

Geonmo Gu

✉ gmgu@theory.snu.ac.kr

🌐 <https://github.com/gmgu>

RESEARCH INTERESTS

My primary research efforts have been devoted to developing fast algorithms. I have developed fast algorithms for graph isomorphism, graph isomorphism query processing, and multiple pattern Cartesian tree matching during my Ph.D. study. At LG Electronics, I have developed AI coding assistant using large language model (LLM).

WORK EXPERIENCE

LG Electronics

Senior Researcher

Seoul, Korea

Apr. 2022 – Present

Artificial Intelligence Lab

- Aug. 2022 – Present: Fine-tuning large language models (of size over 15 billion) for AI coding assistance using PyTorch FSDP and DeepSpeed (for full parameter fine-tuning), and LoRA (for parameter-efficient fine-tuning) on AWS SageMaker. Development of data augmentation techniques for causal language models. Inference server development for large language model using NVIDIA Triton, FastTransformer, and FastAPI. Inference optimization for transformer models using flash attention and paged attention.
- Apr. 2022 – Dec. 2022: Training a small language model (from scratch) for AI coding assistance using PyTorch. Web UI development using Python Streamlit. Inference server development for language models using Flask and ShannonAI/service-streamer.

Seoul National University

Post-Doctoral Assistant

Seoul, Korea

Jan. 2022 – Mar. 2022

Institute of Computer Technology

- Jan. 2022 – Mar. 2022: Algorithm development for graph isomorphism query processing (Efficient Graph Isomorphism Query Processing using Degree Sequences and Color-Label Distributions, IEEE ICDE 2022).

EDUCATION

Seoul National University

Ph.D. in Computer Science and Engineering

Seoul, Korea

Mar. 2014 – Aug. 2021

- Thesis: Fast Graph Isomorphism using Pairwise Color Refinement and Efficient Backtracking
- Advisor: Prof. Kunsoo Park
- GPA: 3.99/4.3

Incheon National University

B.S. in Computer Science and Engineering

Incheon, Korea

Mar. 2010 – Feb. 2014

- GPA: 4.4/4.5 (summa cum laude)

PUBLICATIONS

Geonmo Gu, Yehyun Nam, Kunsoo Park, Zvi Galil, Giuseppe F. Italiano, and Wook-Shin Han. “Efficient Graph Isomorphism Query Processing using Degree Sequences and Color-Label Distributions.” *IEEE International Conference on Data Engineering*, 2022.

Geonmo Gu, Yehyun Nam, Kunsoo Park, Zvi Galil, Giuseppe F. Italiano, and Wook-Shin Han. “Scalable Graph Isomorphism: Combining Pairwise Color Refinement and Backtracking via Compressed Candidate Space.” *IEEE International Conference on Data Engineering*, 2021.

Siwoo Song, **Geonmo Gu**, Cheol Ryu, Simone Faro, Thierry Lecroq, and Kunsoo Park. “Fast Algorithms for Single and Multiple Pattern Cartesian Tree Matching.” *Theoretical Computer Science*, 2020.

Geonmo Gu, Siwoo Song, Simone Faro, Thierry Lecroq, and Kunsoo Park. “Fast Multiple Pattern Cartesian Tree Matching.” *International Conference and Workshop on Algorithms and Computation*, 2020.

Myoungji Han, Hyunjoon Kim, **Geonmo Gu**, Kunsoo Park, and Wook-Shin Han. “Efficient Subgraph Matching: Harmonizing Dynamic Programming, Adaptive Matching Order, and Failing Set Together.” *ACM SIGMOD International Conference on Management of Data*, 2019.

Myoungji Han, Munseong Kang, Sukhyeun Cho, **Geonmo Gu**, Jeong Seop Sim, and Kunsoo Park. “Fast Multiple Order-Preserving Matching Algorithms.” *International Workshop on Combinatorial Algorithms*, 2015.

Seongi Hong, **Geonmo Gu**, Hyunjoon Kim, Kunsoo Park. “Performance Comparison of Adaptive Matching Orders for the Subgraph Isomorphism Problem.” *KIISE Transactions on Computing Practices*, 26.1:38-43. 2020.

Seongi Hong, **Geonmo Gu**, Hyunjoon Kim, Kunsoo Park. “Performance Comparison of Candidate-Size Ordering and Path-Size Ordering for Subgraph Isomorphism Problem.” *Korea Computer Congress*, 2019

PROJECTS

Framework of Practical Algorithms for NP-hard Graph Problems **Seoul National University**
Funded by the Korea government (Ministry of Science and ICT) *Apr. 2018 – Aug. 2021*

- Algorithm development for fast subgraph isomorphism, graph isomorphism, and graph isomorphism query processing.
- Open source contribution for practical graph algorithms (<https://github.com/SNUCSE-CTA>).

Algorithm Development for Scanner/Stage Path Generation **Seoul National University**
Supported by JASTECH *Jul. 2014 – Jun. 2017*

- Sophisticated algorithm that can synchronize Scanner and Stage.
- Development of path simplification method based on chain stabbing (computational geometry).
- Efficient path generation methods by solving the traveling salesman problem (NP-complete).

PROFESSIONAL ACTIVITIES

Reviewer of Information Processing Letters **ELSEVIER**
Dec. 2020 – Sep. 2023

Seminar about Distributed Training Techniques for Large AI Models **LG Electronics**
Jul. 2023

Invited talk at STARLAB Meeting **Korea Computer Congress**
Jun. 2023

Invited talk at 2023 TOPCIT Workshop **IITP**
Mar. 2023

HONORS

The 14th Open SW Developer Contest

Awarded a Gold Prize

- Project Title: GI (Graph Isomorphism Algorithm)
- Source Code URL: <https://github.com/SNUCSE-CTA/GI>

Ministry of Science and ICT

Nov. 2020

The 2nd Test of Practical Competency in IT (TOPCIT)

Awarded a Silver Prize

Ministry of Science and ICT

Sep. 2013

The 1st Test of Practical Competency in IT (TOPCIT)

Awarded a Grand Prize

Ministry of Knowledge Economy

Oct. 2012

SKILLS

Competitive Programming.

- BAEKJOON: <https://www.acmicpc.net/user/gmgu>

Programming Languages. C/C++, Python, CUDA C++, C#, Java, Shell Script, \LaTeX

- C++: <https://github.com/gmgu/GI>
- CUDA C++: <https://github.com/gmgu/study-cuda>

Libraries. PyTorch, TensorFlow, Triton (OpenAI), Seaborn, Pandas, PySpark, HuggingFace Transformers, DeepSpeed, NVIDIA Triton, NVIDIA Faster Transformer, FastAPI, gtest

- Triton: <https://github.com/gmgu/study-trident>

Others. AWS (SageMaker, EC2, Lustre, S3)