Dongyoon Ryu

Combined M.S./Ph.D. Student, OMNIA Lab
Graduate School of Artificial Intelligence, POSTECH
dyryu4w@postech.ac.kr | https://nier4ryu.github.io/Dongyoon-Ryu

RESEARCH INTERESTS

- AI Systems
- Edge Device Systems
- Video Processing

EDUCATION

Pohang University of Science and Technology (POSTECH) Aug. 2024 – Present
Combined M.S./Ph.D. in Artificial Intelligence
OMNIA Lab of POSTECH · Advisor: Myeongjae Jeon

Ulsan National Institute of Science and Technology (UNIST) Feb. 2023 – Aug. 2024
Combined M.S./Ph.D. program in Computer Science Engineering
OMNIA Lab of UNIST · Advisor: Myeongjae Jeon

Ulsan National Institute of Science and Technology (UNIST) Feb. 2017 – Feb. 2023
Bachelor of Science in Computer Science and Engineering

PUBLICATIONS

[1] **DIALER: A Case for Improving Rare-Class Accuracy in Retraining-Free Edge Video Analytics** 2026
Preprint
Dongyoon Ryu, Sungho Jeon, Xinyue Ma, Di Wang, Jonghyun Choi, Minjia Zhang, Myeongjae Jeon
arXiv preprint arXiv:2610.06358

RESEARCH PROJECTS

Semantic-Aware KV Cache Reuse for LLM Serving Feb. 2026 – Present
Collaboration with Samsung Research
This project investigates semantic-aware KV-cache reuse to reduce redundant computation and memory overheads in multi-turn LLM serving. I focus on memory optimization, analyzing KV-cache placement and movement while investigating memory layouts for selective KV reuse and recomputation.

Efficient and Scalable Framework for AI Heterogeneous Cluster Systems Apr. 2025 – Present
Collaboration with Moreh
This project develops a scalable framework for distributed deep learning across heterogeneous AI accelerators, including GPUs, NPUs, and PIMs. I am developing an energy profiler to characterize energy consumption across heterogeneous AI accelerators for energy-aware workload partitioning and optimization.

Development of ML Compiler Framework for On-Device AI Jul. 2024 – Present
Collaboration with ETRI
This project develops an ML compiler and runtime framework for heterogeneous on-device AI systems under resource constraints. I am developing a continual learning (CL) system to support efficient on-device model updates while considering latency, memory, and power constraints.

EXPERIENCES

Teaching Assistant	Introduction to Computer Software Systems (Fall 2026)	*POSTECH, Korea*
	AI Systems (Fall 2024)	
	Computer Architecture (Fall 2023)	*UNIST, Korea*
	Human Computer Interaction (Spring 2023)	

SKILLS

Programming Language	C/C++, Python
Tools/Skillset	PyTorch, Ray
Communication	Korean (native), English (fluent)