Junghye Kim

SEOUL, S.KOREA. 82+010-3178-5793 tweetiijh@naver.com

github.com/junghye01

EDUCATION

B.S. in Dongguk University, Seoul, Korea

- Department of Information & Communication Engineering
- Double major in DataScience software
- **Cumulative GPA: 4.26/4.5**
- **Honors : Summa Cum Laude**

RESEARCH INTEREST

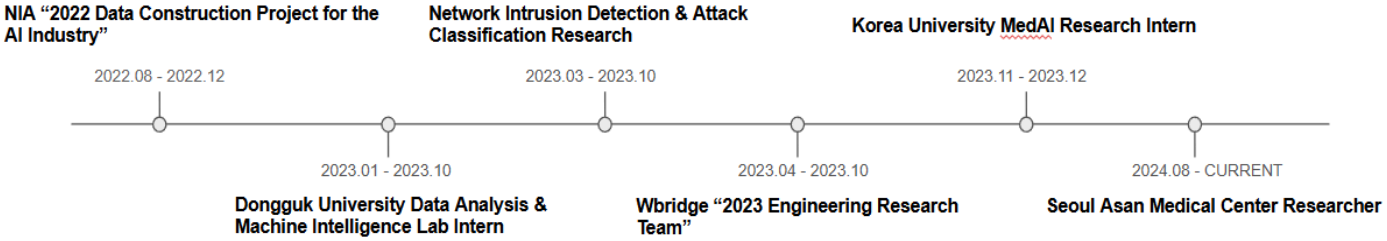
I am interested in **Computer Vision, Multimodal Learning**, and **Video Understanding.** My research goals are:

1. Developing models that are robust on unseen datasets and in challenging conditions
2. Designing a unified architecture capable of processing inputs from multiple modalities (e.g. Multimodal-LLM)

TECHNICAL SKILLS

Deep Learning, Machine Learning	Programming Languages	Version Control & Collaboration Tools	Version Control & Collaboration Tools
- Pytorch, Numpy, Pandas, etc.	- Python, C, Javascript	- Git & Github	- AWS, Docker, Jenkins

RESEARCH EXPERIENCE



Seoul Asan Medical Center (2024.08-2025.04)

Conduct cancer diagnosis research using CT imaging and develop 3D segmentation models

Researcher

- Develop a CT-based benign lymph node segmentation model via multi-resolution approach
- Develop a liver tumor segmentation model without manual annotations by generating synthetic tumors in normal liver scans.

Korea University MedAI Research Intern (2023.11-2023.12)

Focus on incomplete CT reconstruction and integrate clinical, imaging, and pathology data for diagnosis and prognosis prediction

Researcher

- Investigate references related to AI-based incomplete CT reconstruction
- Design a self-supervised CT image interpolation study using Image Masked Modeling

Wbridge "2023 Engineering Research Team" (2023.04-2023.10)

Participates as a team member in the 2023 Women Graduate Student Engineering Research Program

Researcher

- Develop an APMT technique that applies prompts dynamically based on question types
- **Publish "APMT: Adaptive Prompt-Matching Technique to Enhance Your Answers" in KSC 2023**

Dongguk University Data Analysis & Machine Intelligence LAB (2023.01-2023.10)
Conducts research on deep learning, computer vision, and multimodal technologies in data analysis

Research Intern
- Develops a machine learning framework for hierarchical network attack classification using an integrated benchmark dataset
- Develop a multi-label image classification model for long-tailed distributions
- Manage servers by using Docker containers

NIA “2022 Data Construction Project for the AI Industry” (2022.08-2022.12)
Build the DGU-HAO Benchmark Dataset for 3D Human Action Analysis and Validate its effectiveness using the MMNet

Research Intern
- Develop a parallel processing pipeline for 67,505 3D skeleton data
- Evaluate human action recognition performance using the MMNet model
- **Publish “DGU-HAO: A Dataset With Daily Life Objects for Comprehensive 3D Human Action Analysis” in IEEE Access 2024**

PUBLICATIONS

DGU-HAO : A Dataset with Daily Life Objects for Comprehensive 3D Human Action Analysis (IEEE ACCESS, 2024) Jiho Park, **Junghye Kim**, Yujung Gil, Dongho Kim

APMT : Adaptive Prompt-Matching Technique to enhance your answers (Korea Software Congress, 2023)
Yunji Kim, **Junghye Kim***, Yeosol Heo*, Jihie Kim (* : Equal Contribution)

PATENTS

Apparatus and Method of Classifying and Detecting Real-Time Abnormal Behavior Based on Artificial Intelligence
(No. 10-0004987(2024)) (Submitted) Sangsoo Lim, **Junghye Kim**, Junho Choi, Seunghyeon Jung

CERTIFICATIONS

Advanced Data Analytics Semi-Professional(ADsP) Certification FEB 2022

TEACHING EXPERIENCE

‘Machine Learning and Data Science’ - Teaching Assistant SEP 2023- DEC 2023
- Led practical sessions applying machine learning algorithms to real-world datasets

Wbridge,Korean Women in Science and Technology Foundation - Mentor APRIL 2023- OCT 2023
- Mentored high school students on open-source project development and collaboration using Git and GitHub
- Guided students in research team participation and study engagement

‘Open source Software Practice’ - Teaching Assistant MAR 2023- JUNE 2023
- Led practical sessions on Docker-based deployment of Flask web applications
- Mentored 50 students through Q&A and interactive discussions

HONORS AND AWARDS

2024 Dongguk University Summa cum laude
2022 Dongguk University Academic Excellence Scholarship
2021 Top Student Scholarship by Course (‘General Physics and Experiment 2’)

LANGUAGE PROFICIENCY

Fluent in **English(TOEIC 905)** and Native in **Korean**

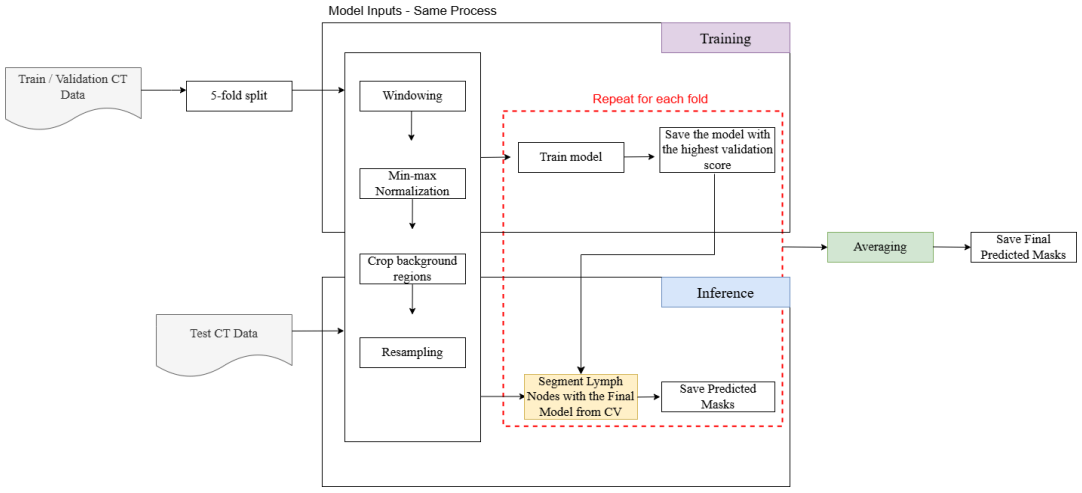
Research & Experiment Overview

ONGOING RESEARCH

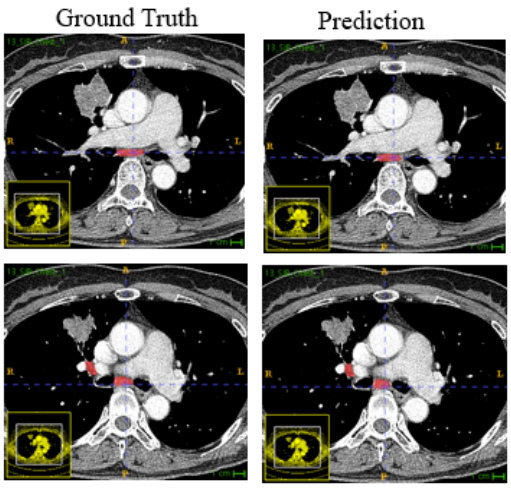
Title	A Dual-Resolution Mamba-based Approach to Mediastinal Lymph Nodes Segmentation in CT with weighted loss for class-imbalanced segmentation
Keywords	3D segmentation , Mediastinal Lymph nodes, CT, Mamba, nnU-Net, Multi-resolution, k-fold cross validation
Role	Experimental Design, Model Development and Testing

Background and Objective: Lymph node segmentation is vital for staging and treatment planning but remains challenging due to low contrast, size variability, and complex anatomical distribution. This study develops a high-performance model that enhances multi-scale feature extraction, refines boundary segmentation despite contrast limitations, and efficiently captures long-range dependencies in 3D CT scans while maintaining computational efficiency.

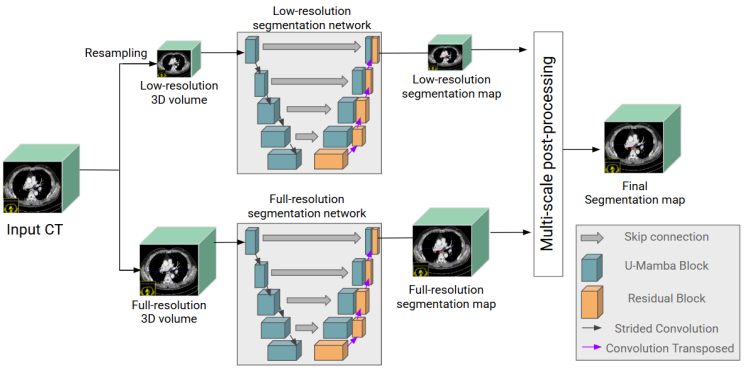
Methods: Lymph node segmentation remains challenging due to significant size and contrast variations, limiting the effectiveness of conventional architectures. CNNs, constrained by fixed receptive fields, struggle with long-range dependencies, while Transformers incur prohibitive computational costs for large 3D medical images. To address these issues, we adopt U-Mamba, a state-space model (SSM)-enhanced convolutional framework that efficiently processes long sequences while preserving local spatial structures. This architecture maintains scalability in volumetric data while mitigating the computational inefficiencies of self-attention mechanisms. A multi-resolution approach is employed, leveraging low-resolution 3D CT scans for coarse localization and high-resolution inputs for precise boundary refinement. Segmentation maps from both scales undergo connected-component analysis, filtering out false positives and enforcing structural consistency. To further improve specificity, foreground regions below 10,000 mm³ are discarded as they are prone to noise and over-segmentation artifacts. For training, we optimize a composite loss function—a weighted sum of Dice, cross-entropy, and Tversky losses—with a strong penalty on false negatives to improve recall, particularly for small or low-contrast lymph nodes.



Model Training and Testing Process



Qualitative results during the test phase

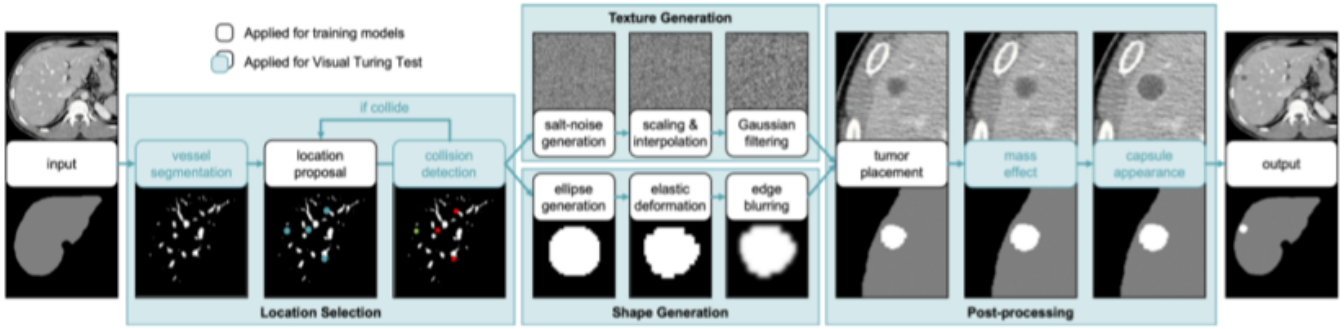


Overview of our multi-scale prediction fusion network Model Architecture

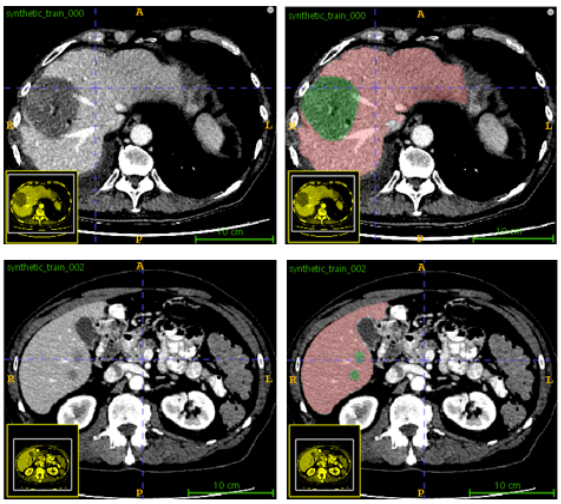
Title	Liver Tumor Segmentation via Tumor Synthesis
Keywords	3D segmentation , Liver and Tumor, Tumor Synthesis, K-Fold Cross Validation
Role	Experimental Design, Model Development and Testing

Background and Objective: Manually annotating medical images is not only time-consuming but also demands extensive expertise, often requiring radiology reports or biopsy confirmation. To reduce this annotation burden, we propose synthesizing realistic tumors—both in shape and texture—directly onto normal liver CT scans, then training a model to segment these synthetic tumors. Additionally, existing datasets often lack sufficient examples of small tumors, which are critical for detecting liver cancer in its early stages. By generating numerous small (or even tiny) synthetic tumors, we aim to improve the model's accuracy in identifying and segmenting early-stage liver tumors.

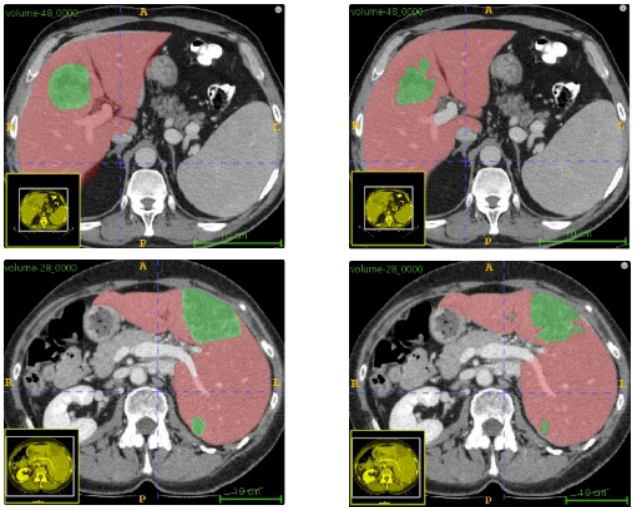
Methods: We trained on 116 normal liver CT scans—20 from CHAOS, 47 from BTCV, 38 from Pancreas-CT, and 11 from LiTS—and reserved 22 LiTS scans for testing. Before generating synthetic tumors, we extracted liver regions using a pretrained nnU-Net (trained on 101 LiTS scans) and introduced tumors through morphological processing. Tumors were categorized into five sizes (tiny, small, medium, large, mix), each selected with equal probability (20%).
The tumor synthesis process includes:(1) Selecting a location within 30–70% of the liver's depth, eroding edges, and picking a random (x, y, z) coordinate. (2) Generating texture using Gaussian noise matching normal liver intensity, smoothed with cubic interpolation and a Gaussian filter. (3) Shaping tumors as ellipsoids, deformed with elastic transformation and Gaussian blur for realism. (4) Integrating tumors into CT scans, simulating mass effect and capsule edges by adjusting HU intensities.
The final tumor mask and CT volume serve as labels and images for segmentation. The model employed is U-Mamba, evaluated under 5-Fold Cross Validation.



Synthesizing Tumors in Normal Liver Scans



Example of a Synthetic Tumor (Left : CT Volume, Right : CT with Mask Overlay)



Qualitative results during the test phase
(Left : Ground Truth, Right : Prediction)

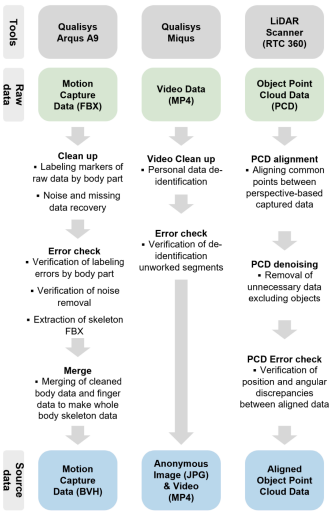
PUBLICATIONS

IEEE ACCESS(2024)	
Title	DGU-HAO : A Dataset with Daily Life Objects for Comprehensive 3D Human Action Analysis
Author	Jiho Park, Junghye Kim, Yujung Gil, Dongho Kim
Keywords	3D human action analysis, human action recognition, human activity understanding, motion capture, multi-modal dataset.
Role	Data Preprocessing, Model Training, Thesis Writing and Submission

Background and Objective:Existing 3D human action datasets often lack detailed object interactions, limiting their real-world applicability. To address this issue, DGU-HAO introduces a dataset that incorporates 63 action classes interacting with 60 common furniture and electronic devices. This dataset integrates RGB videos, 3D skeleton data, and point cloud (PCD) data into a single JSON file, significantly reducing preprocessing efforts and making it more accessible for researchers.

Methods:DGU-HAO was built using 67,505 motion capture samples collected from 126 subjects. Each action class consists of three distinct phases (approach → interaction → retreat), and the 3D skeleton data includes precise coordinates for 25 joints. During preprocessing, BVH files were converted into CSV and JSON formats, ensuring a unified global coordinate system for enhanced usability.
For dataset validation, the MMNet model was trained on 80% of the data and tested on the remaining 20%, using both skeletal joints and bone connections as input, thereby improving motion representation and recognition accuracy.

Results and Conclusion:The MMNet model achieved Top-1 classification accuracy of 91.51% using the Skeleton Joint method and 92.29% using the Skeleton Bone method, outperforming or matching existing 3D action datasets.
Additionally, the lowest classification accuracy across all 63 action classes remained above 98.67%, demonstrating the dataset's robustness and reliability.
Beyond action recognition, DGU-HAO is highly applicable for various research areas, including human-object interaction analysis, action generation, and real-time detection.



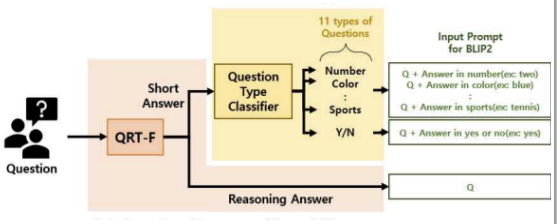
Comprehensive Setup for Capturing Multi-Modal Human Action Data in DGU-HAO Dataset

Korea Software Congress (2023)	
Title	APMT : Adaptive Prompt-Matching Technique to enhance your answers
Author	Yunji Kim, Junghye Kim*, Yeosol Heo*, Jihie Kim (* : Equal Contribution)
Keywords	Visual Question Answering, Vision-Language Model, Large Language Model, Prompt Engineering
Role	Experimental Design, Model Development, Prompt Application Experiments, Thesis Writing and Submission

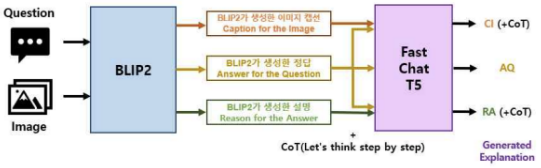
Background and Objective: Visual Question Answering (VQA) requires models to process both visual and textual information, but optimizing for both short factual answers and complex reasoning remains challenging. Existing approaches struggle to balance accuracy across different question types. To address this, we propose the Adaptive Prompt-Matching Technique (APMT), which dynamically adjusts prompts based on question type. APMT leverages Vision-Language models (BLIP2) for short-answer questions and refines descriptive responses using a Large Language Model (FastChat-T5) to enhance logical consistency and contextual awareness.

Methods:APMT classifies questions through a two-step process. The Question Response Type-Filter (QRT-F) distinguishes between short-answer and descriptive-answer types. Short-answer questions, such as yes/no, color, and counting, are further categorized by the Question Type Classifier (QTC) for optimized response generation using structured prompts in BLIP2. Descriptive questions are initially processed by BLIP2 and then refined by FastChat-T5 for improved clarity and coherence. To further enhance response quality, APMT employs adaptive prompt engineering. Short-answer questions are structured to maximize factual accuracy, while descriptive-answer questions incorporate image-caption-based prompting (CI, CI+CoT) to strengthen reasoning and coherence.

Results and Conclusion: Evaluations on two benchmark datasets demonstrate APMT's effectiveness. On the TDIUC dataset, APMT significantly improved short-answer accuracy, with BLEU scores increasing by a factor of 60. For descriptive-answer questions on the VCR dataset, image-caption-based prompts (CI, CI+CoT) yielded the highest BLEU, METEOR, ROUGE, and CIDEr scores, enhancing explanation quality. These results confirm that APMT improves both factual precision and reasoning-driven answers, leading to overall better performance in VQA tasks.



Question Processing with APMT



Explanation Generation through APMT

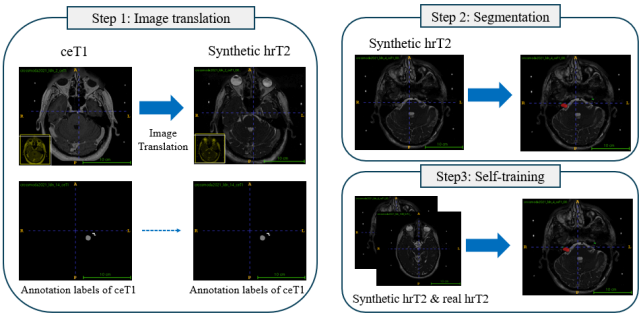
RESEARCH

Title	Cross-Modality Domain Adaptation for Medical Image Segmentation
Keywords	Unsupervised Domain Adaptation, Medical Image Segmentation, Brain MRI
Role	Experimental Design, Model Development and Testing

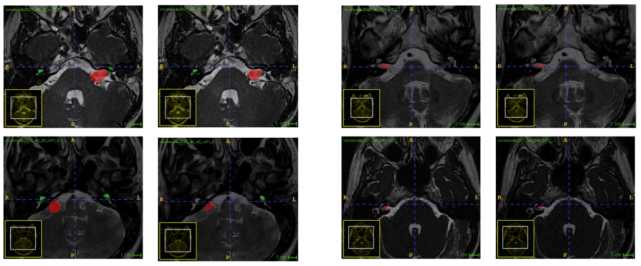
Background and objective: Unsupervised domain adaptation (UDA) is crucial for addressing performance drops with unseen modalities in medical image segmentation. While ceT1 scans offer high accuracy, they rely on costly gadolinium agents with safety risks. hrT2 scans are safer and cost-effective but lack sufficient contrast for accurate segmentation. This study addresses these challenges by using unpaired ceT1-hrT2 data with domain translation and self-training to enable accurate 3D segmentation of vestibular schwannoma (VS) and cochlea.

Methods: We propose a target-aware domain translation network using QSAttn to convert T1 scans into pseudo T2 images while preserving structures like VS and cochlea. A teacher-student self-training scheme bridges the gap between pseudo and real T2 data by leveraging T1 annotations to refine segmentation performance on T2 images.

Results and Conclusion: QSAttn with a segmentation decoder achieved Dice scores of 66.45% for VS and 68.75% for cochlea using pseudo T2 scans, while combining pseudo and real T2 data revealed domain gaps, reducing scores (VS: 58.13%, cochlea: 40.95%). This study highlights the effectiveness of target-aware translation and self-training but underscores the need for enhanced 3D translation networks and extended self-training to improve segmentation accuracy and clinical utility.



Overview of the Proposed UDA Framework



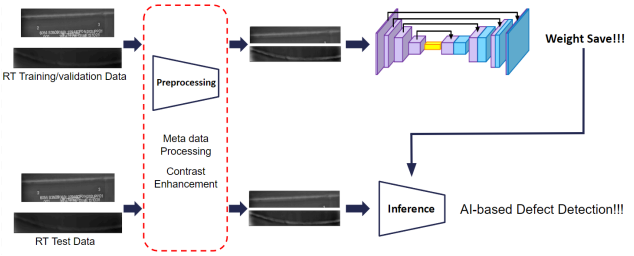
Qualitative test results trained on both pseudo T2 and real T2 data(Left : Ground Truth, Right : Prediction)

Title	AI Model Development for Welding Defect Detection
Keywords	Automated Welding defect detection, Radiographic images, Object Detection
Role	Experimental Design, Data Preprocessing, Model Development and Testing

Background and objective: Conventional radiography testing (RT) is slow, costly, and relies on manual interpretation. This study proposes an AI-driven Digital Radiography Test (DRT) to streamline workflows, reduce costs, and improve defect detection and classification efficiency.

Methods: We classify four defect types (Normal, Others, Porosity, Slag) using EasyOCR for metadata removal and CLAHE for contrast enhancement. Detection models were trained with data augmentation and evaluated by F1-scores. Non-max suppression refined bounding boxes, and weighted scores prioritized Porosity and Slag for model selection.

Results and Conclusion: Using test sets, Our model achieved the highest F1-scores for Porosity (78.2%), Slag (89%), and overall (89%), with a final weighted score of 84.68. These improvements, attributed to CLAHE and non-max suppression, enhanced defect visibility and reduced redundancy. AI-driven DRT offers faster, more cost-effective defect detection with superior performance and streamlined workflows.



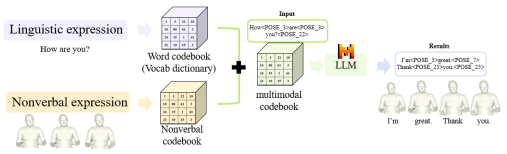
Model Development Strategy

Title	Development of a Multimodal Conversational Model
Keywords	Conversational AI, Human-Agent Interaction, Multimodal Learning
Role	Experimental Design, Data Preprocessing and Analysis of Collected Data, Model Training and Testing

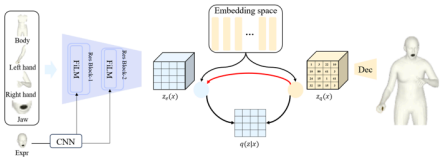
Background and objective: Nonverbal expressions like gestures and facial cues enrich communication but are often overlooked in text-only conversational AI, limiting natural interaction and engagement. To address this, we propose a multimodal conversation agent that integrates speech recognition and facial expression analysis to generate gestures, facial responses, and dialogue.

Methods: Using 10,000 hours of dual-speaker YouTube videos, we extract word, sentence, and utterance-level timestamps with whisperx and capture 3D nonverbal cues (e.g., gestures, facial expressions) via OSX. These are quantized into a unified multimodal codebook using VQ-VAE with FiLM conditioning layers, merging verbal and nonverbal inputs. The LLM (Mistral 7B with QLoRA) is trained on combined cues using reinforcement learning.

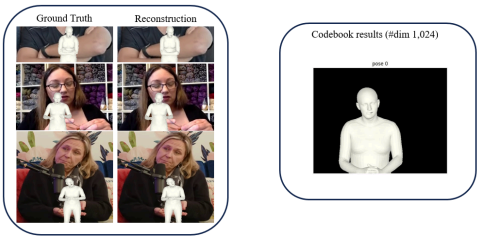
Results and Conclusion: Our multimodal agent processes 10,000 hours of video, generating 60,000 segments and 1.7M utterances with 3D nonverbal parameters. Using VQ-VAE with FiLM-based convolution and a joint loss setup, the model achieves superior reconstruction and perplexity. By appending quantized nonverbal tokens to text, it effectively unifies linguistic and nonverbal inputs, advancing conversational AI beyond text-only systems.



Multimodal Data Processing in Conversational AI



VQ-VAE Training Workflow



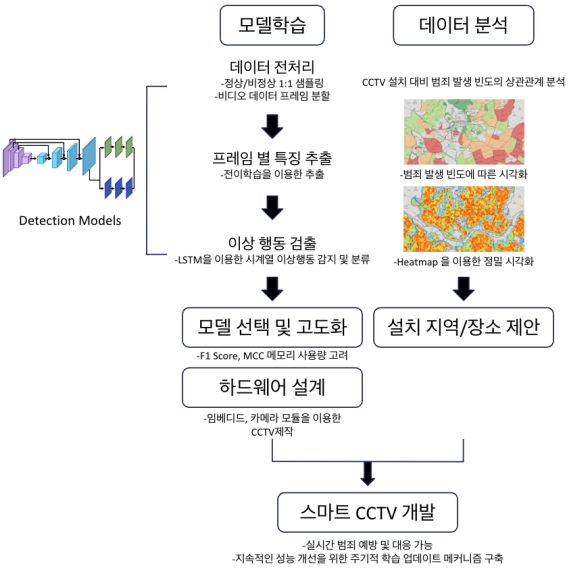
VQ-VAE Qualitative Results (Left : Reconstruction , Right: Codebook)

Title	Development of AI-Based Real-Time Anomalous Behavior Detection and Classification Model
Keywords	Anomaly Detection, Data Analysis and Visualization, Smart CCTV, Model Selection
Role	Experimental Design, Data Preprocessing, Data Analysis and Visualization, Model FineTuning, Model Testing and Selection

Background and objective: Seoul's 92,991 CCTV units primarily record footage without real-time crime detection. This study proposes an AI-based anomaly detection model for real-time CCTV analysis, targeting an F1-score of at least 70% under class-imbalanced conditions. Additionally, we analyze domestic crime data to develop smart CCTV deployment strategies for optimal coverage and crime prevention.

Methods: Using the UCF-Crime dataset, we conduct two experiments. Experiment 1 tests six CNN backbones with LSTM to classify normal vs. anomalous behavior, evaluating AUC and model size. Experiment 2 focuses on five major crimes (Abuse, Assault, Fighting, Robbery, Stealing) and adjusts class ratios based on crime statistics, using F1-Score, MCC, and model size to select the final model. Crime and CCTV data are also analyzed to model the impact of installations and identify high-risk areas.

Results and Conclusion: EfficientNet Lite3 outperforms ResNet18 with 34% higher F1-score, 5% higher MCC, and nine times less memory usage, emerging as the best backbone. Linear regression shows that one additional CCTV camera reduces crime by 1.26 incidents, with apartment complexes having the highest crime count (~140,000 cases). Vulnerable districts like Geumcheon and Gwanak are identified for AI-driven CCTV deployment to improve crime prevention and rapid response.



Workflow for Model Training and Data