SangLyul Cho

E-mail: chosanglyul@gmail.com / chosanglyul@snu.ac.kr

Co-designing efficient deep learning algorithm and software
My primary interest lies in machine learning (ML) and the acceleration of large-scale deep learning (DL) models. I have a diverse background in ML, systems, competitive programming, and web development. Having observed the escalating growth and complexity of DL models, I've developed an interest in tackling the challenges of efficient training and serving these models. To deepen my understanding and explore these interests further, I am double majoring in math.

Education

Bachelor's degree in CSE and Math — *Seoul National University*
Current GPA: 4.27/4.3, CSE: 4.27/4.3, Math: 4.30/4.3 — *Mar 2022 - Present*

High School Diploma — *Seoul Science High School*
Final GPA: 4.24/4.3 — *Mar 2019 - Feb 2022*
Senior project: Improvement of backpropagation time and stability of RNN using Tree-RNN

Publications

Any-Precision LLM: Low-Cost Deployment of Multiple, Different-Sized LLMs
Yeonhong Park, Jake Hyun, **SangLyul Cho**, Bonggeun Sim, Jae W. Lee
The 41st *International Conference on Machine Learning* (**ICML**), 2024

- **Selected as oral presentation**, GitHub Repository
- Designed and implemented the Table Lookup Merge (TLM) technique to enhance the efficiency of the quantized GEMV, achieving **up to 19% faster** GEMV time through optimization
- Developed an efficient CUDA kernel with tiling for batched and quantized GEMV

VGA: Hardware Accelerator for Scalable Long Sequence Model Inference
SeungYul Lee, Jihoon Hong, Hyunseung Lee, **SangLyul Cho**, Jae W. Lee
The 57th *IEEE/ACM International Symposium on Microarchitecture* (**MICRO**), 2024

- Implemented and profiled LLMs in NVIDIA GPU and Google Cloud TPU

Work and Research experience

ARC Lab — *Jul - Aug 2023 / Jan - Feb 2024*
Intern — *Seoul, Korea*

- Published two papers under the supervision of Prof. Jae W. Lee.

Samsung Electronics — *Jul 2022 - Aug 2022*
Summer Intern, SW Development Team in Memory Business — *Hwaseong, Korea*

- Improved performance of SSD(Solid State Drive) using machine learning

Scholarships and Awards

Semiconductor Track Scholarship (4000+ USD)	2024 - Present
Presidential Science Scholarship (30000+ USD)	2022 - Present
ICPC Seoul Regional, Fifth Award(12th place)	2023
SCPC 2023, Fifth Award	2023
UCPC 2023, Fourth Award(9th place)	2023
2023 Undergraduate Deep Learning Product Recognition Contest, Excellence Prize(2nd place)	2023
ICPC Seoul Regional, Fifth Award(13th place)	2022
Korea Olympiad in Informatics, Qualifier Gold Prize(4th place), Finals Silver Prize(8th place)	2021
Korea Olympiad in Informatics, Qualifier Bronze Prize, Finals Bronze Prize	2019

Technical skills

Proficient	C, C++, Python, PyTorch, TensorFlow, Git, LaTeX, CUDA
Experienced	Rust, Verilog, Bluespec System Verilog, Java, Kotlin, Javascript, Docker

Relevant coursework

Computer Science	Computer architecture, System programming, Operating system, Automata Theory, Compiler, High-performance computing
Mathematics	Linear algebra*, Calculus, Abstract algebra 1, Introduction to Analysis 1, Real analysis
Statistics	Mathematical statistics

*Studied in Seoul Science High School

†Currently studying