Jiawei Zhao

Ph.D. student, Department of Computing + Mathematical Sciences   jiawei@caltech.edu
California Institute of Technology   jiawei-zhao.netlify.com
1200 E California Blvd
Pasadena, CA 91125

EDUCATION

- **California Institute of Technology**
  Ph.D. student, Computing and Mathematical Sciences   *Sep. 2019 - Present*

- **Nanjing University of Aeronautics and Astronautics**
  BE, Computer Science and Technology   *Sep. 2015 - Jun. 2019*

PUBLICATIONS

- **Jiawei Zhao**, Florian Tobias Schaefer, and Anima Anandkumar. 'ZerO Initialization: Initializing Neural Networks with Only Zeros and Ones'. Transactions on Machine Learning Research, 2022.

- **Jiawei Zhao**, Steve Dai, Rangharajan Venkatesan, Brian Zimmer, Mustafa Ali, Ming-Yu Liu, Brucek Khailany, William J. Dally, and Anima Anandkumar. 'LNS-Madam: Low-Precision Training in Logarithmic Number System Using Multiplicative Weight Update'. IEEE Transactions on Computers, 2022, 1–12.

- Jeremy Bernstein, **Jiawei Zhao**, Markus Meister, Ming-Yu Liu, Anima Anandkumar, and Yisong Yue. 'Learning Compositional Functions via Multiplicative Weight Updates'. In Advances in Neural Information Processing Systems, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, 33:13319–30. Curran Associates, Inc., 2020.

- Jeremy Bernstein[*], **Jiawei Zhao**[*], Kamyar Azizzadenesheli, and Anima Anandkumar. 'SignSGD with Majority Vote Is Communication Efficient and Fault Tolerant'. In International Conference on Learning Representations, 2019.

- Shengjun Huang, **Jiawei Zhao**, Zhaoyang Liu, *"Cost-Effective Training of Deep CNNs with Active Model Adaptation"*, Proc. of 2018 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining **(KDD'18)**, London, United Kingdom, August 2018

RESEARCH EXPERIENCE

- **Research Internship**
  NVIDIA   *June.2021 - Present*

  - Proposed LNS-Madam low-precision training framework that co-designs Madam learning algorithm and Logarithmic number system to achieve high accuracy with low precision.
  - Demonstrated that the framework achieves comparable accuracy to full-precision counterparts with only 8 bits on popular tasks. Compared to FP32 and FP8, it reduces the energy consumption by over 90% and 55%, respectively.

- **Visiting Undergraduate Researcher**
  California Institute of Technology   *July.2018 - Oct.2018*
  - Proposed a novel distributed training system with signSGD that can reduce time cost of communication between machines and accelerate convergence performance.
  - Realized a 25% reduction in time for training ResNet-50 on ImageNet when using 15 AWS p3.2xlarge machines, compared with the state-of-the-art "Nvidia Collective Communications Library" (NCCL).

AWARDS AND HONORS

First Prize Innovation Award, NUAA 2018
Principal Excellent Scholarship, NUAA 2018