Gray System Lab**, California USA

*Research Intern*

**Summer 2021**

Worked on machine learning system research, focusing on factorized in-DBMS machine learning. Built a highly scalable system on top of Apache Spark that can outperform its MLlib by orders of magnitude.

**VMware**, Palo Alto, California USA

*Software Engineer Intern*

**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 Database, boosting efficiency by over 10x. Contributed to the Apache MADlib project in Python and SQL. Lead the development of a major release. This project has been incorporated into Greenplum and production-ready for VMware's customers.

**Opera Solutions**, San Diego, California USA

*Data Scientist Intern*

**Summer 2018**

Worked on a theatre scheduling & recommender system. Proposed new models and optimized the existing system.

PUBLICATIONS

See Full CV.

PRESENTATIONS

See Full CV.

SERVICE

**External reviewer:** SIGMOD 2020, SIGMOD 2021, VLDB 2022

**Reviewer:** JMLR MLOSS 2022, SIGMOD 2024

MISC

SIGMOD 2021 *Students and Postdocs in DB Panel Discussion*

**June 2021**

HONORS AND  
AWARDS

Best Thesis Award, School of Physics, Nankai University, 2016

Wang Kechang Scholarship for Academic Distinction, 2014

Gong-Neng Scholarship, 2013

Poling Academy Scholarship, 2012-2016