Sunwoo Lee

Postdoctoral Researcher
Department of Electrical and Computer Engineering
University of Southern California

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

Research Interests	1. Scalable Deep Learning algorithms for large-scale applications 2. Communication-efficient model aggregation in Federated Learning 3. Applied Deep Learning in domain problems	
Education	Northwestern University, USA Ph.D. in Computer Engineering Advisors: Prof. Alok Choudhary and Prof. Wei-keng Liao	Sep 2020
	Hanyang University, Seoul, South Korea M.S. and B.S. in Computer Engineering Advisor: Prof. Minsoo Ryu	Feb 2009
Research Experience	University of Southern California Postdoctoral Researcher Advisor: Prof. Salman Avestimehr	Oct 2020 – Present
	Lawrence Berkeley National Laboratory Research Intern	Jun 2020 – Aug 2020
	Fermi National Laboratory Research Intern	Jul 2019 – Sep 2019
	Argonne National Laboratory W.J.Cody Associate (Research Intern)	Jun 2018 – Aug 2018
Professional Experience	Samsung Electronics, Memory Solutions Lab. Software Researcher	May 2013 – Jan 2015
	Humax, Software Lab. Software Engineer (alternative military service)	Feb 2009 – Mar 2013
Teaching Experience	University of Southern California, Mentor • AEOP Scholarship Program in Data Science	Summer 2021
	Northwestern University, Teaching Assistant • CE303: Advanced Digital Design • CE501: Social Media Mining	Fall 2019 Spring 2020
Honors & Awards	IEEE HiPC Best Paper Award Finalist, 2017 • Parallel Deep Convolutional Neural Network Training by Exploiting the Overlapping of Computation and Communication	

IEEE BigData Travel Grant, 2019

- Improving Scalability of Parallel CNN Training by Adjusting Mini-Batch Size at Run-Time

---

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, 2022

2. **Sunwoo Lee**, Kai-yuan Hou, Kewei Wang, Saba Sehrish, Marc Paterno, James Kowalkowski, Quincey Koziol, Ross Robert, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao, A Case Study on Parallel HDF5 Dataset Concatenation for High-Energy Physics Data Analysis. Elsevier Parallel Computing, 2022

3. Kai-yuan Hou, Qiao Kang, **Sunwoo Lee**, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao, Supporting Data Compression in PnetCDF, IEEE International Conference on BigData, December 2021 (19.9%)

4. **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 (22.9%)

5. 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

6. 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

7. **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%)

8. 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

9. 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

10. 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

11. 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

12. **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%)

13. **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

14. **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%)

15. **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

16. 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

17. **Sunwoo Lee**, Byung Kwan Jung, Minsoo Ryu, Seungwon Lee, Extending Component-based Approaches for Multi-threaded Design of Multiprocessor Embedded Software. IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing 2009

Preprints

1. **Sunwoo Lee**, Tuo Zhang, Chaoyang He, and Salman Avestimehr, Layer-wise Model Aggregation for Scalable Federated Learning. *arXiv 2021* (*Under review by ML top-tier conference*)

2. **Sunwoo Lee**, Salman Avestimehr, Partial Model Aggregation in Federated Learning: Performance Guarantees. (*Under review by ML top-tier conference*)

3. **Sunwoo Lee**, Salman Avestimehr, Two-Phase Large-Batch Training for Scalable Deep Learning. (*Under review by ML top-tier conference*)

4. 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* (*Under review by ML top-tier conference*)

5. **Sunwoo Lee**, Jaeyoung Jeon, Kitae Eom, Chaehwa Jeong, Yongsoo Yang, Ji-Yong Park, Chang Beom Eom, and Hyungwoo Lee, Multi-level Memristors based on Two-dimensional Electron Gases in Oxide Heterostructures for High-Precision Neuromorphic Computing. (*Under review by Nature Communications*)

6. Kewei Wang, **Sunwoo Lee**, Alex Sim, Ankit Agrawal, Alok Choudhary, Kesheng Wu, and Wei-keng Liao, Using Mixed-Resolution Data for Neural Network Training in Deep Learning Applications. (*Under review by ML top-tier conference*)

Invited Talks

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

- HDF5 User Group Meeting: A Case Study on Parallel HDF5 Dataset Concatenation for Scientific Data Analysis, 10/21/2021

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

Skills & Qualifications

Programming Languages
C/C++, Python

Deep Learning Software Frameworks
TensorFlow, PyTorch, Caffe

Parallelization Libraries
MPI, OpenMP, Pthreads

I/O Libraries
MPI-I/O (ROMIO), HDF5, NetCDF

References

Alok Choudhary
Henry and Isabelle Dever Professor
Department of Electrical and Computer Engineering
Northwestern University, IL, USA
Email: a-choudhary@northwestern.edu
Phone: +1-847-467-4129

Salman Avestimehr
Dean's Professor
Department of Electrical and Computer Engineering
University of Southern California, CA, USA
Email: avestime@usc.edu
Phone: +1-213-740-7326

Wei-keng Liao
Research Professor
Department of Electrical and Computer Engineering
Northwestern University, IL, USA
Email: wkliao@northwestern.edu
Phone: +1-847-491-2916