Sunwoo **Lee**

Assistant Professor
Department of Computer Engineering
Inha University, South Korea

Tel: (+82) 032-860-7445
Email: sunwool@inha.ac.kr
https://sites.google.com/view/sunwoolee

Research Interest

- Large-Scale Distributed Deep Learning
- Communication-efficient Federated Learning on heterogeneous devices
- Applied Machine Learning (E.g., Physics + ML)

Education

Northwestern University — Evanston, IL USA
Ph.D. in Computer Engineering — 2020
Advisors: Prof. Alok Choudhary and Prof. Wei-keng Liao
Thesis: Scalable Parallelization Strategy for Large-Scale Deep Learning

Hanyang University — Seoul, South Korea
B.S. and M.S. in Computer Engineering — 2009
Advisor: Prof. Minsoo Ryu

Employment

Inha University — Incheon, South Korea
Assistant Professor of Computer Engineering — 2022 – now

University of Southern California — Los Angeles, CA USA
Postdoctoral Researcher — 2020 – 2022
Advisor: Prof. Salman Avestimehr

Samsung Electronics, Memory Solutions Lab. — Hwaseong, South Korea
System Software Researcher — 2013 –2015

Humax — Bundang, South Korea
Software Engineer (Alternative Military Service) — 2009 –2013

Research Internship

Lawrence Berkeley National Laboratory — Berkeley, CA USA
Research Intern — Jun 2020 – Aug 2020

Fermi National Accelerator Laboratory — Batavia, IL USA
Research Intern — Jul 2019 – Sep 2019

Argonne National Laboratory — Lemont, IL USA
Research Intern — Jun 2018 – Aug 2018

Teaching Experience

Inha University
- CSE3209: System Programming — Fall 2022
- CSE3313: Linux Programming — Fall 2022

University of Southern California
- AEOP Scholarship Program in Data Science — Summer 2021

Northwestern University
- CE501: Social Media Mining — Spring 2020
- CE303: Advanced Digital Design — Fall 2019

Honors & Awards

FL-AAAI Workshop Best Paper Award — 2022
- SSFL: Tackling Label Deficiency in Federated Learning via Personalized Self-Supervision

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

Publications

1. Sunwoo Lee, Chaoyang He, and Salman Avestimehr, Achieving Small-Batch Accuracy with Large-Batch Scalability via Hessian-Aware Learning Rate Adjustment. *Elsevier Neural Networks* — 2023

2. Sunwoo Lee, Jaeyong Jeon, and Hyungwoo Lee, Probing Oxygen Vacancy Distribution in Oxide Heterostructures by Deep Learning-based Spectral Analysis of Current Noise. *Applied Surface Science*, p154599 — 2022

3. Sunwoo Lee, Jaeyoung Jeon, Kitae Eom, Chaehwa Jeong, Yongsoo Yang, Ji-Yong Park, Chang Beom Eom, and Hyungwoo Lee, Variance-aware Weight Quantization of Multi-level Resistive Switching Devices based on Pt/LaAlO3/SrTiO3 Heterostructures. *Scientific Reports*, 12, 1-10 2022

4. Kewei Wang, Sunwoo Lee, Jan Balewski, Alex Sim, Peter Nugent, Ankit Agrawal, Alok Choudhary, Kesheng Wu, and Wei-keng Liao, Using Multi-resolution Data to Accelerate Neural Network Training in Scientific Applications. *International Symposium on Cluster, Cloud and Internet Computing (CCGrid)* 2022

5. 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*, 159, 10-23 2022

6. 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. *Parallel Computing*, 110, 102877 2022

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

8. 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)* (22.9%) 2021

9. 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* 2021

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

11. 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*, 2, 025010 2020

12. 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

13. 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 2020

Requests with Intra-node Aggregation. *IEEE Transactions on Parallel and Distributed Systems*, 31, 11, 2682-2695

14. 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 — 2020

15. 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. *International Conference on Big Data* (18.7%) — 2019

16. Sunwoo Lee, Ankit Agrawal, Prasanna Balaprakash, Alok Choudhary, and Wei-keng Liao. Communication-Efficient Parallelization Strategy for Deep Convolutional Neural Network Training. *Machine Learning in High-Performance Computing Environments (MLHPC)* — 2018

17. 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)* (22.8%) — 2017

18. Sunwoo Lee, Wei-keng Liao, Ankit Agrawal, Nikos Hardavellas, and Alok Choudhary. Evaluation of K-Means Data Clustering Algorithm on Intel Xeon Phi. *International Conference on Big Data* — 2016

19. 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. *International Conference on High Performance Computing (ISC)* — 2016

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

Pre-prints

1. Tuo Zhang, TianTian Feng, Samiul Alam, **Sunwoo Lee**, Mi Zhang, Shrikanth S. Narayana, Salman Avestimehr, FedAudio: A Federated Learning Benchmark for Audio Tasks. *arXiv* 2022 (*Under review by a top-tier ML conference*).

2. Yue Niu, Saurav Prakash, Souvik Kundu, **Sunwoo Lee**, Salman Avestimehr, Federated Learning of Large Models at the Edge via Principal Sub-Model Training. *FL-NeurIPS* 2022.

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

4. Sunwoo Lee, Anit Sahu, Chaoyang He, and Salman Avestimehr, Partial Model Aggregation in Federated Learning: Performance Guarantees. *arXiv* 2022 *(Under review in Neurocomputing).*

5. 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

- Department of Computer Engineering at Gachon University, South Korea: Scalable Federated Learning Strategies on Real-world Edge Computing Environments, 11/18/2022
- Department of Electrical Engineering at Hanyang University, South Korea: Partial Model Training Strategies in Federated Learning, 10/04/2022
- 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

Service

- Program Committee of Association for the Artificial Intelligence (AAAI) 2021, 2022, 2023
- Program Committee of Federated Learning for Natural Language Processing (FL4NLP) workshop held in conjunction with Association for Computational Linguistics (ACL) 2022
- Program Committee of International Conference on Learning Representations (ICLR) 2021, 2022
- Program Committee of International Conference on Machine Learning (ICML) 2021, 2022
- Program Committee of International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, 2022

Skills and Qualifications

Programming Language
C/C++, Python

Deep Learning Software Frameworks
TensorFlow, PyTorch, Caffe

Parallelization Libraries
MPI, OpenMP

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