Dongsu Song
Email: raister2873@gmail.com
Homepage: https://sharpit1.github.io/my-website/

EDUCATION

Korea Aerospace University, Gyeonggi-do, Republic of Korea
Master of Science in Artificial Intelligence |***GPA: 4.50/4.50*** *Mar. 2023 – Feb. 2025*

- *Quantum Artificial Intelligence (QAI) Laboratory under the supervision of Prof. J.H. Jung*
- Coursework*:* Advanced Neural Networks, Advanced Computer Vision, Autonomous Driving Cars

Korea Aerospace University, Gyeonggi-do, Republic of Korea
Bachelor of Science in Mechanical Engineering |***GPA: 4.24/4.50*** *Mar. 2017 – Feb. 2023*

- **Magna Cum Laude**
- Coursework*:* Machine Learning, C Programming & Training, Introduction to Quantum Computing

PUBLICATIONS

- **Dongsu Song**, Daehwa Ko, and Jay Hoon Jung. "Amnesia as a catalyst for enhancing black box pixel attacks in image classification and object detection." ***Advances in Neural Information Processing Systems (NeurIPS)***, 2024
- **Dongsu Song,** and Jay Hoon Jung**.** “A Quantitative Comparison of LIME and SHAP using Stamp-Based Distance Method on Image Data.” (in Korean), *Journal of Korean Institute of Information Scientists and Engineers (KIISE)*, 2023

RESEARCH INTERESTS

- **Trustworthy Machine Learning:** adversarial robustness in vision, language, and multimodal models
- **Reinforcement Learning for Robustness:** RL-based adversarial attacks in black-box settings
- **Generative AI for Robustness:** generating OOD datasets and semantic adversarial examples

PROFESSIONAL ACTIVITIES

Reviewer, Conference on Neural Information Processing Systems **(NeurIPS)** *2025*

- **Reviewed** submissions on adversarial robustness and trustworthy ML

Invited poster presentation, Korean Conference on Computer Vision **(KCCV),** Busan**,** Republic of Korea *Aug. 2025*

- **Selected** to present “Amnesia as a Catalyst for Enhancing Black Box Pixel Attacks in Image Classification and Object Detection,” sparking discussions on realistic sparse perturbations and robust vision systems

Undergraduate AI Mentor, in Quantum AI Lab, Korea Aerospace University *Fall 2023*

- **Organized and** led a seminar on machine learning fundamentals for junior lab members

Research Assistant, Industry-Academia Collaboration Project, Korea Aerospace University *Fall 2023 - Fall 2024*

- **Mentored** and **guided** undergraduate students throughout a capstone project

RESEARCH EXPERIENCE

Airbus Institute for Engineering Research (AIER), an Airbus-funded collaboration with USC
Graduate Researcher at Korea Aerospace University
Project: Robust, Verified, and Trusted AI in Aviation *Mar. 2023 - Feb. 2025*

- **Led** the development of an RL-based sparse pixel attack framework for query-efficient black-box adversarial attacks, addressing the severe query inefficiency of prior methods
- **Designed** a CNN-based reinforcement-learning agent with an “Amnesia” mechanism that periodically resets parts of the agent’s state to escape over-exploitation and extended the framework from image classification to object detection
- **Achieved** on average ~20% fewer queries and a 15% higher attack success rate compared to existing black-box baselines on ImageNet models
- **First author** of the NeurIPS 2024 poster “Amnesia as a Catalyst for Enhancing Black Box Pixel Attacks in Image Classification and Object Detection.”
- Ongoing project extending this framework to semantic segmentation with a new loss for sparse pixel perturbations on per-pixel prediction

Industry Innovation Talent Program, funded by KIAT
Graduate Researcher
Project: Analysis of the Robustness of Quantized Tiny Vision Models *Mar. 2023 - Dec. 2023*

- **Proposed** and **led** a 3-person research team to investigate the robustness of tiny vision models
- **Converted** both standard and adversarially trained CNN models into "tiny" models using quantization techniques
- **Analyzed** the vulnerability of both model types to adversarial attacks, demonstrating that quantized adversarial-trained models became more susceptible to attacks
- **Won an Honorable Mention** at the 21st World Embedded Software Contest for this project

Basic Research Program, funded by NRF
Student Researcher
Program: Quantitative Comparison Methods for LIME and SHAP *Mar. 2022 - Feb. 2023*

- **Adapted** and **applied** a quantitative comparison methodology originally used in Gradient-based XAI to evaluate the fidelity of LIME and SHAP
- **Demonstrated** that LIME provides higher explanation fidelity than SHAP for image data
- **First author** of a paper **accepted in the *Journal of KIISE***: "*A Quantitative Comparison of LIME and SHAP using Stamp-Based Distance Method on Image Data*"

ACADEMIC PROJECTS

Autonomous Outdoor Drone Navigation (Undergraduate Capstone) *Mar. 2021 - Dec. 2021*

- **Led** the development of a custom drone from scratch, implementing a noise-robust algorithm that enabled successful obstacle avoidance and a 25-minute autonomous flight

Land-Cover Classification with XAI (Undergraduate Project) *Mar. 2022 - Dec. 2022*

- **Applied** LIME/SHAP to diagnose misclassifications in satellite imagery; this analysis identified limitations in existing metrics and evolved into my first journal publication on quantitative XAI evaluation

Generative Model Implementation (Graduate Course Project) *Spring 2023*

- **Implemented** a DDPM from scratch and compared it against GAN baselines in terms of training stability and mode collapse

MILITARY SERVICE

Republic of Korea Army, Hwajeon, Gyeonggido *July 2018 – Feb. 2020*
Squad Leader (Headquarters Squad) *Jan. 2019 - Jan. 2020*

- **Awarded** the Division Commander’s Award (highest honor for exemplary service) and Battalion Commander’s Commendation for outstanding teamwork and leadership
- **Led and managed** a 6-person squad, overseeing headquarters administration, communications, and operational support

AWARDS AND SCHOLARSHIPS

- **Honorable Mention, 21st World Embedded Software Contest** *Dec. 2023*
- **Research Excellence Scholarship** (Journal of KIISE) *Fall 2023*
- **Advisor-Selected Graduate Scholarship** (half tuition; Sole Recipient in Lab) *Fall 2023 – Fall 2024*
- **Academic Excellence Scholarship** (half tuition; 2nd-highest GPA in department) *Spring 2021, Fall 2017*
- **Academic Merit Scholarship** (one-sixth tuition; top 10% in department) *Spring 2022, Fall 2021, Spring 2019*

SKILLS

- **Programming:** Python, C
- **AI Frameworks & Libraries**: PyTorch, TensorFlow, Hugging Face, timm, MMSegmentation
- **Tools:** Git, Docker
- **Research Expertise:** Adversarial attacks on vision models, explainable AI, reinforcement learning
- **Models & Methods:** CNN and Transformer-based Vision Models, LIME, SHAP, RISE
- **Datasets & Benchmarks:** ImageNet-2012, PASCAL VOC, ADE20K, Cityscapes, Argoverse Sample