CVPR
JUNE 3-7, 2026
DENVER
COLORADO

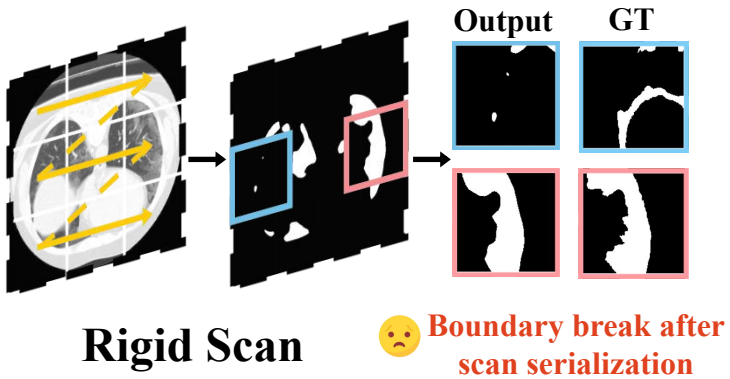
GeoSemba

Reconstructing State Space Model for Cross Paradigm Representation in Medical Image Segmentation

Xutao Sun[1], Jiarui Li[1], Junwen Liu[1], Yonggong Ren[1]

[1]School of Computer Science and Artificial Intelligence
Liaoning Normal University

Motivation: from scan order to interaction quality



Limitation 1

Mamba serializes 2D images into 1D token streams. Spatially adjacent tokens in the scan may be semantically mismatched, blurring discriminative evidence.

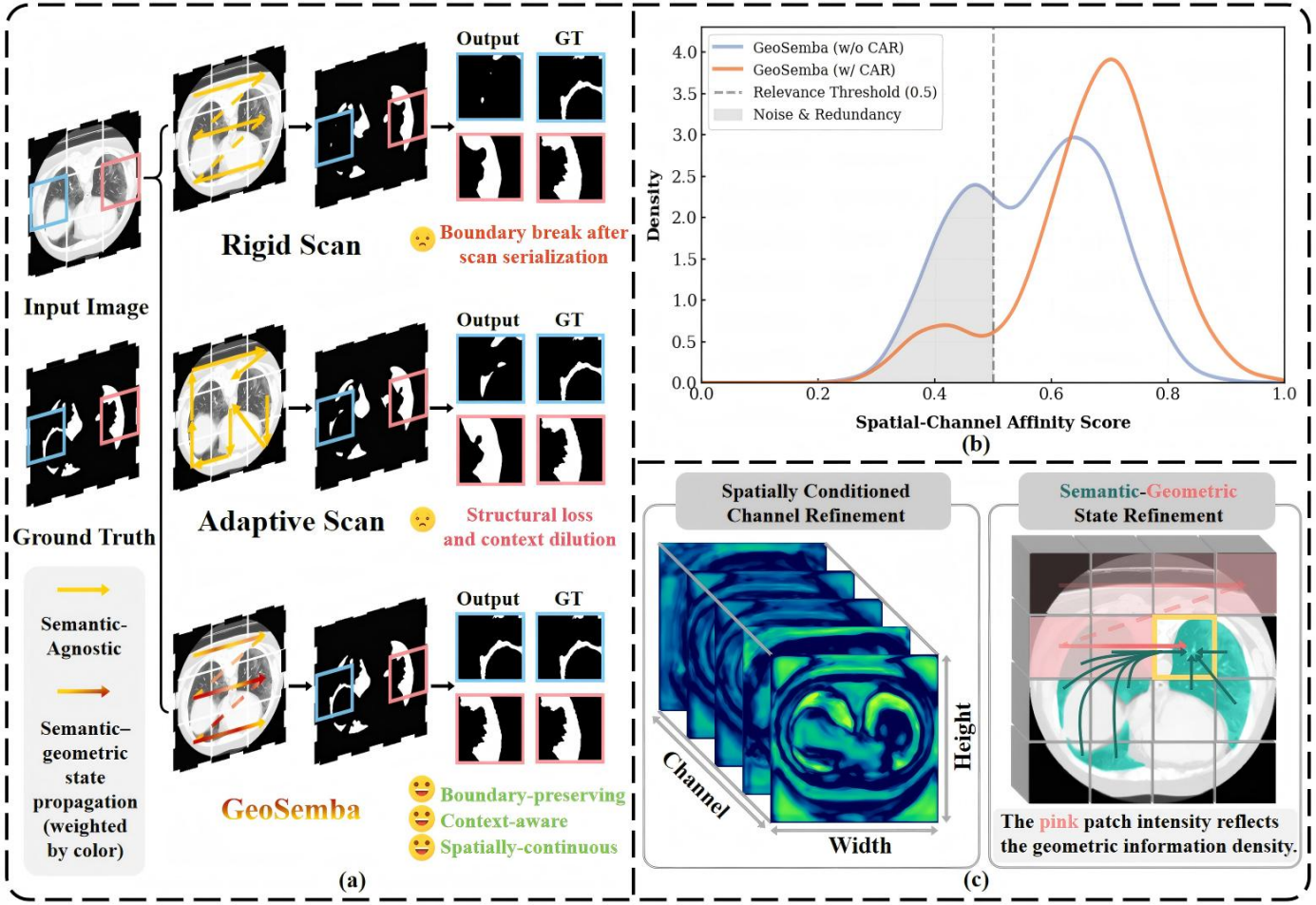
Limitation 2

Spatial interaction and channel discrimination are often treated separately, although medical cues arise from their joint effect.

Our question

Can a single-scan SSM preserve geometry-aware semantic propagation while modeling spatial-channel dependencies efficiently?

Main contributions

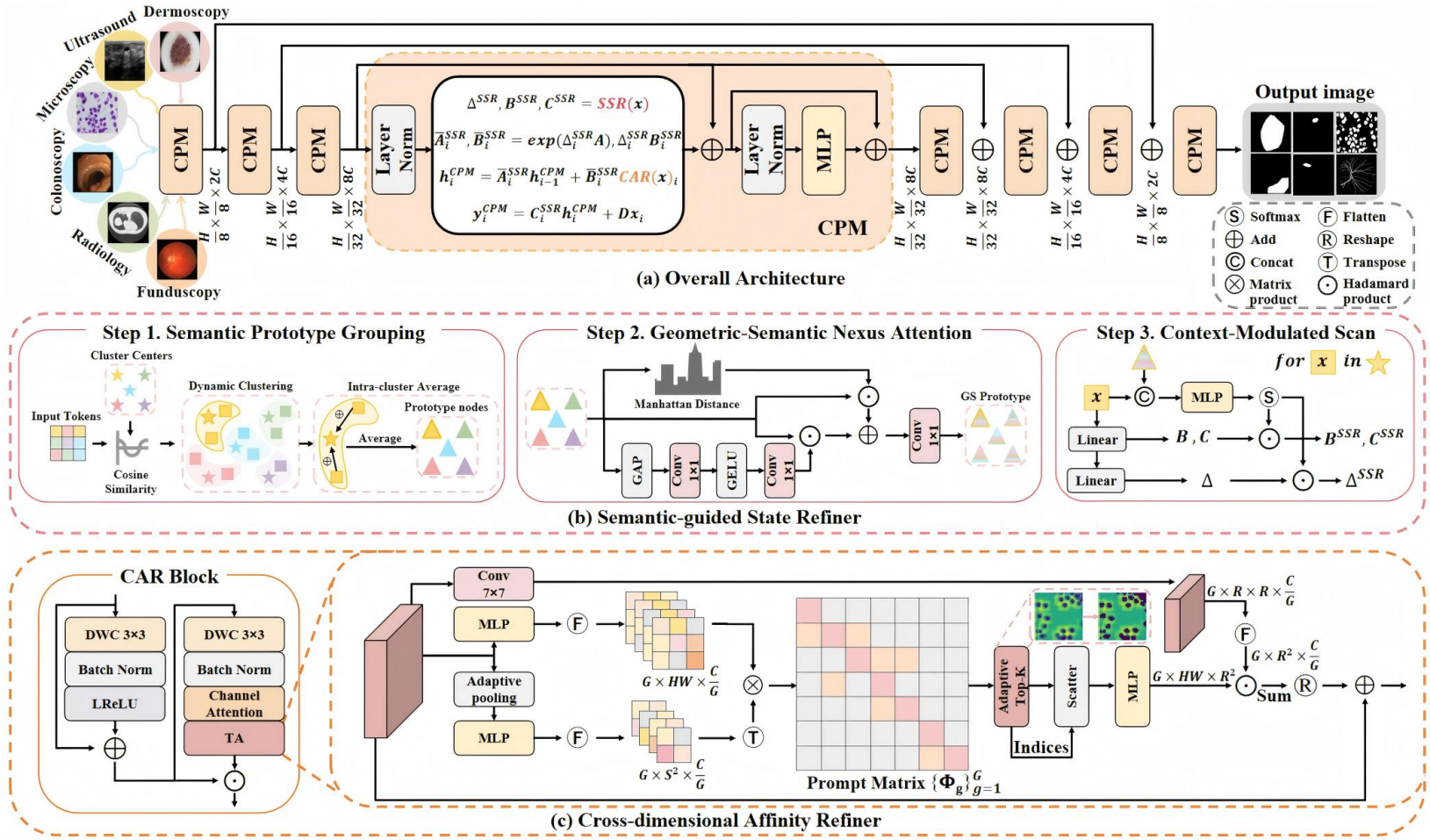


We present GeoSemba, a Mamba-based framework for medical image segmentation.

Reconstructs the state-space update rather than only redesigning scan paths.

Couples geometric-semantic propagation with selective spatial-channel interaction within a single scan.

Overall framework of GeoSemba



$$\Delta^{SSR}, B^{SSR}, C^{SSR} = SSR(x),$$

$$\overline{A}_i^{SSR}, \overline{B}_i^{SSR} = \exp(\Delta_i^{SSR} A), \Delta_i^{SSR} B_i^{SSR}$$

$$h_i^{CPM} = \overline{A}_i^{SSR} h_{i-1}^{CPM} + \overline{B}_i^{SSR} \mathrm{CAR}(x)_i,$$

$$y_i^{CPM} = C_i^{SSR} h_i^{CPM} + Dx_i.$$

Overview of GeoSemba, highlighting the core Cross-Paradigm Module (CPM). CPM consists of two key components: the Semantic-guided State Refiner (SSR) and the Cross-dimensional Affinity Refiner (CAR).

Comparison with State-of-the-art Models

Table 1. Quantitative comparison of GeoSemba with SOTA methods and ablation studies across six diverse medical modalities. The best and second-best results are highlighted in **red** and **blue**, respectively. P-values from one-tailed t-tests are also provided to assess the statistical significance of GeoSemba's improvements over competitors.

Methods	Dermoscopy		Radiology		Ultrasound		Microscopy		Colonoscopy		Fundus		P-value
	ISIC2018 [9]		COVID19-1[19]		BUSI [1]		DSB2018 [3]		ClinicDB [2]		DRIVE [31]		
	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU	
UNet [27]	87.3	80.2	47.7	38.6	69.5	60.2	91.1	84.3	76.5	69.1	81.4	73.2	2.1E-07
TransUNet [5]	87.3	81.2	75.6	68.8	75.5	68.4	91.8	85.2	87.4	82.9	81.5	74.7	1.8E-06
MADGNet [25]	90.2	**83.7**	**83.7**	**76.8**	81.3	73.4	92.0	85.5	93.9	89.5	85.2	78.3	3.2E-03
PVT-EMCAD [26]	**90.9**	83.4	83.6	76.5	**81.8**	**73.5**	**92.7**	85.7	**95.2**	**90.4**	86.3	80.1	4.7E-03
VM-UNet [28]	89.7	81.4	80.0	74.1	79.2	72.1	91.5	84.6	91.9	86.8	84.0	78.2	1.1E-04
SegMamba [37]	90.4	83.2	82.9	75.6	81.3	72.7	92.0	85.1	91.2	86.1	85.6	79.3	5.6E-03
Spatial-Mamba [36]	90.1	82.5	83.3	76.1	81.2	73.0	91.9	84.9	91.8	86.6	85.8	79.6	2.8E-03
LocalMamba [17]	88.4	82.6	82.7	75.4	80.8	72.6	91.9	85.0	91.1	86.5	85.1	78.5	5.3E-05
Mamba-Sea [7]	89.6	82.9	82.9	75.7	80.7	72.5	92.4	**85.9**	92.6	87.3	85.7	79.8	2.8E-04
MambaIRv2 [14]	90.5	83.5	83.0	75.8	81.2	73.2	92.0	85.3	92.1	87.0	**86.5**	80.1	6.1E-03
SCSegamba [22]	90.7	**83.7**	83.3	75.9	81.5	**73.5**	92.5	85.8	92.3	87.2	**86.6**	80.3	1.9E-03
2DMamba [44]	90.2	83.1	82.7	75.6	81.0	73.1	91.3	84.4	91.5	86.7	85.4	79.7	4.4E-03
DefMamba [23]	90.4	83.6	82.8	75.7	81.1	73.3	91.6	84.6	91.7	86.9	85.8	**80.6**	7.2E-03
GeoSemba(w/o CAR)	90.5	82.7	82.6	75.9	81.5	72.3	91.9	85.5	92.7	88.0	85.0	80.3	2.3E-03
GeoSemba(w/o SSR)	90.7	82.9	82.5	75.7	81.8	72.9	92.4	85.7	93.3	88.6	85.4	80.6	3.6E-03
GeoSemba	**91.1**	**84.1**	**83.8**	**77.0**	**82.1**	**73.7**	**93.2**	**86.6**	**94.8**	**90.6**	**86.6**	**81.2**	-

Comparison with State-of-the-art Models

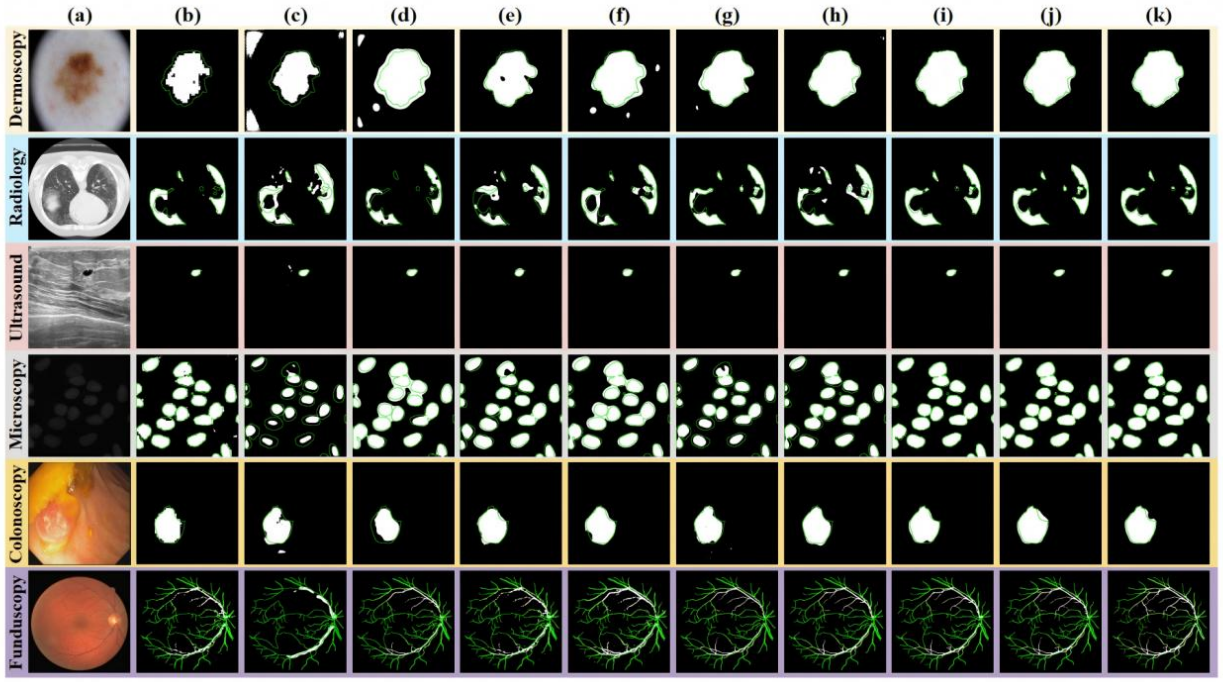


Figure 4. Qualitative comparison of different methods. (a) Input images. (b) UNet [27]. (c) TransUNet [5]. (d) VM-UNet [28]. (e) LocalMamba [17]. (f) MambaIRv2 [14]. (g) SCSegamba [22]. (h) DefMamba [23]. (i) GeoSemba (w/o CAR). (j) GeoSemba (w/o SSR). (k) GeoSemba (ours). Green lines denote the boundaries of the ground truth.

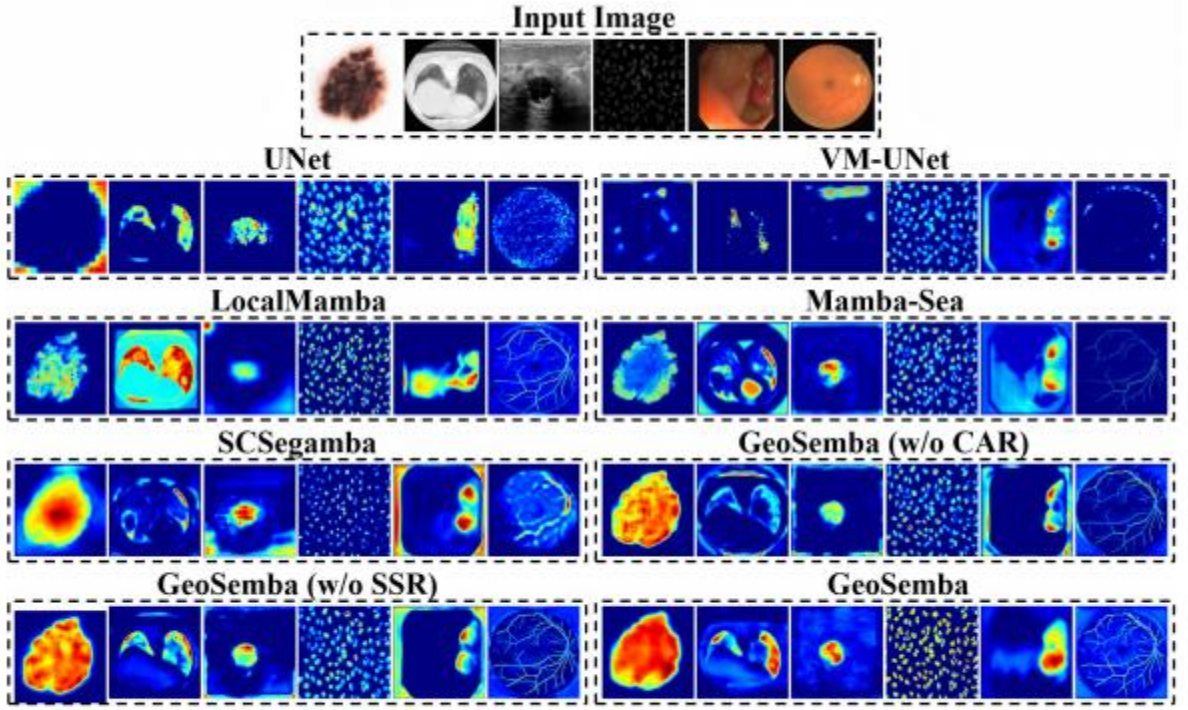


Figure 5. The Grad-CAM visualization of different methods.

Ablation Studies

Table 2. Ablation study on the components of SSR.

Strategy	ISIC2018	COVID19-1	BUSI	DSB2018	ClinicDB	DRIVE
w/o SPG	88.6	81.1	80.2	90.4	91.5	83.7
w/o GSNA	88.3	80.5	79.9	90.0	91.1	83.2
Full	**90.5**	**82.6**	**81.5**	**91.9**	**92.7**	**85.0**

Table 3. Ablation study on the number of prototype nodes in SSR.

M	ISIC2018[9]		COVID-19[19]		BUSI[1]		Params (M)	Infer (ms)
	DSC	IoU	DSC	IoU	DSC	IoU		
3	89.2	82.3	82.1	75.6	80.3	71.8	38.92	14.3
7	**91.1**	**84.1**	**83.8**	**77.0**	**82.1**	**73.7**	41.11	17.0
12	90.1	83.1	83.1	76.3	81.6	73.1	44.67	19.8

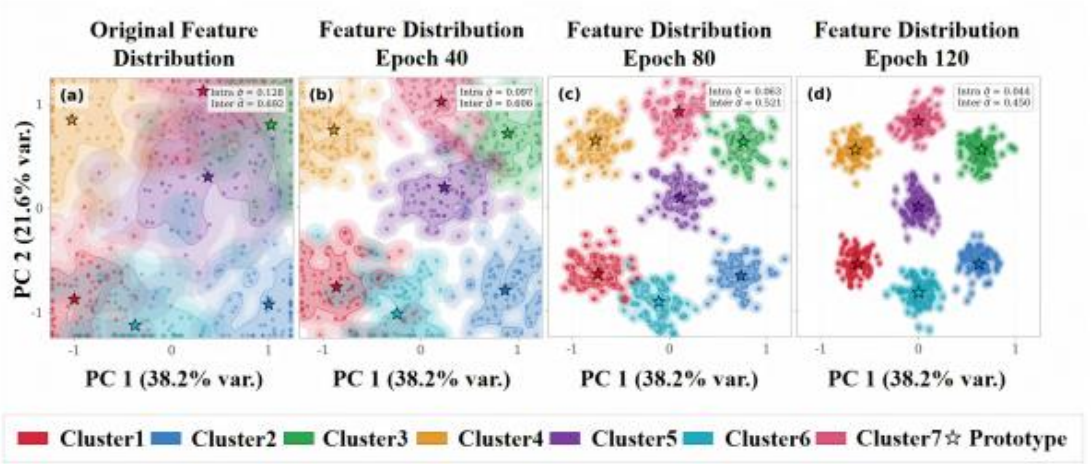


Figure 7. Prototype-guided feature distribution during training.

Ablation Studies

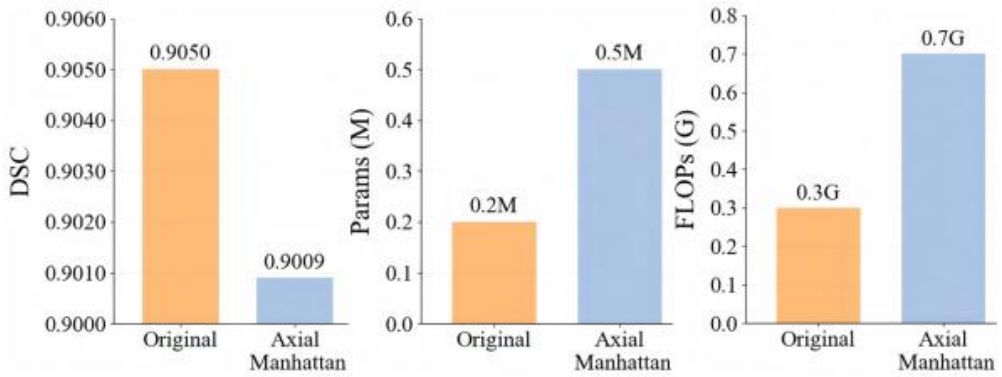


Figure 3. Comparison of GeoSemba with axial decomposition-based methods in terms of DSC, Params, and FLOPs.

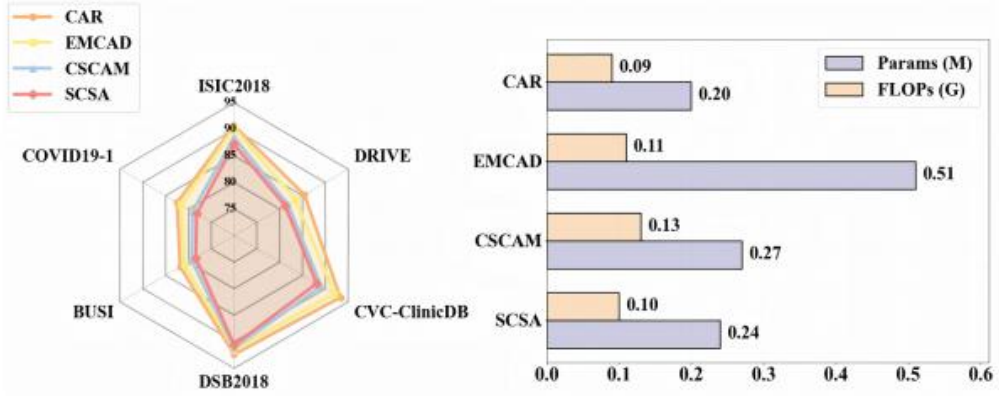


Figure 8. Segmentation performance and computational efficiency of GeoSemba and representative spatial–channel aggregators.

Table 4. Ablation study on the components of CAR.

Strategy	ISIC2018	COVID19-1	BUSI	DSB2018	ClinicDB	DRIVE
w/o MP	88.4	80.6	80.0	90.8	90.8	82.7
w/o CA	88.6	80.9	80.1	91.1	91.0	83.1
w/o Top-K	89.8	81.4	80.4	91.9	92.6	84.7
w/o W^d	90.3	81.8	80.7	92.1	92.8	85.0
Full	**90.7**	**82.5**	**81.8**	**92.4**	**93.3**	**85.4**

Table 5. Ablation study on the number of channel groups in CAR.

G	ISIC2018[9]		COVID-19[19]		BUSI[1]		Params (M)	Infer (ms)
	DSC	IoU	DSC	IoU	DSC	IoU		
6	89.5	82.6	82.7	75.8	81.1	72.4	40.73	15.8
12	**91.1**	**84.1**	**83.8**	**77.0**	**82.1**	**73.7**	41.11	17.0
18	90.2	83.2	83.0	76.3	81.5	73.0	41.58	18.4

Table 6. Ablation study on the Top-K ratio in CAR.

K	ISIC2018[9]		COVID-19[19]		BUSI[1]		Params (M)	Infer (ms)
	DSC	IoU	DSC	IoU	DSC	IoU		
0.25	90.0	83.2	83.1	76.3	81.7	73.0	41.05	16.4
0.50	**91.1**	**84.1**	**83.8**	**77.0**	**82.1**	**73.7**	41.11	17.0
0.75	90.3	83.4	83.4	76.6	81.8	73.3	41.30	18.2

Gradient Flow under Top-K Sparsification

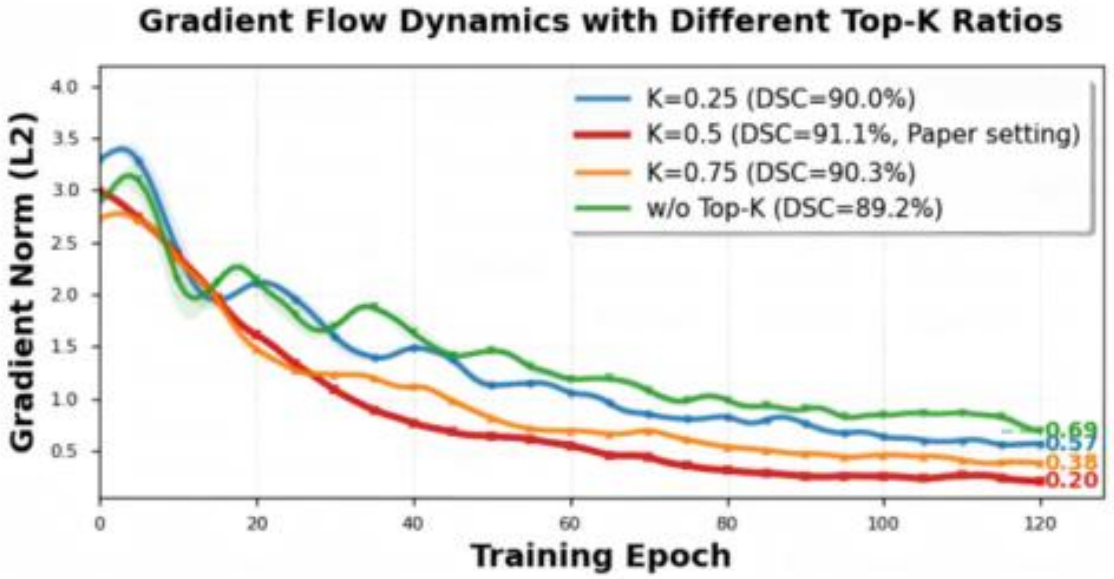


Figure 10. Non-zero gradients under different Top-K settings

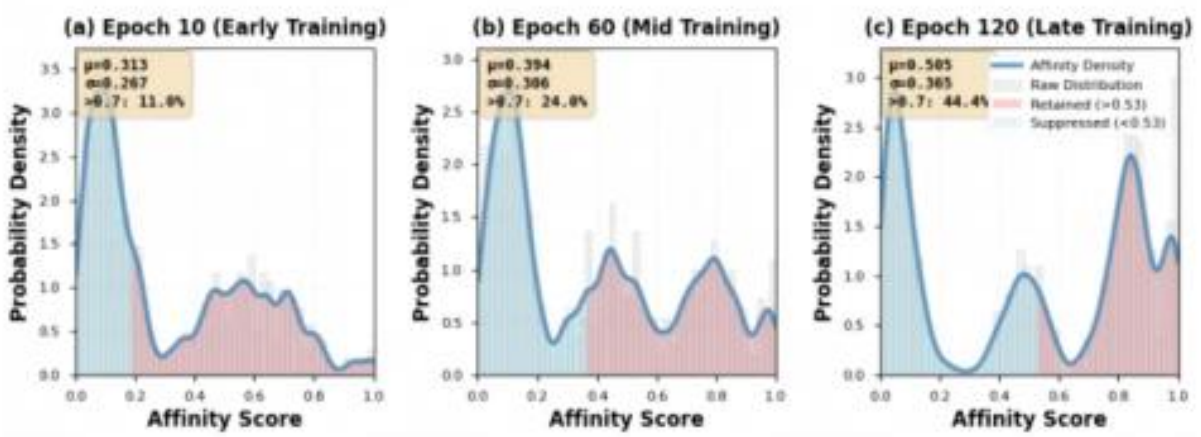


Figure 11. Visualization of affinity patterns at different training stages under Top-K sparsification.

Ablation Studies

Table 8. Ablation study of BCE/Dice loss weighting ratios on representative medical segmentation benchmarks.

Loss Ratio (α:β)	ISIC2018[9]		COVID19-1[19]		BUSI[1]	
	DSC (%)	IoU (%)	DSC (%)	IoU (%)	DSC (%)	IoU (%)
Dice	88.9	81.6	81.2	74.0	80.7	72.1
BCE	89.3	81.9	81.5	74.2	81.0	72.4
1:1	**91.1**	**84.1**	**83.8**	**77.0**	**82.1**	**73.7**
1:2	90.6	83.2	83.0	76.3	81.8	73.3
2:1	90.4	83.1	82.9	75.8	81.5	73.0

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right],$$

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \hat{y}_i^2 + \epsilon},$$

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{Dice}} + \beta \mathcal{L}_{\text{BCE}}.$$

Computational scalability analysis

Table 7. Asymptotic complexity comparison of representative token mixers.

Method	Complexity
Global Self-Attention	$\mathcal{O}(N^2)$
Local / Window Attention (window size w tokens)	$\mathcal{O}(Nw)$
Multi-Directional Mamba (k scans)	$\mathcal{O}(kN)$
CPM (SSR + CAR)	$\mathcal{O}(\mathbf{N})$

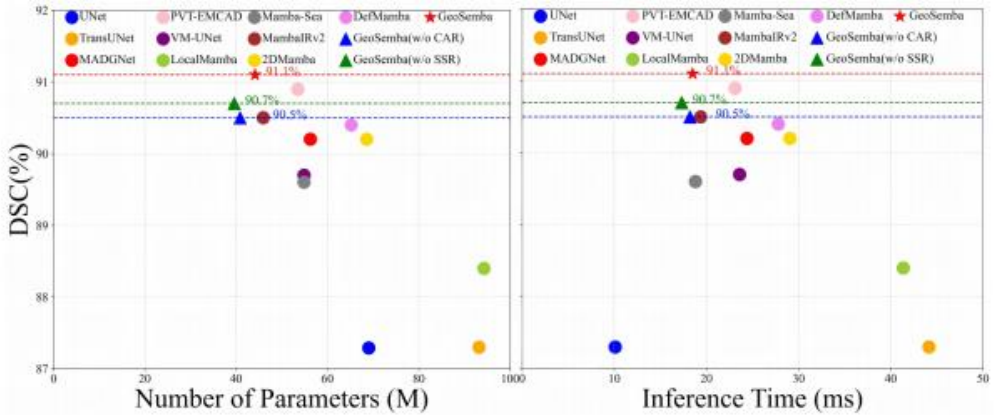


Figure 6. Comparison of parameters and inference speed vs DSC performance for ISIC2018 [9].

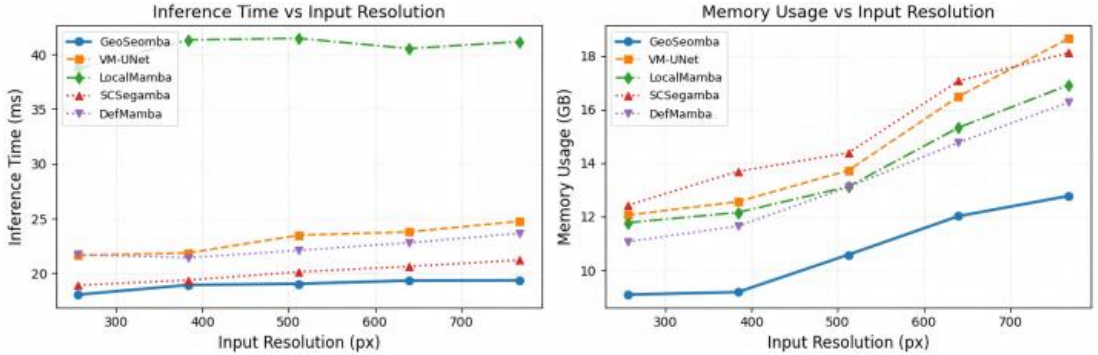


Figure 9. Computational scalability analysis of GeoSemba and representative Mamba variants under different input resolutions.

CVPR JUNE 3-7, 2026 DENVER COLORADO

Thank you !