Hierarchical Region-Aware Multi-Granularity Mamba for White Matter Lesion Segmentation

Dahye Lee[1] and Kwanseok Oh[1,2,*]

[1]DEEPNOID Inc., Seoul, Korea [2]Department of Artificial Intelligence, Korea University, Seoul, Korea *Corresponding Author

Email: {**dhlee, gsoh**}@deepnoid.com

GitHub

MICCAI 2026 29TH INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING & COMPUTER ASSISTED INTERVENTION Sept. 27- Oct. 1, 2026 Strasbourg Convention Center, FRANCE

Motivation

Clinical Motivations

- White matter hyperintensities (WMH) and multiple sclerosis (MS) lesions are important imaging biomarkers of cerebrovascular and demyelinating disease.
- On T2-FLAIR, they are small, spatially scattered and low-contrast with indistinct boundaries.

Technical Limitations

- **CNN / ViT** — Large-scale annotations required
- **SAM** — Large token size loses local detail → Weak on faint, diffuse lesions
- **+ CNN** — Attention-based feature aggregation → Local awareness added, quadratic cost
- **+ Mamba** — Linear cost, fixed causal scan → Misaligned with non-causal MRI & weak distant lesion interactions

Problem Definition

Existing methods lack efficient, fine-grained regional propagation for subtle lesions and indistinct boundaries.

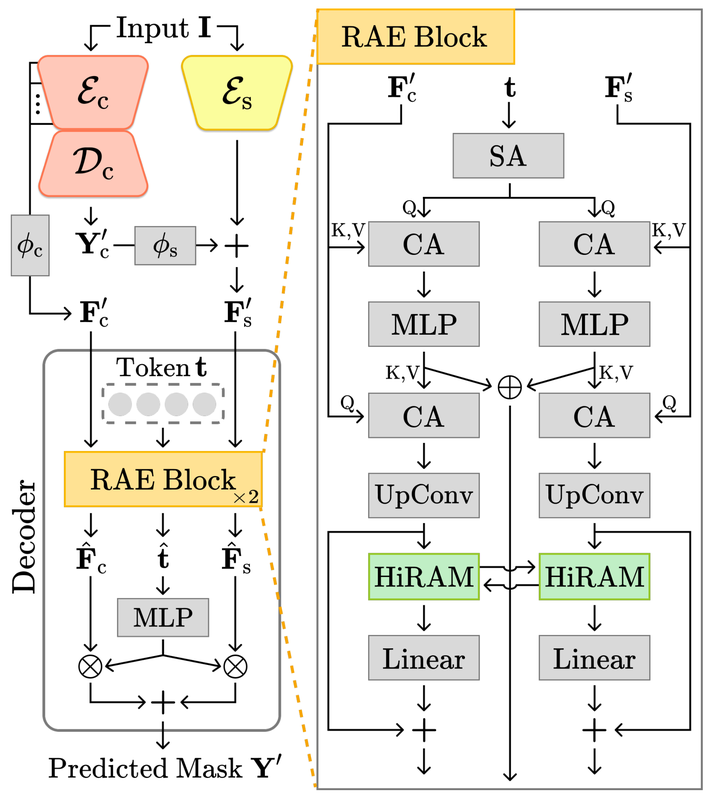
Our Contributions

1. **HiRAM —** Hierarchical region-aware multi-granularity Mamba inside the decoder.
2. **RPG —** Region-tailored priors for interior (in) / boundary (bd) / background (bg) regions.
3. **MUG —** Confidence-driven non-causal region-specific SSM with holistic SS2D propagation.
4. **Results —** SOTA on two public benchmark datasets, outperforming 10 baselines.

Overall Framework

Two complementary branches and unified decoder

- CNN branch: Multi-scale local features → Resize & Fuse → $\mathbf{F'}_c$
- SAM branch: Global semantic features → CNN mask injection→ $\mathbf{F'}_s$
- RAE blocks + Learnable tokens: Joint feature refinement
- Lightweight MLP: Final mask prediction



Training Objective

- Segmentation loss: Dice + BCE on final mask and coarse CNN mask
- Region decomposition loss: KL divergence between GT interior / boundary masks and every RPG log-probability map → Emphasize foreground structure

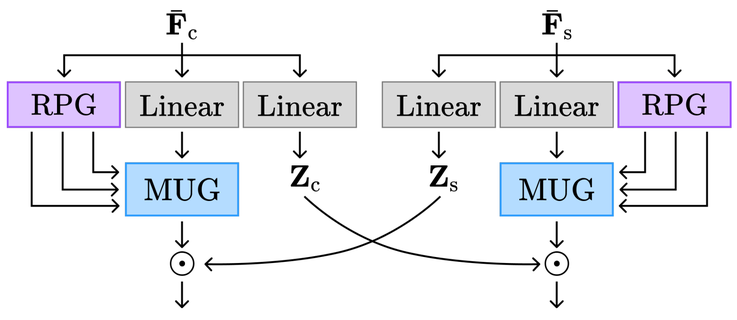
$$\mathcal{L}_{\text{dec}} = \sum_{l}\sum_{b}\sum_{r\in\{\text{in,bd}\}} \frac{1}{|\Omega_r^l|}\sum_{i\in\Omega_r^l} \text{KL}\left(g(\mathbf{Y}_r(i),l)||\mathbf{M}_{b,r}^l(i)\right)$$

$$\mathcal{L} = \mathcal{L}_{\text{DiceBCE}}(\mathbf{Y}',\mathbf{Y}) + \mathcal{L}_{\text{DiceBCE}}(\mathbf{Y}'_c,\mathbf{Y}) + \mathcal{L}_{\text{dec}}$$

- $\mathbf{M}_{b,r}^l$: RPG log-probability map for region r
- Ω_r^l : Spatial positions assigned to region r
- $\mathbf{Y}_r$: GT region mask
- $g(\cdot,l)$: Interpolation function

HiRAM

- HiRAM acts symmetrically on the two branch features.
- Each branch is decomposed by RPG, propagated by MUG, and finally modulated by the cross-branch embeddings $\mathbf{Z}_c$ and $\mathbf{Z}_s$ for adaptive interaction.



RPG (Region Prior Generator)

Confidence-based region decomposition for region-tailored prior generation

- **Region decomposition**

  Linear projection with LogSoftmax yields a log-probability map $\mathbf{M}_r$ over interior / boundary / background.

- **Region assignment**

  Gumbel-Softmax turns $\mathbf{M}_r$ into a differentiable one-hot routing map, yielding region-wise index set $\mathcal{S}_r$.

- **Confidence ordering**

  Region confidences within each region-wise $\mathcal{S}_r$ are sorted in descending order as $\mathbf{c}_r^{\downarrow}$, defining the permutation π_r

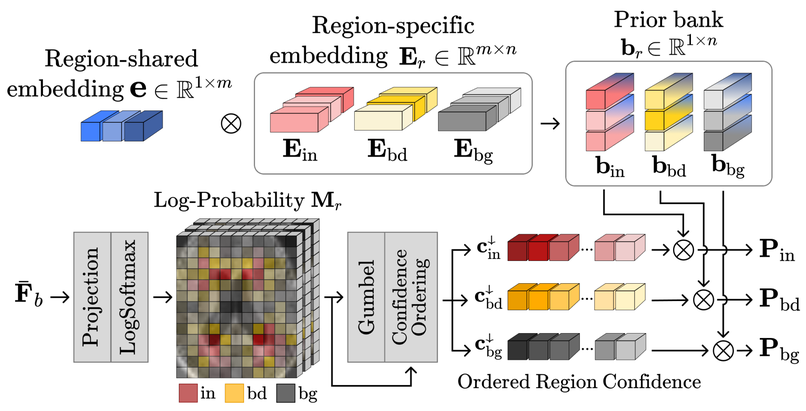
- **Prior bank**

  Prior bank $\mathbf{b}_r$ come from semantic decoupling; a shared basis $\mathbf{e}$ and region-specific coefficients $\mathbf{E}_r$ → Unified semantics with region-aware specialization.

- **Region-tailored prior**

  Sorted confidence $\mathbf{c}_r^{\downarrow}$ is combined with prior bank $\mathbf{b}_r$ via matrix multiplication to produce region-tailored prior $\mathbf{P}_r$ → Confident pixels impose *stable and meaningful prior*.

$$\mathbf{b}_r = \mathbf{e}\otimes\mathbf{E}_r,\quad \mathbf{e}\in\mathbb{R}^{1\times m},\quad \mathbf{E}_r\in\mathbb{R}^{m\times n}$$
$$\mathbf{P}_r := \mathbf{c}_r^{\downarrow}\otimes\mathbf{b}_r$$



MUG (Multi-Granularity Mamba)

Multi-granularity propagation with RA-SSM and global SS2D

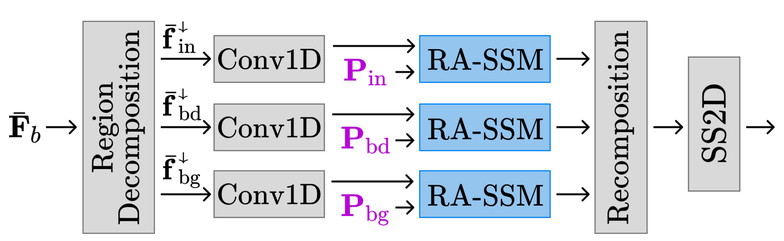
Regional propagation

- Features are partitioned by permutation π_r, → High-certainty tokens enter the recurrence first and stabilize early hidden states.
- Each ordered regional sequence passes a 1D convolution and a region-adaptive SSM (RA-SSM).
- Prior $\mathbf{P}_r$ acts as a region-conditioned readout bias → Re-weights the latent states without altering the state-transition dynamics.

$$h_t = \bar{\mathbf{A}}h_{t-1} + \bar{\mathbf{B}}x_t,\quad y_t = (\mathbf{C}+\mathbf{P}_r)\,h_t + \mathbf{D}x_t$$

Global propagation

- The inverse permutation restores spatial order, then SS2D enforces global contextual coherence.



Implementation Details

Dataset

- MICCAI'17 WMH Segmentation Challenge (MWSC) : 170 scans, 60 train / 110 test split
- MS Lesion Segmentation (MSLesSeg) : 115 scans, 93 train / 22 test split

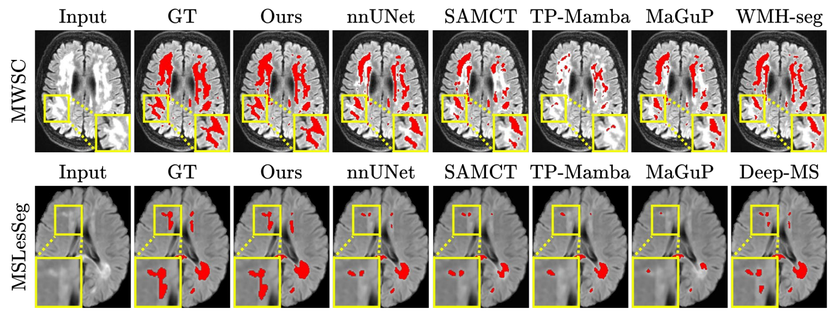
Evaluation Metrics

- Dice similarity coefficient (DSC)
- Absolute volume difference (AVD)

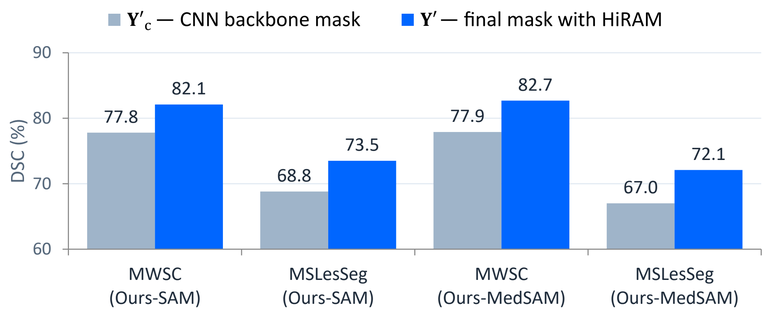
Results

- **Comparison with SOTA Methods**

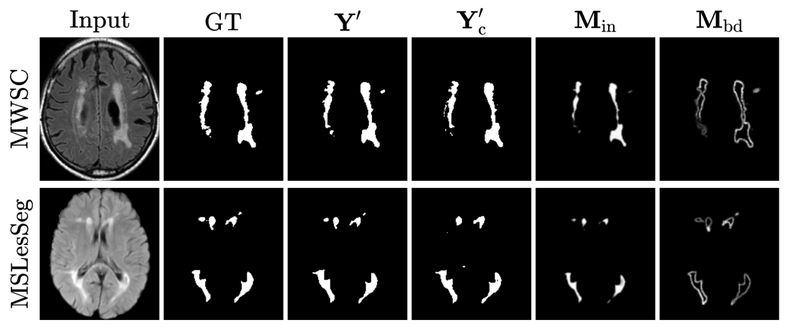
Method	MWSC (WMH)		MSLesSeg (MS)	
	DSC↑(%)	AVD↓(%)	DSC↑(%)	AVD↓(%)
nnUNet	74.9	23.3	68.7	30.1
SwinUNETR	69.7	26.6	67.5	29.4
SAM	35.8	68.5	27.6	69.8
MedSAM	56.9	57.1	45.1	42.4
SAMUS	74.0	18.9	63.8	33.2
SAMCT	75.8	21.5	66.5	25.3
TP-Mamba	74.6	20.6	66.9	26.0
MaGuP	72.0	21.8	63.8	29.9
WMH-seg	77.0	18.1	65.4	34.4
Deep-MS	75.1	20.8	70.2	25.8
Ours - SAM	**82.1**	**13.1**	**73.5**	**16.5**



- **Impact of HiRAM on the Backbone Mask**



- **Quality of Segmentation Mask & Region Decomposition**



- **Ablation Study on Key Components**

Case	HiRAM	RA-SSM	$\mathcal{L}_{\text{dec}}$	π_r	$\mathbf{P}_r$	MWSC DSC (%)	MWSC AVD (%)	MSLesSeg DSC (%)	MSLesSeg AVD (%)
1	–	–	–	–	–	69.9	28.9	66.4	30.5
2	✓	–	–	–	–	70.2	26.4	66.3	31.4
3	✓	✓	–	–	–	74.6	20.1	68.5	27.9
4	✓	✓	✓	–	–	77.5	16.3	70.2	20.3
5	✓	✓	✓	✓	–	79.0	15.8	71.8	18.8
6	✓	✓	✓	–	✓	80.4	15.2	72.1	18.6
Ours	✓	✓	✓	✓	✓	**82.1**	**13.1**	**73.5**	**16.5**

Conclusion

- HiRAM enables fine-grained WM lesion segmentation through region-aware feature modeling.
- HiRAM achieves SOTA performance across WMH (+5.1% DSC) and MS lesion (+3.3% DSC) segmentation.

Acknowledgments: This work was supported by the Technology Innovation Pro gram (or Industrial Strategic Technology Development Program) (RS-2025-02221011, Development of Medical-Specialized Multimodal Hyperscale Generative AI Technology for Global Integration) funded by the Ministry of Trade Industry & Energy (MOTIE, Korea).