DONGHYEOK SHIN

Ph.D. Candidate @ KAIST, ISE
tlsehdgur0@kaist.ac.kr ⋄ sdh73987114@gmail.com
[Google Scholar] [Semantic Scholar] [Github]

RESEARCH INTERESTS

My research interest primarily focuses on developing efficient knowledge transfer algorithms in real-world scenarios. I currently focus on the following topics:

- **Data-level efficiency**: Dataset Distillation, Coreset Selection
- **Network-level efficiency**: Knowledge Distillation
- **Learning with Non-stationary data**: Continual Learning, Lifelong Learning

EDUCATION

KAIST, Daejeon, Korea *Mar. 2022 - Present*

- Ph.D., Department of Industrial & Systems Engineering
- Advisor: Prof. Il-Chul Moon

KAIST, Daejeon, Korea *Mar. 2020 - Feb. 2022*

- M.S., Department of Industrial & Systems Engineering
- Thesis: Dataset Distillation via Loss Approximation for Continual Learning
- Advisor: Prof. Il-Chul Moon

KAIST, Daejeon, Korea *Mar. 2015 - Feb. 2020*

- B.S., Department of Mathematical Sciences
- B.S., Department of Industrial & Systems Engineering (Double Major)

PREPRINTS

[P5] **AMiD: Knowledge Distillation for LLMs with α-mixture Assistant Distribution**
Donghyeok Shin, Yeongmin Kim, Suhyeon Jo, Byeonghu Na, and Il-Chul Moon
Under review, 2026.

[P4] **Distillation of Large Language Models via Concrete Score Matching**
Yeongmin Kim, Donghyeok Shin, Mina Kang, Byeonghu Na, and Il-Chul Moon
Under review, 2026.

[P3] **AC-Sampler: Accelerate and Correct Diffusion Sampling with Metropolis-Hastings Algorithm**
Minsang Park, Gyuwon Sim, Hyungho Na, Jiseok Kwak, Sumin Lee, Richard Lee Kim, Donghyeok Shin, Byeonghu Na, Yeongmin Kim, and Il-chul Moon
Under review, 2026.

[P2] **Towards Adversarially Robust VLMs with an Information-Theoretic Approach**
Jiseok Kwak, Donghyeok Shin, Richard Lee Kim, Minsang Park, Zeynep Altiner, and Il-chul Moon
Under review, 2026.

[P1] **Towards Pareto-Optimality for Test-Time Adaptation**
JoonHo Jang, DongHyeok Shin, Byeonghu Na, HeeSun Bae, and Il-chul Moon
Preprint, 2024.

PUBLICATIONS

(*: equal contribution)

[C7] **Diffusion Adaptive Text Embedding for Text-to-Image Diffusion Models**
Byeonghu Na, Minsang Park, Gyuwon Sim, Donghyeok Shin, HeeSun Bae, Mina Kang, SeJung Kwon, Wanmo Kang, and Il-Chul Moon
Neural Information Processing Systems (**NeurIPS**), 2025

[C6] **Distilling Dataset into Neural Field**
Donghyeok Shin, HeeSun Bae, Gyuwon Sim, Wanmo Kang, and Il-Chul Moon
International Conference on Learning Representations (**ICLR**), 2025

[C5] **Diffusion Rejection Sampling**
Byeonghu Na, Yeongmin Kim, Minsang Park, Donghyeok Shin, Wanmo Kang, and Il-Chul Moon
International Conference on Machine Learning (**ICML**), 2024

[C4] **Frequency Domain-based Dataset Distillation**
Donghyeok Shin*, Seungjae Shin*, Il-Chul Moon
Neural Information Processing Systems (**NeurIPS**), 2023

[C3] **Loss Curvature Matching for Dataset Selection and Condensation**
Seungjae Shin*, Heesun Bae*, DongHyeok Shin, Weonyoung Joo, Il-Chul Moon
Artificial Intelligence and Statistics (**AISTATS**), 2023

[C2] **Hierarchical Multi-Label Classification with Partial Labels and Unknown Hierarchy**
Suhyeon Jo, DongHyeok Shin, Byeonghu Na, JoonHo Jang, and Il-Chul Moon
ACM International Conference on Information and Knowledge Management (**CIKM**), 2023

[C1] **Unknown-Aware Domain Adversarial Learning for Open-Set Domain Adaptation**
JoonHo Jang, Byeonghu Na, DongHyeok Shin, Mingi Ji, Kyungwoo Song, and Il-Chul Moon
Neural Information Processing Systems (**NeurIPS**), 2022

AWARDS & HONORS

- Qualcomm Innovation Fellowship 2025 South Korea Winner
- Top Reviewer of NeurIPS 2025
- Grand Prize (1st Prize), Undergraduate Paper Competition in the Korea Software Congress 2017 (KSC 2017), 2017

SERVICES

Reviewer

- International Conference on Learning Representations (ICLR): 2025, 2026
- Neural Information Processing Systems (NeurIPS): 2025

PROJECTS

Synthetic Data Generation *Mar. 2025 - Present*

- Funded by Ministry of Science and ICT of Korea
- Developed a method for synthetic data generation and advanced AI prediction models in exceptional situations.

Semi-supervised Continual Learning *Mar. 2025 - Present*

- Funded by SK hynix
- Developed semi-supervised continual learning algorithms for semiconductor images.

Drone Task Planning via LLM *Dec. 2024 - Mar. 2025*

- Funded by Information Technology Research Center of Korea
- Developed a pipeline to perform drone task planning with Large Language Model (LLM).

Multi-Modal Dataset Distillation *Nov. 2023 - Dec. 2024*

- Funded by Samsung Advanced Institute of Technology (SAIT)
- Developed a parameterization method for multi-modal dataset distillation.

Removed Thickness Prediction *Mar. 2023 - Aug. 2023*

- Funded by SK hynix
- Developed a prediction model for removed thickness in CMP process using internal and external factors.

10K Report Analysis *Jun. 2021 - Feb. 2023*

- Funded by National Research Foundation of Korea
- Developed an entity-relation extraction model for a comprehensive report using GNN.

Critical Dimension Prediction *Mar. 2021 - May. 2021*

- Funded by Samsung Display
- Developed a data-driven virtual metrology model using VAE and Gaussian process regression.

Card Delinquent Profiling *Mar. 2020 - Feb. 2021*

- Funded by Hyundai Card
- Developed a prediction model for card delinquents using a neural network with LSTM and VAE.

TEACHING EXPERIENCES

Teaching Assistant

- IE481 Data Structures and Analysis *Fall. 2021*

- IE260 Data Structures and Analysis *Spring. 2022*
- KOOC Introduction to Artificial Intelligence and Machine Learning I *2021*
- KOOC Data Structures and Analysis: Non-Linear Structure, Optimization, and Algorithms *2022, 2023*
- KOOC Data Structures and Analysis: Linear Structure and Dynamic Programming *2023*
- 2nd LG AI Advanced Advanced Course *Oct. 2020 - Dec. 2020*