

RESEARCH STACK	Computer Vision: Classification, Semantic Segmentation, Object Detection Data Scarcity, Privacy and Efficiency: Semi/Self Supervised Learning, Privacy Preserving, Dataset Distillation On Device: Federated Learning, Quantization, Network Pruning Generative AI: Large Language Model, Simulation-to-Real Transfer	
EDUCATION	<b>Korea University</b> M.S. in Computer Science and Engineering ( <i>Advisor: Prof. Seung Jun Baek</i> ) <b>University of Seoul</b> B.S. in Statistics and Data Science	Mar 2019 - Feb 2022 Mar 2012 - Feb 2019
EXPERIENCE	<i>Machine Learning Researcher</i> — <b>Deeping Source</b> - Research on Large Language Model - Research on Simulation-to-Real Transfer - Research on Model Compression <i>M.S. Candidate</i> — <b>System INtelligence Group (SING) Lab, Korea University</b> - Research on Federated Semi-Supervised Segmentation - Research on AI system for Rehabilitation Medicine - Research on Medical Image Segmentation	Jun 2022 - present Mar 2019 - Feb 2022
PUBLICATION	<ul style="list-style-type: none"> <li>• <b>Minhyeong Yu</b>, Federica Spinola, Myeongjun Kim, Philipp Benz, Tae-hoon Kim, “Rethinking of Straight-Through Estimator: Quantization-Bias Aware Training”, (under revision), 2024.</li> <li>• Federica Spinola, Philipp Benz, <b>Minhyeong Yu</b>, Tae-hoon Kim, “Knowledge Assembly: Semi-Supervised Multi-Task Learning from Multiple Datasets with Disjoint Labels”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023.</li> <li>• <b>Minhyeong Yu</b>, Sunwoo Kim, Seungjun Baek. “Federated Semi-Supervised Segmentation with Randomized Weight Perturbation”, International Symposium on Biomedical Imaging (ISBI), 2023.</li> <li>• Beom Suk Kim*, <b>Minhyeong Yu*</b>, Sunwoo Kim, Joon Shik Yoon, Seungjun Baek, “Scale-Attentional U-Net for the Segmentation of the Median Nerve in Ultrasound Images”, Ultrasonography, 2022.</li> <li>• Minki Kim* <b>Minhyeong Yu*</b>, “Selection and Proposal of Vertical Building Forest Sites in preparation for the implementation of the Seoul Park Cancellation”, Review of Korean Society for Internet Information, 2018.</li> </ul>	
PROJECT	<ul style="list-style-type: none"> <li>• <b>Generative Agents for Sales Solution Generation</b>, Deeping Source Using a Large Language Model (LLM), this system creates a simulator where a generative agent makes shopping decisions in an offline environment. The simulator models various shopping scenarios, mimicking real-world customer behaviors in physical store settings. Using this simulator, LLM tests and generates solution reports aimed at increasing sales, providing store managers with insights to enhance customer experiences and develop more effective sales strategies.</li> <li>• <b>Simulation-to-Real Transfer with Diffusion Model</b>, Deeping Source Using NVIDIA Omniverse, a realistic retail store simulation is constructed, and data for multi-camera multi-tracking is gathered by setting various camera angles. This data is processed through a Diffusion model to resemble actual store CCTV footage, and the enhanced data is then utilized to train object tracking algorithms, contributing to the security and customer experience management in retail stores.</li> <li>• <b>Cumulative Model Compression</b>, Deeping Source Many model compression methods are often not compatible with each other for cumulative use. For instance, applying quantization may preclude the possibility of network pruning, and compressing weights to lower precision can impose additional burdens on activation functions. This incompatibility of techniques for cumulative application poses practical challenges in real-world development. We have developed a model compression method that is both accumulable and deployable, addressing these practical challenges.</li> <li>• <b>Implementing Quantization on Multiple Hardware Systems</b>, Deeping Source We have performed quantization in various frameworks, including Pytorch, Onnx, TensorRT, Openvino, AIMET, and Furiosa SDK, to enable the use of models for different tasks (classification, object detection) across a range of hardware platforms, such as NVIDIA, Intel, Qualcomm, and Furiosa.</li> <li>• <b>Quantization Aware Training for Object Detection</b>, Deeping Source Quantization is a promising technique for faster speed of inference. However, it often struggles to maintain its performance. To address this issue, we conducted a study on quantization aware training. Our findings suggest</li> </ul>	Nov 2023 - present Nov 2023 - present Feb 2023 - Oct 2023 Feb 2023 - Oct 2023 Jun 2022 - Jan 2023

that the quantization bias between fake quantized activation and full precision one can be reduced when the interaction in matrix multiplication is taken into account. We have documented our observations in a paper for further reference.

- **Federated Semi-Supervised Segmentaton**, Korea University Feb 2022 - Mar 2022  
Medical Image Segmentation is challenging due to limited annotated data and privacy concerns. Federated Learning and Semi-Supervised Learning help train models in a private way. We introduce FedWeP, a Federated Semi-Supervised Segmentation method using Randomized Weight Perturbation, where the server adds Gaussian noise to model weights for client training.
- **AI system for Rehabilitation Medicine**, Korea University Sep 2020 - Dec 2021  
We have been developing an AI-based system for rehabilitation medicine, supported by the Ministry of Science and ICT (MSIT) of Korea and supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP). During the first year of the ICT Creative Consilience program, we developed a system to assess hemiplegic patients and recommend suitable exercises. In the second year, we created an automated system for the detection of videofluoroscopic swallowing studies in stroke patients.
- **Medical Image Segmentaton**, Korea University May 2020 - Nov 2021  
We collaborated with Korea Guro Hospital to study nerve segmentation on ultrasound imaging modality, for which we were awarded the Excellence Prize at the Korean Academy of Neuromusculoskeletal Sonography. Subsequent to this, we applied for a patent for this research and further studied it to propose a novel convolution, namely Scale Attentional Convolution, specialized in ultrasound nerve image segmentation.

#### PATENT & HONOR

- “Method and apparatus for automatically recognizing peripheral nerves and measuring nerve indicators in ultrasound images based on deep learning algorithms”, 10-2020-0067199, Rep. of Korea Jun 2020
- Excellence Prize, Korean Academy of Neuromusculoskeletal Sonography Nov 2020
- Excellence Prize, Seoul Digital Foundation Nov 2018