WORKTYPE

AI/ML
Multi-modal Learning
Medical Imaging

PORTFOLIO

Soojin Hwang

CONTACT

+82 010 5427 6415

soojin0622@postech.ac.kr

AI Researcher

Soojin Hwang.

B 2000.06.22

P 010 5427 6415

M soojin0622@postech.ac.kr

W https://developing-soosoo.tistory.com/

EDUCATION

2024.03 - Present **POSTECH**
GPA 3.93/4.30

2019.03 - 2024.02 **Inha Unversity**
GPA 4.22/4.50

2022.09 - 2023.02 **University of Hull, England**

RESEARCH INTEREST

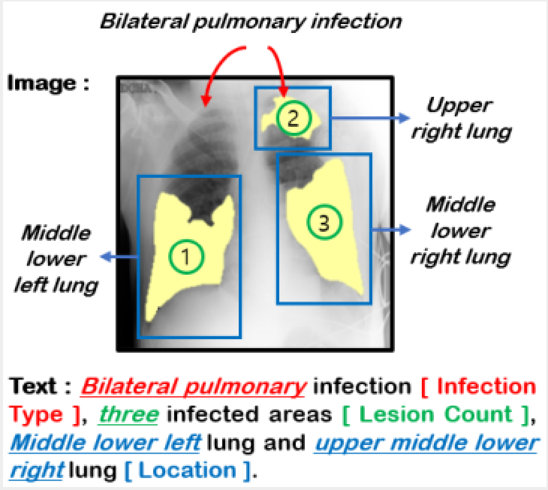
- Multi-modal Learning
- Computer Vision
- Medical Imaging
- Large Language Model

Publication

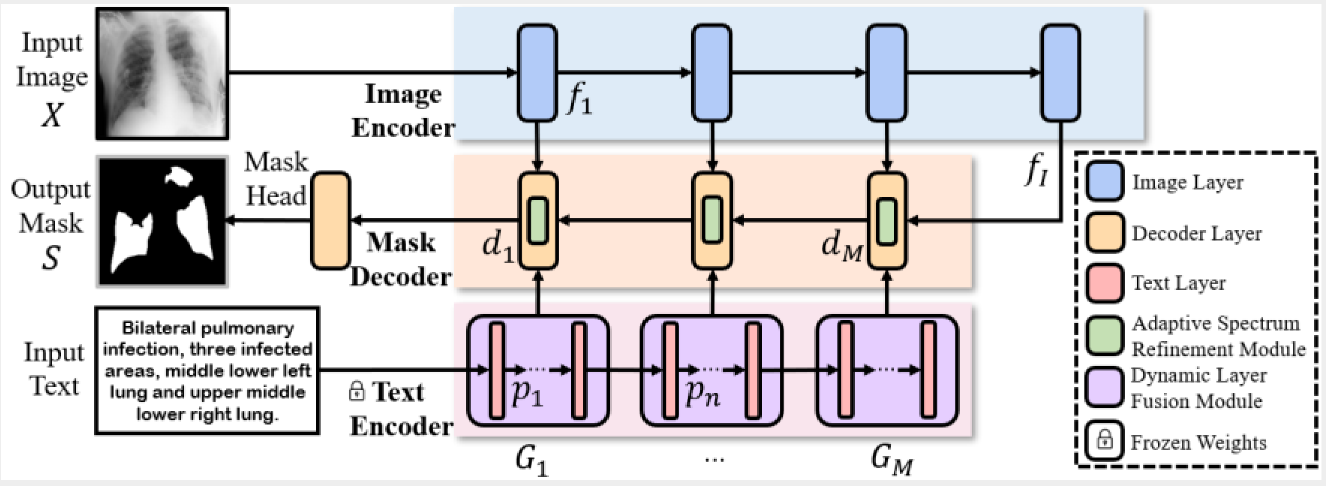
HiMix : Hierarchical Visual-Textual Mixing for Segmentation (WACV 2026)

Soojin Hwang*, Jaeyoon Sim*, Won Hwa Kim

- Overview
  - Lesion segmentation aids in diagnosing diseases but is limited by the cost and availability of annotated medical datasets.
  - Although recent work has utilized clinical texts to guide segmentation, most methods rely on a single text embedding, which limits the use of rich linguistic information.
- Key contributions
  - Hierarchical Fusion: HiMix integrates high-level semantics from clinical text with fine-grained visual features throughout the decoding process.
  - Dynamic Layer Fusion (DLFM)
    - Applies dynamic, hierarchy-aware routing with learnable fusion of multi-layer text embeddings across decoder stages.
  - Adaptive Spectrum Refinement (ASRM)
    - Enhances global and local visual representations by adaptively adjusting feature resolution at each decoding stage.



Example of an image-prompt describing structured radiologic findings



Architecture of HiMix

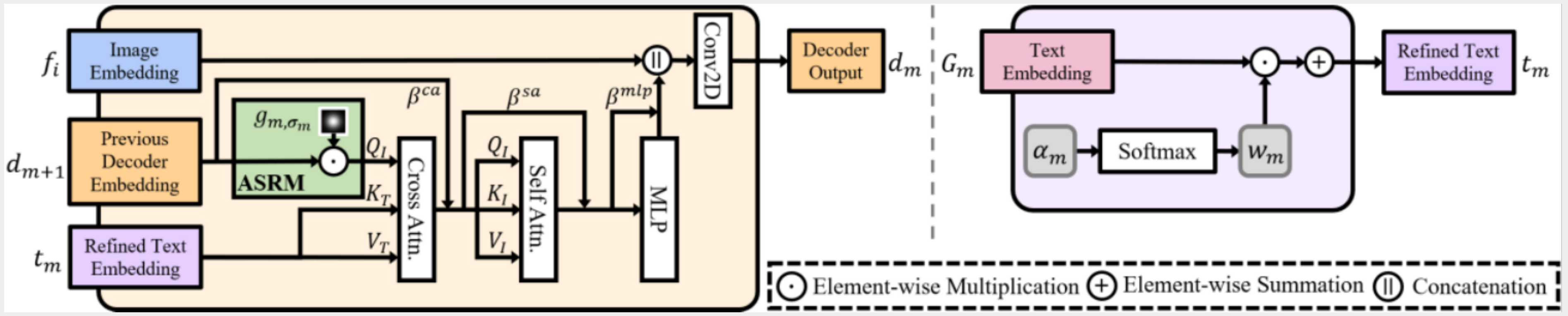


Illustration of key modules in HiMix. Left : Decoder with adaptive spectrum refinement module (ASRM). Right : Dynamic layer fusion module (DLFM)

Publication

HiMix : Hierarchical Visual-Textual Mixing for Segmentation (WACV 2026)

Soojin Hwang*, Jaeyoon Sim*, Won Hwa Kim

- Results
  - **Superior Performance:** HiMix outperforms both uni-modal and multi-modal baselines, including SOTA models.
  - **Few Parameters:** HiMix performs well with only 44.7M parameters, significantly fewer than other models.

Approach	Type	Method	Param ↓ (M)	QaTa-COV19 DSC ↑	QaTa-COV19 IoU ↑	MosMedData+ DSC ↑	MosMedData+ IoU ↑	Kvasir-SEG DSC ↑	Kvasir-SEG IoU ↑
Uni-Modal (Image-Only)	CNN	UNet [37]	14.8	79.02	69.46	64.60	50.73	82.94	74.47
		UNet++ [56]	74.5	79.62	70.25	71.75	58.39	80.43	72.13
		AttnUNet [33]	34.9	79.31	70.04	66.34	52.82	81.31	73.74
		nnUNet [19]	19.1	80.42	70.81	72.59	60.36	85.06	74.01
	Transformer	Swin-UNet [5]	82.3	78.07	68.34	63.29	50.19	75.97	67.45
	Hybrid	UCTransNet [46]	65.6	79.15	69.60	65.90	52.69	78.21	65.25
		TransUNet [6]	105.0	78.63	69.13	71.24	58.44	79.67	71.14
	Mamba	VMUNet [38]	31.0	86.31	75.92	75.85	61.09	79.54	66.04
		H-vmunet [49]	31.0	86.26	75.84	76.37	61.78	82.25	69.85
Multi-Modal (Image-Text)	CNN	GLoRIA [17]	45.6	79.94	70.68	72.42	60.18	84.73	73.51
		ConVIRT [53]	35.2	79.72	70.58	72.06	59.73	84.27	72.82
		RecLMIS [18]	23.7	85.22	77.00	77.48	65.07	87.73	78.15
	Transformer	CLIP [35]	87.0	79.81	70.66	71.97	59.64	81.27	68.45
		MedCLIP [47]	137.0	86.54	76.27	69.14	52.84	78.84	65.08
		DMMI [15]	131.5	79.25	68.87	72.57	60.78	74.79	59.73
		RefSegformer [48]	195.0	84.09	75.48	74.98	61.70	79.78	66.37
		MedSAM [30]	**4.5**	78.49	69.11	54.22	42.22	86.69	79.24
	Hybrid	ViLT [23]	87.4	79.63	70.12	72.36	60.15	70.33	54.23
		LAVT [51]	118.6	79.28	69.89	73.29	60.41	70.59	54.55
		LViT [27]	29.7	83.66	75.11	74.57	61.33	87.17	77.26
		SLViT [34]	114.6	84.13	75.66	75.01	61.83	86.61	76.38
		GuideDecoder [55]	44.0	89.78	81.45	77.75	63.60	88.31	79.07
		MMI-UNet [4]	56.2	90.88	83.28	78.42	64.50	91.43	84.57
		HiMix (Ours)	44.7	**91.17**	**83.78**	**79.44**	**65.90**	**92.18**	**85.50**

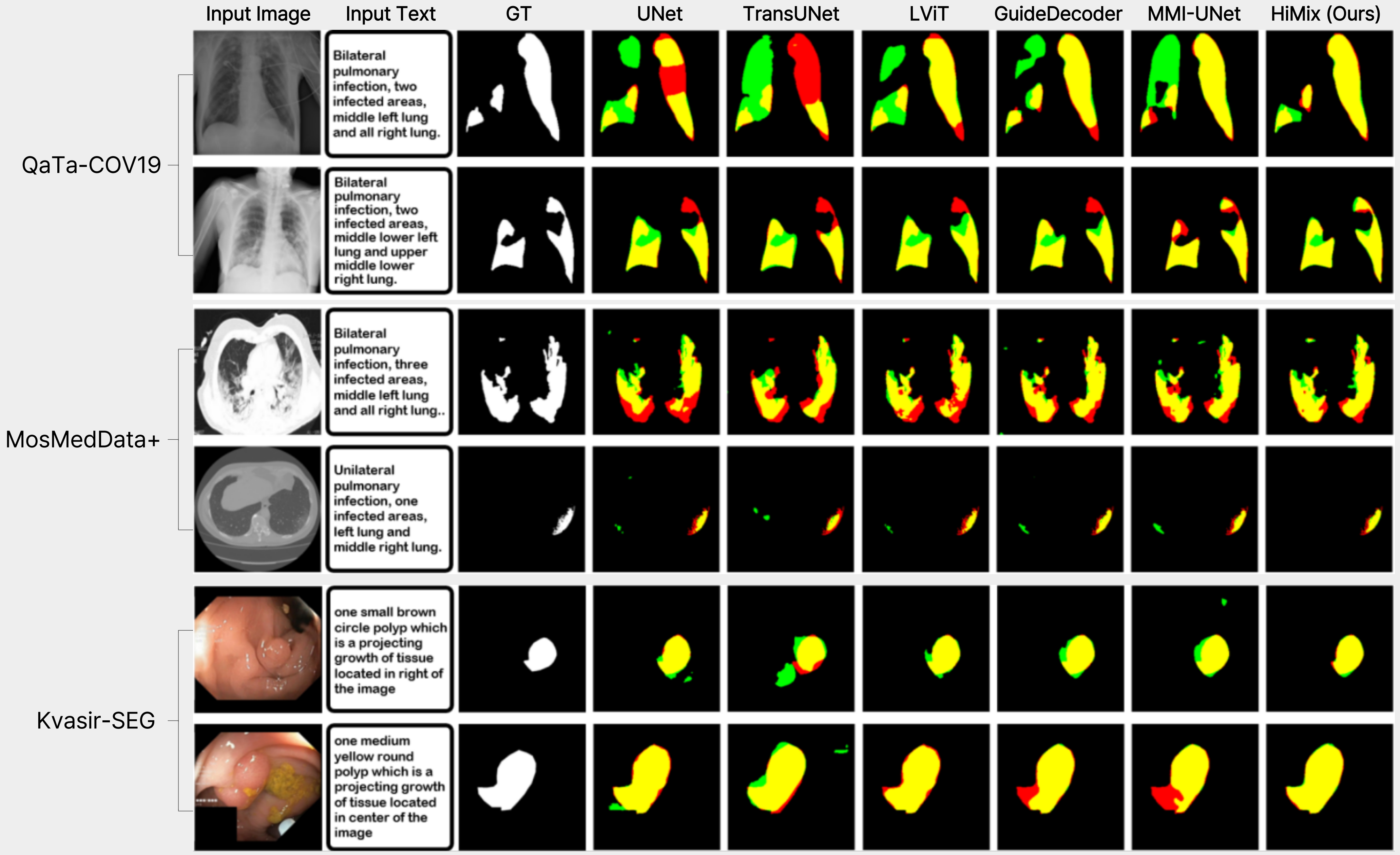
The best and second-best results are highlighted in **bold** and underlined, respectively

Publication

HiMix : Hierarchical Visual-Textual Mixing for Segmentation (WACV 2026)

Soojin Hwang*, Jaeyoon Sim*, Won Hwa Kim

- Key Findings
  - Hierarchical Alignment: Aligning hierarchical text features with visual decoding stages provides more accurate and contextually grounded segmentation.
  - Generalization: The model generalizes well to unstructured text formats, showing robustness in real-world clinical applications.
- Conclusion
  - HiMix shows that hierarchical fusion of text and image modalities leads to fine-grained and context-aware lesion segmentation, across diverse datasets and text formats.

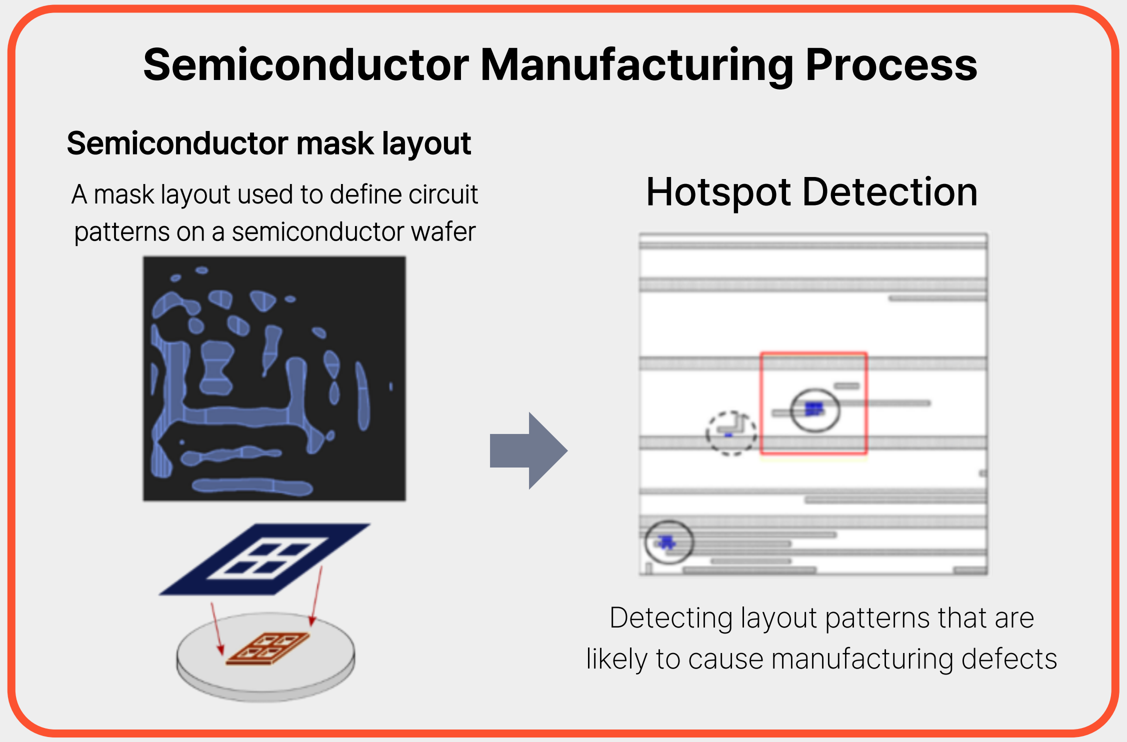


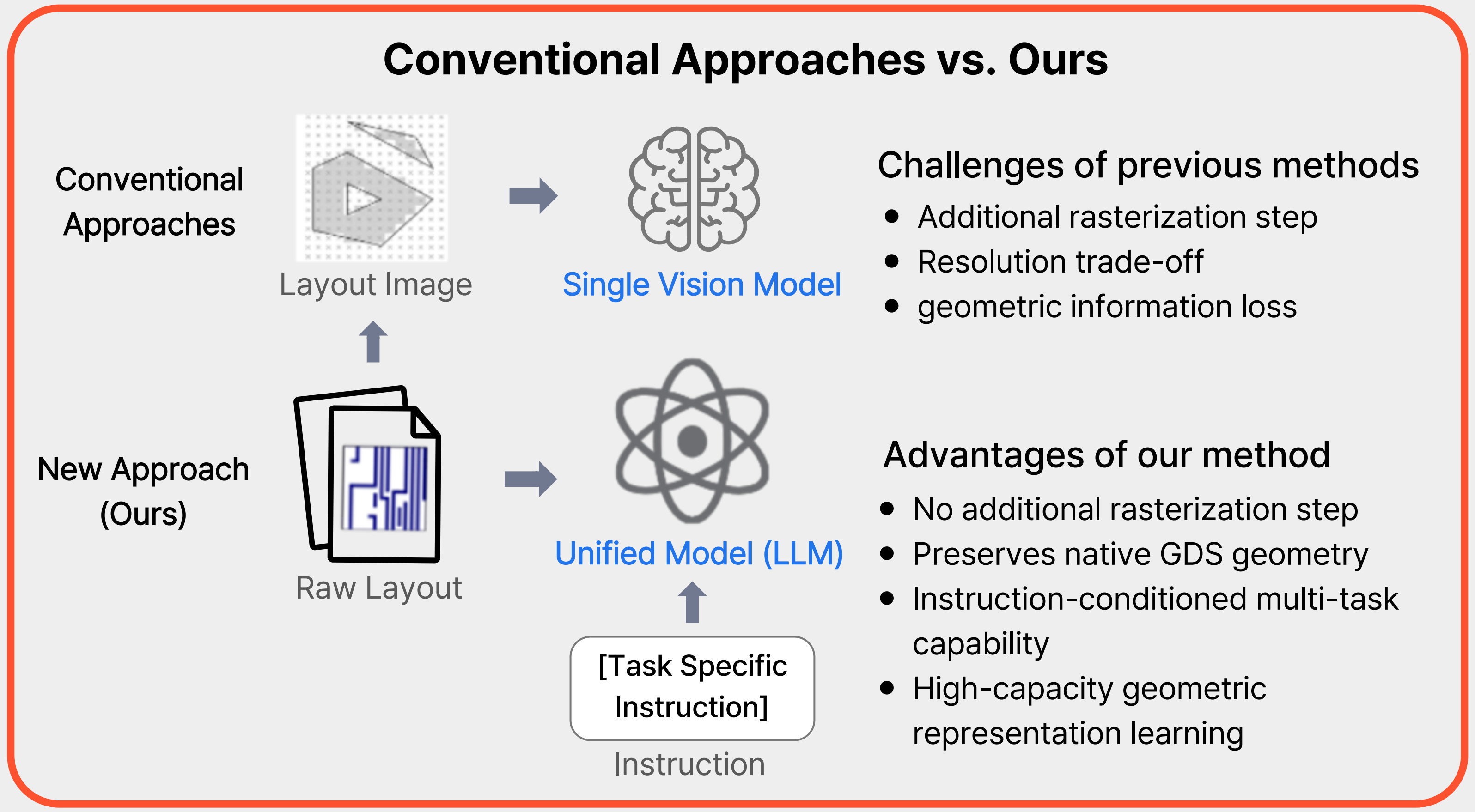
Project

Semiconductor unified task

Joint Research with Samsung DS AI center (2025.04 - Present)

- Goal
  - Develop a unified model for semiconductor layout (GDSII) design tasks, with a focus on lithography hotspot detection and extensibility to DRC (Design Rule Check) and OPC (Optical Proximity Correction)
- Key Contributions
  - Point-based Representation for GDS Layout Data
    - Proposed a native point-based representation that directly models GDSII Polyon vertives
    - Adapted a point-cloud foundation model to learn geometric patterns in 2D seminconductor layouts, and validate the approach on the ICCAD 2019 hotspot detection benchmark
  - Instruction-based Learning for Layout Hotspot Detection
    - Constructed an instruction-sample pair dataset tailored for hotspot detection and fine-tuned a multimodal LLM to perform layout reasoning under diverse linguitic instructions
    - Demonstrated robustness and generalization through instruction diversity via ablation studies



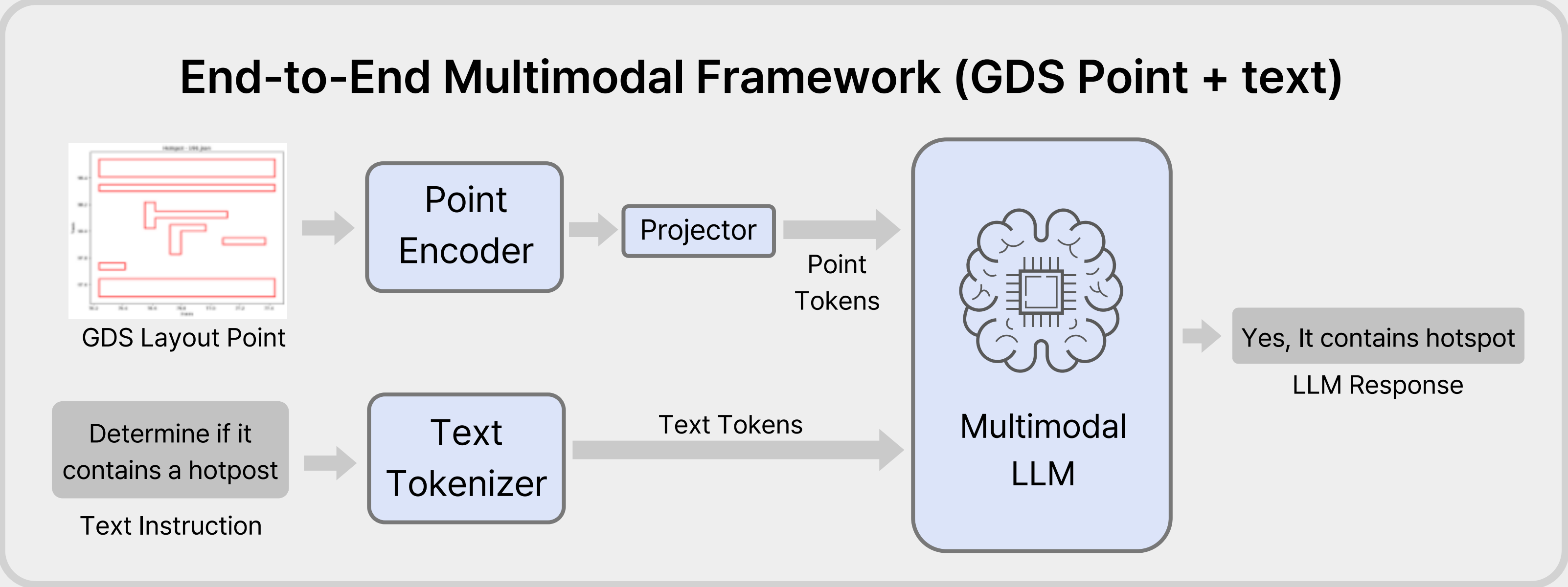


Project

Semiconductor unified task

Joint Research with Samsung DS AI center (2025.04 - Present)

- Impact
    - First to directly apply GDS layouts to an LLM, achieving reasonable performance for lithography hotspot detection without rasterization.
    - Demonstrated effective geometry understanding by learning directly from GDS layout representation.
    - Established an end-to-end, scalable training pipeline that integrates point-based geometry encoding with instruction-tuned LLMs (GPU A6000, A100, H100)
    - Validated the extensibility of instruction-tuned LLMs beyond hotspot detection,



Hotspot Detection Result (ICCAD 2019 benchmark [1])

Split	DAC'17 [2]		TCAD'18 [3]		Ours	
	Recall ↑	FA ↓	Recall ↑	FA ↓	Recall ↑	FA ↓
Test set 1	67.83	2.00	77.02	4.10	80.39	66.42
Test set 2	99.44	97.82	99.43	97.48	77.11	69.27

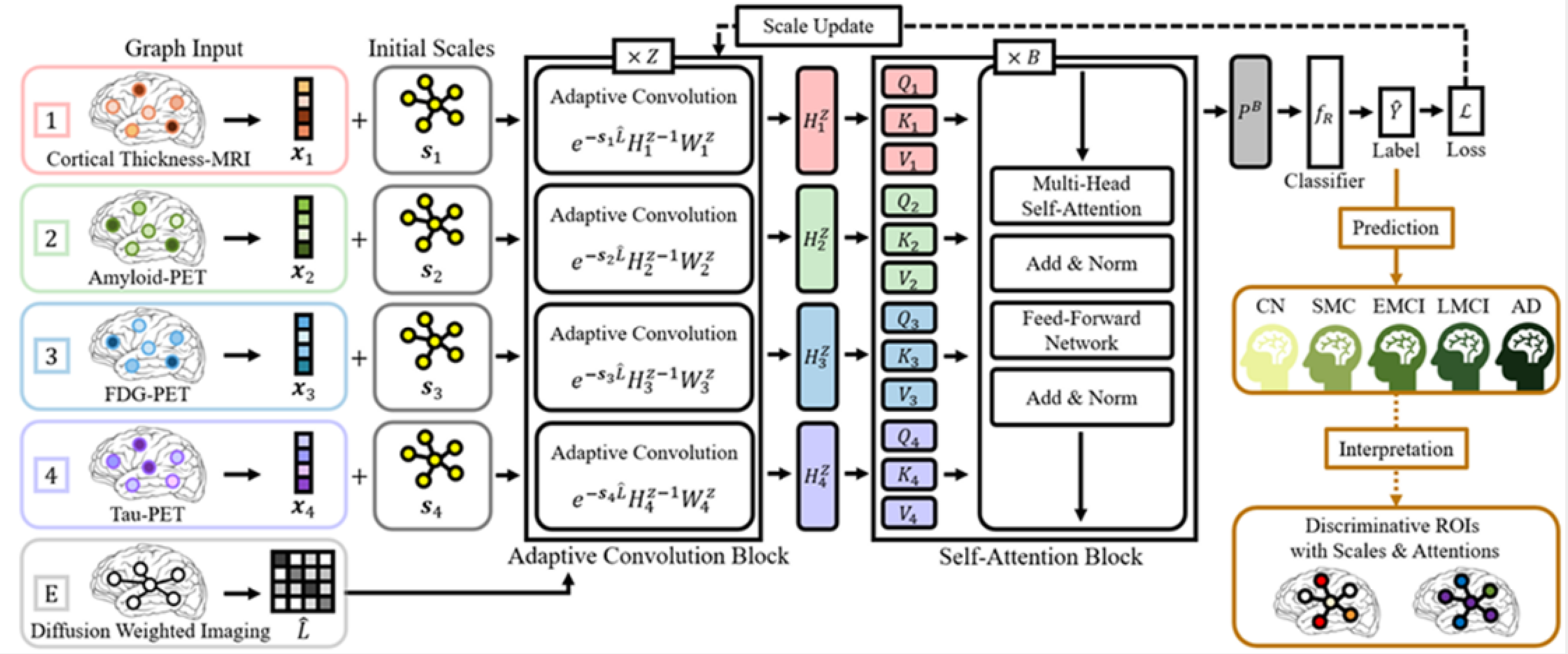
[1] Reddy, Gaurav Rajavendra, and Kareem Madkour "Machine learning-based hotspot detection: Fallacies, pitfalls and marching orders." ICCAD. IEEE, 2019.
[2] H. Yang, J. Su, Y. Zou, B. Yu, and E. F. Young, “Layout hotspot detection with feature tensor generation and deep biased learning,” DAC, 2017.
[3] H. Yang, J. Su, Y. Zou, and Y. Ma, “Layout hotspot detection with feature tensor generation and deep biased learning,” DAC, 2018.

Publication

GTAD : Transformer-Guided Adaptive Diffusion for Multi-Modal Brain Networks (under review)

Jaeyoon Sim*, Soojin Hwang*, Guorong Wu, Won Hwa Kim

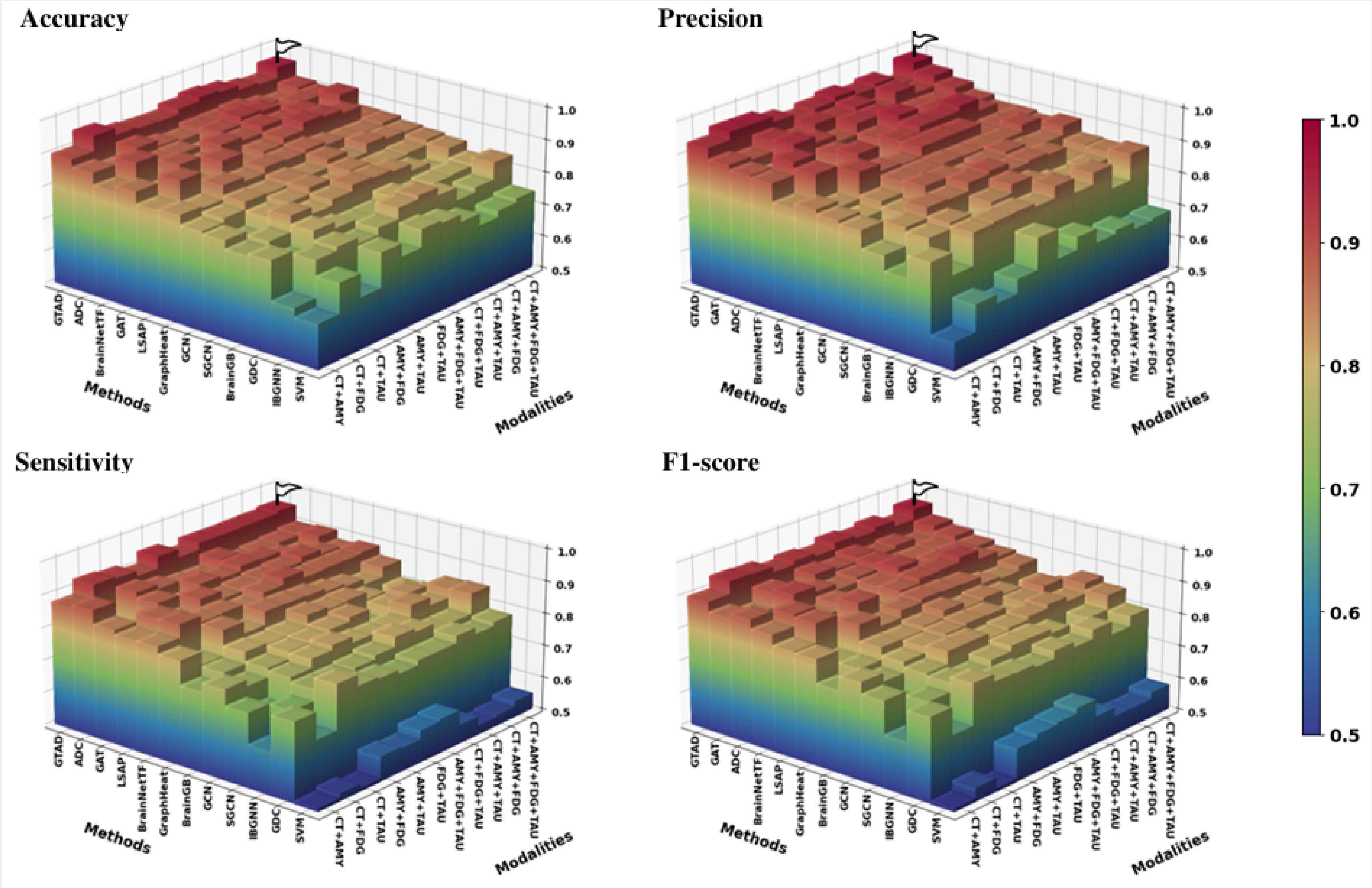
- Overview
  - GTAD is a multi-modal graph neural network for Alzheimer's Disease (AD) classification, integrating multiple neuroimaging modalities (e.g., MRI, PET).
  - It uses a transformer-guided diffusion process to capture both local and global dependencies in brain networks.
- Key Contributions
  - Graph Diffusion: Aggregates local features using heat kernels.
  - Transformer Guidance: Uses multi-head self-attention to capture long-range dependencies.
  - Modality-Specific Analysis: Improves interpretability by processing each modality separately.
- Results
  - GTAD outperforms SOTA models in AD classification, identifying key ROIs linked to AD progression across modalities.
  - It offers an interpretable framework, leveraging multi-modal data and adaptive learning for improved diagnosis and interpretation of Alzheimer's Disease.

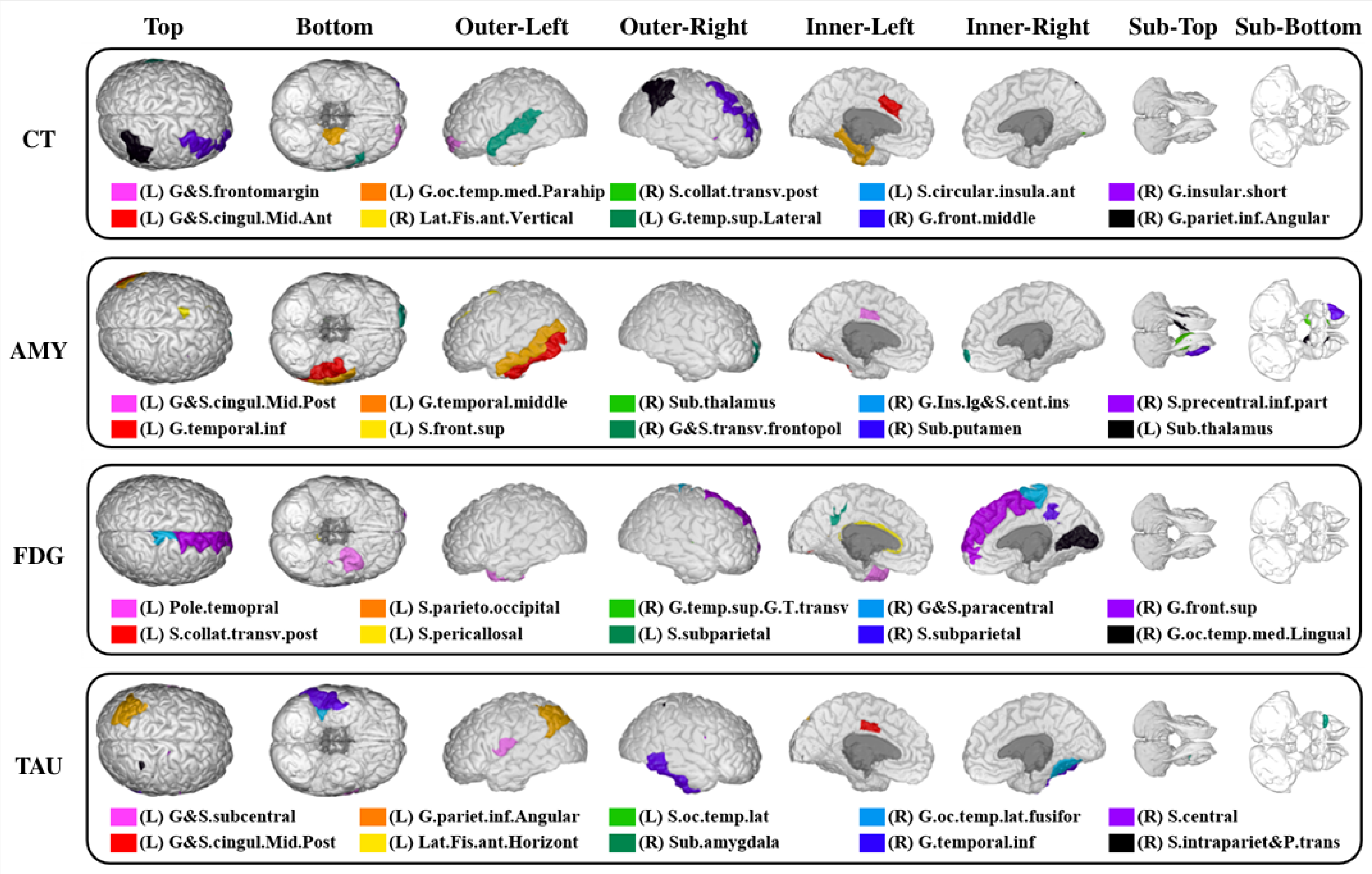


Publication

GTAD : Transformer-Guided Adaptive Diffusion for Multi-Modal Brain Networks (under review)

Jaeyoon Sim*, Soojin Hwang*, Guorong Wu, Won Hwa Kim



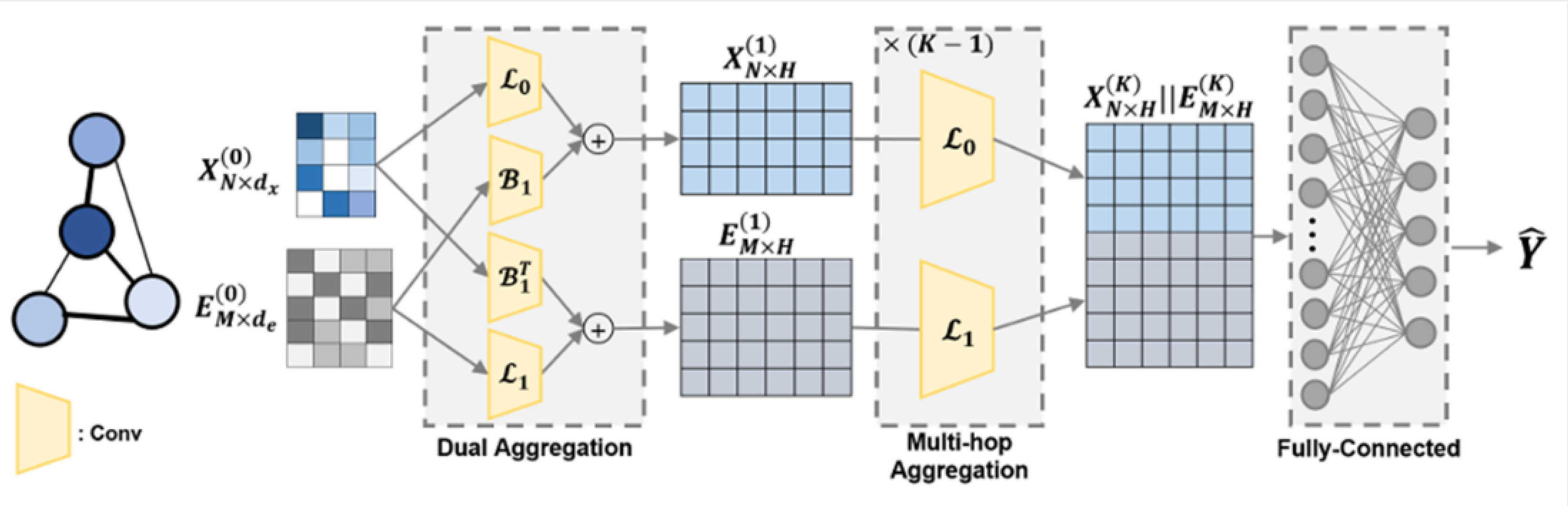


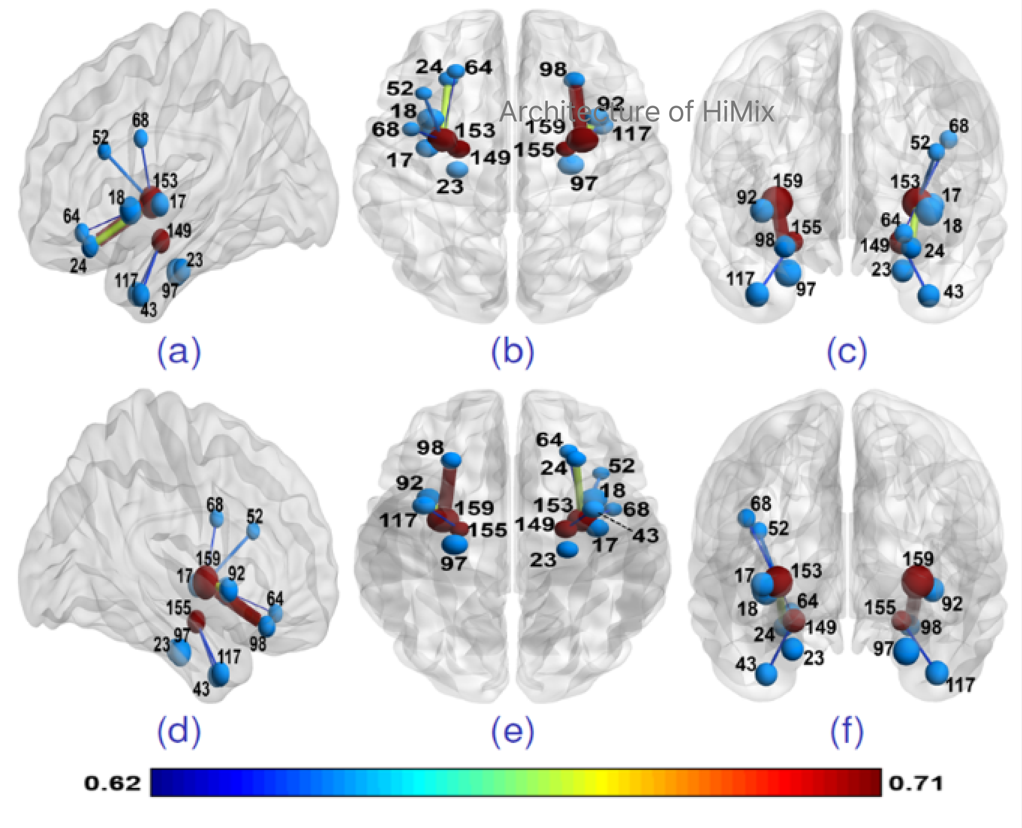
Publication

Multi-order Simplex-based GNN for Brain Networks (MICCAI 2024)

Yechan Hwang, Soojin Hwang, Guorong Wu, Won Hwa Kim

- Overview
  - The paper introduces a Multi-order Simplex-based Graph Neural Network for brain network analysis, incorporating both node and edge features via a dual aggregation framework using an incidence matrix.
- Key Contributions
  - Dual Aggregation: Simultaneous aggregation of node and edge features.
  - Separate Node/Edge Representation: Direct learning of node and edge embeddings.
  - Interpretability: Uses Grad-CAM to identify key ROIs and connectivities linked to Alzheimer's disease.
- Results
  - Outperforms SOTA models on the ADNI dataset with improved classification performance and interpretability.





Project

KIST Internship : AWS-based Video Analysis

- Overview
  - Built an end-to-end AI pipeline on AWS : S3, Rekognition, Kinesis Video Stream, SNS/SQS
  - Integrated Gstreamer + Ubuntu for real-time webcam streaming
- Key Contributions
  - Automated PPE(Personal Protective Equipment) detection and face recognition using Rekognition APIs
  - Designed custom label models for domain-specific object detection
- Impact
  - Hands-on experience in cloud-based AI service development
  - Practical understanding of real-time multimodal pipeline

Thank you for your time

AI Researcher Porfolio
by Soojin Hwang

CONTACT

Phone Number | 010 5427 6415

e-mail | soojin0622@postech.ac.kr