Sunwoo Lee

330 W 11th ST APT C09
Los Angeles CA, 90015

sunwool@usc.edu
+1) 224-999-5923 https://sites.google.com/view/sunwoolee

Education

Northwestern University, USA Sep 2020
Ph.D. in Computer Engineering
Advisors: Prof. Alok Choudhary and Prof. Wei-keng Liao
Thesis: Scalable parallelization strategy for large-scale deep learning

Hanyang University, Seoul, South Korea Feb 2009
M.S. in Computer Engineering

Hanyang University, Seoul, South Korea Feb 2007
B.S. in Computer Engineering

Research Experience

University of Southern California Oct 2020 – Present
Postdoctoral Researcher

Lawrence Berkeley National Laboratory Jun 2020 – Aug 2020
Research Intern

Fermi National Laboratory Jul 2019 – Sep 2019
Research Intern

Argonne National Laboratory Jun 2018 – Aug 2018
W. J. Cody Associate (Research Intern)

Professional Experience

Samsung Electronics, Memory Solutions Lab. May 2013 – Jan 2015
Software researcher

Humax (*alternative military service*) Feb 2009 – Mar 2013
Software developer

Teaching Experience

University of Southern California, Mentor
- AEOP Summer Scholarship Program: Data Science — Summer 2021

Northwestern University, Teaching Assistant
- CE303: Advanced Digital Design — Fall 2019
- CE510: Social Media Mining — Spring 2020

Honors and Awards

Best Paper Award Finalist — 2017
- Parallel Deep Convolutional Neural Network Training by Exploiting the Overlapping of Computation and Communication, IEEE HiPC 2017

Travel Grants
- IEEE International Conference on BigData 2019 Travel Grant — 2019
- Northwestern TGS Travel Grant — 2016, 2018, 2019

Publications

1. **Sunwoo Lee**, Qiao Kang, Reda Al-Bahrani, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao, Improving Scalability of Parallel CNN Training by Adaptively Adjusting Parameter Update Frequency, Journal of Distributed and Parallel Computing, 2021

2. **Sunwoo Lee**, Qiao Kang, Kewei Wang, Jan Balewski, Alex Sim, Kesheng Wu, Ankit Agrawal, Alok Choudhary, Peter Nugent, and Wei-keng Liao, Asynchronous I/O Strategy for Large-Scale Deep Learning Applications. International Conference on High-Performance Computing, Data, and Analytics (HiPC), December 2021

3. Yue Niu, Zalan Fabian, **Sunwoo Lee**, Mahdi Soltanolkotabi, and Salman Avestimehr, SLIM-QN: A Stochastic Light, Momentumized, Quasi-Newton Optimizer for Deep Neural Networks, International Conference on Machine Learning workshop, 2021

4. Reda Al-bahrani, Dipendra Jha, Qiao Kang, **Sunwoo Lee**, Zijiang Yang, Wei-keng Liao, Ankit Agrawal, and Alok Choudhary, SIGRNN: Synthetic minority Instances Generation in imbalanced datasets using a Recurrent Neural Network, International Conference on Pattern Recognition Applications and Methods, February 2021

5. **Sunwoo Lee**, Qiao Kang, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao, Communication-Efficient Local SGD for Scalable Deep Learning. IEEE International Conference on BigData, December 2020 (15.7%)

6. Sandeep Madireddy, Ji Hwan Park, **Sunwoo Lee**, Prasanna Balaprakash, Shinjae Yoo, Wei-keng Liao, Cory Hauck, M. Paul Laiu, and Richard Archibald, In Situ Compression Artifact

Removal in Scientific Data Using Deep Transfer Learning and Experience Replay. Machine Learning: Science and Technology 2020

7. Qiao Kang, **Sunwoo Lee**, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao, Improving All-to-many Personalized Communication in MPI I/O. International Conference for High Performance Computing, Networking, Storage, and Analysis (SC) 2020

8. Qiao Kang, **Sunwoo Lee**, Kai-yuan Hou, Robert Ross, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao, Improving MPI Collective I/O for High Volume Non-contiguous Requests with Intra-node Aggregation. IEEE Transactions on Parallel and Distributed Systems 2020

9. Qiao Kang, Alex Sim, Peter Nugent, **Sunwoo Lee**, Wei-keng Liao, Ankit Agrawal, Alok Choudhary, and Kesheng Wu. Predicting Resource Requirement in Intermediate Palomar Transient Factory Workflow. International Symposium on Cluster, Cloud and Internet Computing (CCGrid) 2020

10. **Sunwoo Lee**, Qiao Kang, Sandeep Madireddy, Prasanna Balaprakash, Ankit Agrawal, Alok Choudhary, Richard Archibald, and Wei-keng Liao. Improving Scalability of Parallel CNN Training by Adjusting Mini-Batch Size at Run-Time. IEEE International Conference on BigData, December 2019 (18.7%)

11. **Sunwoo Lee**, Ankit Agrawal, Prasanna Balaprakash, Alok Choudhary, and Wei-keng Liao. Communication-Efficient Parallelization Strategy for Deep Convolutional Neural Network Training. In Workshop on Machine Learning in High-Performance Computing Environments (MLHPC), held in conjunction with International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), November 2018

12. **Sunwoo Lee**, Dipendra Jha, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao. Parallel Deep Convolutional Neural Network Training by Exploiting the Overlapping of Computation and Communication. International Conference on High-Performance Computing, Data, and Analytics (HiPC), December 2017 (22.8%)

13. **Sunwoo Lee**, Wei-keng Liao, Ankit Agrawal, Nikos Hardavellas, and Alok Choudhary. Evaluation of K-Means Data Clustering Algorithm on Intel Xeon Phi. In Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery, held in conjunction with the IEEE International Conference on BigData, December 2016

14. Diana Palsetia, William Hendrix, **Sunwoo Lee**, Ankit Agrawal, Wei-keng Liao, and Alok Choudhary. Parallel Community Detection Algorithm Using a Data Partitioning Strategy with Pairwise Subdomain Duplication. In the 31st International Conference on High Performance Computing (ISC), June 2016

Preprints

1. **Sunwoo Lee**, Kai-yuan Hou, Kewei Wang, Saba Sehrish, Marc Paterno, James Kowalkowski, Quincey Koziol, Robert B. Ross, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao, A Case Study on Parallel HDF5 Dataset Concatenation for High-Energy Physics Data Analysis. (*Under review by Elsevier Parallel Computing*)

2. Chaoyang He, Zhengyu Yang, Erum Mushtaq, **Sunwoo Lee**, Mahdi Soltanolkotabi, Salman Avestimehr, SSFL: Tackling Label Deficiency in Federated Learning via Personalized Self-Supervision, arXiv 2021

Invited Talks

Asynchronous I/O Strategy for Large-Scale Deep Learning Applications
U.S. Department of Energy, SciDAC, RAPIDS Institute, Tech Talk, 12/01/2021

A Case Study on Parallel HDF5 Dataset Concatenation for Scientific Data Analysis
HDF5 User Group Annual Meeting, 10/12/2021

Communication-Efficient Local SGD for Scalable Deep Learning
U.S. Department of Energy, SciDAC, RAPIDS Institute, Tech Talk, 7/7/2021

Skills and Qualifications

Programming Language:
C/C++, Python

Deep Learning Software Framework:
Caffe, TensorFlow, Horovod, PyTorch

Parallelization Library:
MPI, OpenMP, Pthreads

Compiling, Debugging, and Analyzer:
GNU and Intel compilers, Intel VTune

I/O Library:
MPI-IO (ROMIO), HDF5, PNetCDF

Contributions to Open-Source Software

- [PCNN]: A software framework for distributed Convolutional Neural Network training
- [ph5concat]: A parallel HDF5 file concatenation program for large-scale scientific data
- [ROMIO]: A module for pipelined two-phase I/O on Lustre parallel file system