# Sang-Ho An

**M.S. student**

Department of Computer Science, Kookmin University
Seoul, 02707, Republic of Korea

: dksgh00@naver.com
: github.com/Ahnho
: Google Scholar
: 0009-0007-3904-5467


## Education

**M.S. in Computer Science**, Kookmin University, Seoul, Republic of Korea — Mar. 2024 – Present
Supervisor: Prof. [Jang Ho Kim]

**B.S. in Computer Science**, *Kookmin University, Seoul, Republic of Korea* — Mar. 2018 – Feb. 2024

## Military Service

*Republic of Korea Army (ROKA)* — Mar. 2020 - Sep. 2021

## Research Areas & Expertise

- **Model Compression**
  - Network Pruning: Structured/unstructured pruning for model efficiency and deployment optimization
  - Hardware-aware Pruning: Structured pruning techniques for accelerated inference
  - Knowledge Distillation: Teacher-student frameworks and feature transfer techniques
  - Quantization: Reducing the bit precision of weights and activations (e.g., FP32 to INT8) to decrease model size and accelerate inference
- **Computer Vison, Natural Language Processing, Parameter-Efficient Fine-Tuning, Machine Unlearning**

## Programming Skills & Tools

**Primary Languages**
- Python, C++, Nim

**Deep Learning & ML**
- PyTorch, TensorFlow
- Weights & Biases (wandb)
- ONNX, TorchScript

**Scientific Computing**
- NumPy, SciPy, Pandas
- Matplotlib, Plotly

**Optimization & Deployment**
- TensorRT, ONNX Runtime
- OpenVINO, cuBLAS

**Edge AI Experience**
- NVIDIA Jetson (Orin, Xavier, Spark)
- Intel NUC
- Edge Device Deployment

**Development Tools**
- Git, Docker
- Linux/Ubuntu

## Publications

2025 *"Quantization-Aware Training With Dynamic and Static Pruning"*, **Sangho An**, Jongyun Shin, Jangho Kim [IEEE Access 2025, Accepted]

2025 *"Sparse Structure Exploration and Re-optimization for Vision Transformer"*, **Sangho An**, Jinwoo Kim, Jangho Kim et al [UAI 2025, Accepted]

2025 *"Exploring Diverse Sparse Network Structures via Dynamic pruning with Weight Alignment"*, Jinwoo Kim, Jongyun Shin, **Sangho An**, Jangho Kim, [CIKM 2025, Accepted, (Oral)]

## Publications (Under Review)

2025 *"Hardware-Aware Grouped Pruning for Efficient Vision Transformers"*, **Sangho An**, Hyunjoon Cho, Jangho Kim
2025 *"Unlearning-Aware Optimization"*, Hyunjoon Cho, **Sangho An**, Jangho Kim
2025 *"ESFP: Effective Soft Prompt Fine-Tuning using Parameter-efficient Mixture-of-Experts"*, Jongyun Shin, **Sangho An**, Jangho Kim
2025 *"Temporal-Aware Quantization via Sequence-Wise Quantizer Bounds Ensembling for Real-World Video Super-Resolutions"*, Jinwoo Chung, **Sangho An**, Jangho Kim
2025 *"SharedKD: In-Place Knowledge Distillation for Efficient 3D Object Detection"*, Hyunjoon Cho, **Sangho An**, Jangho Kim

## Recent Talks

2025 *"The Conference on Uncertainty in Artificial Intelligence (UAI)"* at UAI 2025, Rio de Janeiro, Brazil

2025 *"MIPAL 2025 Summer Collaborative Research Workshop"* at Konjiam Resort, Korea

2025 *"The Conference on Information and Knowledge Management (CIKM)"* at CIKM 2025, COEX, Seoul, Korea

## Research Project

2024.4 *"Model Lightweighting Research was supported by **Hyundai Motor Company**"*
~2025.3

2024.12 *"Autonomous Driving Technology Research was supported by **Hyundai Motor Company**"*
~2025.11

2024.3 *"Algorithm Development for Shade Matching was supported by **Small and Medium Business Administration**"*
~2025.3

2024.11 *"Model Lightweighting Research was supported by **LG CTO**"*
~2025.11

2025.07 *"Anomaly Detection Research was supported by **POSCO**"*
~ Present

2025.09 *"Large Model Lightweighting Research was supported by **Hyundai Motor Company**"*
~ Present

## Patent

2024 "동적 및 정적 프루닝 기법을 활용한 양자화 인식 학습 장치 및 학습 방법" , 김장호 , **안상호**, [10-2024-0068090]

## Study Groups

**SYTEARK Machine Learning Study Group** [SYTEARK; MLStudy, ML2022, ML2021] 2019 – 2024

- Weekly rotational presentations on machine learning journals, topics, and personal research

## Teaching Experience

2025 Fall TA *Computer Programming II, 1131001, Kookmin University, Seoul, Republic of Korea*