Jack Kosaian

jackkosaian.github.io ⋄ jkosaian@cs.cmu.edu

Education

Carnegie Mellon University Aug. 2017 - Present
Ph.D. in Computer Science
Thesis topic: Resource-efficient and reliable machine learning systems

University of Michigan, Ann Arbor Sept. 2013 - Dec. 2016
B.S.E. in Computer Science & Engineering

Awards

TCS Presidential Fellowship (2021)
NSF Graduate Research Fellowship (2017)
Angell Scholar (2015)
Branstrom Prize (2014)

Publications

Arithmetic-Intensity-Guided Fault Tolerance for Neural Network Inference on GPUs
Jack Kosaian, K. V. Rashmi
International Conference on High Performance Computing, Networking, Storage and Analysis (SC), 2021

Boosting the Throughput and Accelerator Utilization of Specialized CNN Inference Beyond Increasing Batch Size
Jack Kosaian, Amar Phanishayee, Matthai Philipose, Debadeepta Dey, K. V. Rashmi
International Conference on Machine Learning (ICML), 2021

ECRM: Efficient Fault Tolerance for Recommendation Model Training via Erasure Coding
Kaige Liu*, Jack Kosaian*, K. V. Rashmi
Preprint: arXiv:2104.01981 (*Equal contribution)

Learning-Based Coded Computation
Jack Kosaian, K. V. Rashmi, Shivaram Venkataraman
IEEE Journal on Selected Areas in Information Theory (JSAIT), 2020

Parity Models: Erasure-Coded Resilience for Prediction Serving Systems
Jack Kosaian, K. V. Rashmi, Shivaram Venkataraman
ACM Symposium on Operating Systems Principles (SOSP), 2019

Vantage: Optimizing Video Upload for Time-shifted Viewing of Social Live Streams
Devdeep Ray, Jack Kosaian, K. V. Rashmi, Srinivasan Seshan
ACM SIGCOMM 2019

EC-Cache: Load-Balanced, Low-Latency Cluster Caching with Online Erasure Coding
K. V. Rashmi, Mosharaf Chowdhury, Jack Kosaian, Ion Stoica, Kannan Ramchandran
USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2016

Industry Experience

Microsoft Research, *Research Intern* June 2021 - Sep. 2021
Mentor: Amar Phanishayee
- Researched techniques to design more hardware-efficient CNNs in various settings

Microsoft Research, *Research Intern* May 2019 - Aug. 2019
Mentor: Amar Phanishayee
- Researched strategies to improve the throughput and hardware utilization of CNNs on accelerators

Google, *Software Engineering Intern* May 2017 - July 2017
- Analyzed performance and scalability bottlenecks of BigQuery high-throughput read/write API

Google, *Software Engineering Intern* May 2016 - Aug. 2016
- Explored hardware virtualization extensions for gVisor sandboxing system

Epic Systems, *Software Development Intern* May 2015 - Aug. 2015
- Developed dashboard for physicians to explore changes in patient health

Skills

Languages: C++, Python, CUDA
Frameworks: PyTorch, TensorFlow, TensorRT

Teaching Experience

CMU 15-712: Advanced Operating Systems and Distributed Systems Spring 2021
Teaching Assistant

CMU 15-440: Distributed Systems Spring 2020
Head Teaching Assistant

UofM EECS 370: Introduction to Computer Organization Fall 2015, Winter 2016
Teaching Assistant

Outreach

CMU SCS Creative Technology Nights
- Assist in teaching STEM concepts to middle school girls in the Pittsburgh area