Jinsol Park

+82-10-7143-6155 | jinsolp@cs.cmu.edu | LinkedIn

Education

Carnegie Mellon University (CMU) — Pittsburgh, PA
MS in Computer Science — *Aug 2023 – Dec 2024 (Expected)*

Seoul National University (SNU) — Seoul, Korea
BS in Computer Science and Engineering, Graduated Cum Laude — *Mar 2018 – Feb 2023*

Technical Skills

Languages: Python, C/C++, Rust, Java
Frameworks / Libraries: PyTorch, TensorFlow, Transformers, Onnx, DVC, Hydra, NumPy, Matplotlib, Weights & Biases, Flink, Nsight Compute
Developer Tools: Git, Docker, GCP, AWS, LaTex

Experience

Undergraduate Research Assistant @ SNU — Seoul, Korea
Architecture and Code Optimization Lab — *Jun 2022 – Dec 2022*

- Optimized inter-device communication for Mixture-of-Experts training with PyTorch, reducing All-to-All latency.
- Established a multi-node environment using Docker on GCP, enabling efficient experiments.
- Constructed system for federated learning of Mixture-of-Experts considering model privacy via weight permutation.

Software Platform Lab — *Jan 2021 - Jun 2022*

- Analyzed sparse attention efficiency using Transformers, emphasizing the need of integration into our system.
- Implemented custom TensorFlow Ops to exploit Triton kernels in Python environment.
- Designed a special prompt tuning method using PyTorch, increasing performance by 11.9%.
- Participated in a project on specialized store backend for streaming workloads, improving throughput by 4.12x.
- Implemented disk compaction in C for a stream-processing targeted system, minimizing memory overhead.
- Identified memory leaks within our system's Java Native Interface, enabling experiments without crashing.

Thunder Research Group — *Aug 2020 - Dec 2020*

- Analyzed runtime breakdowns of different DL frameworks using Nsight Compute.

Software Research Intern @ NCSoft — Pangyo, Korea
Natural Language Understanding Team, NLP Lab — *Jul 2021 – Sep 2021*

- Established a novel DL model training scheme for zero anaphora resolution, contributing to company's chatbot.

Projects

FCR Detector | *Python, PyTorch, Docker* — *Sep 2022 – Dec 2022*

- Led team of 4 to collaborate with Sherpa Space Corp. to develop a DL model predicting FCR disease in crops.
- Performed data augmentation on time series data in Python for training a DL model.

Deepest Model | *Pytorch-Lightning, Weights & Biases, Hydra, DVC, Onnx, Docker, AWS* — *Mar 2022 – Jul 2022*

- Taught the team frameworks and packages essential to the MLOps flow.

Dataset Corruption Detection | *Python, PyTorch, CrypTen* — *Mar 2022 – Jul 2022*

- Experimented impact of corrupted dataset in multi-party DL, emphasizing need for detection methods.
- Applied Zero Knowledge Proof to detect dataset corruption in multi-party DL without harming privacy.

LLVM Compiler Optimization | *C++, LLVM Compiler, Git, Docker* — *Mar 2021 – Jul 2021*

- Designed customized optimization passes for LLVM compilers in C++, reducing 40% of cost on average.

Publications

- Ha, H.; Jung, S.; **Park, J.**; Seo, M.; Hwang, S.; Chun, B. G.,"Two Examples are Better than One: Context Regularization for Gradient-based Prompt Tuning", *ACL Findings* (2023)
- Lee, G.; Maeng, J.; **Park, J.**; Seo, J.; Cho, H.; Yang, Y.; Um, T.; Lee, J.; Lee, J. W.; Chun, B. G.,"FlowKV: A Semantic-Aware Store for Large-Scale State Management of Stream Processing Engines", *Proceedings of the Eighteenth EuroSys Conference* (2023)
- **Park, J.**; Koo, Y.; Chun, B. G., "Analyzing Computational Efficiency of Diverse Sparse Attention Mechanisms", *Korea Software Congress* (2021)
- **Park, J.**; Choi, M.; Matteson, A.; Lee, C., "Optimizing ELECTRA-based model for Zero Anaphora Resolution", *Human and Language Technology* (2021)
- (*Under Review*) Han, W.; **Park, J.**; Lee, K., Designed a unified system that enables large language models to handle any type of information-seeking question by exploiting presuppositions.