Non-Stationary Spectral State-Space Networks with Region-Wise Frequency Routing for Robust Fundus Vessel Segmentation

Thanaporn Viriyasaranon and Jang-Hwan Choi(✉)

Department of Artificial Intelligence, Ewha Womans University, Seoul, South Korea
choij@ewha.ac.kr

Abstract. Retinal vessel segmentation and artery/vein (A/V) separation are central to fundus-based vascular biomarker analysis, yet performance often degrades under domain shift when acquisition-dependent appearance overwhelms thin-vessel evidence. We propose NS^{3}Net, a frequency-explicit segmentation framework that couples Mamba-based state-space modeling with region-adaptive spectral processing. Its core NS^{3}MoE performs frequency-plane mixture-of-experts routing to refine high-frequency Fourier magnitudes while preserving phase, enhancing vessel boundaries and thin capillary cues. A low-frequency residual path preserves global appearance to reduce artifact-driven false positives. We further introduce supervised two-view routing-consistency training, which aligns latent routing distributions between original and photometrically perturbed labeled views without unlabeled data or target-domain adaptation. Experiments on six public datasets show consistent ID/OOD gains for vessel segmentation and A/V separation, particularly in boundary-sensitive performance, while maintaining a lightweight profile of 10.03G FLOPs. Our code is available at https://github.com/Thanaporn09/NS3Net.git.

Keywords: Retinal vessel segmentation · Frequency-routed mixture-of-experts · State-space model · Out-of-distribution generalization · Routing-consistency regularization.

1 Introduction

Retinal vessel segmentation is foundational to quantitative vascular biomarker analysis in fundus imaging, supporting clinical workflows such as retinopathy of prematurity (ROP) screening [19] and diabetic retinopathy (DR) assessment [24]. While deep learning models can directly predict vessel and artery/vein (A/V) maps from raw images, their robustness often degrades under real-world variations in contrast, illumination, pathology, and acquisition domain.

Despite rapid advances in medical image segmentation [21, 16, 12, 20], retinal vessel analysis remains challenging due to regional and multi-scale morphological heterogeneity. Thick vessels require topology-consistent tracing, whereas thin capillaries are low-contrast structures that depend on subtle edge cues. Moreover, illumination changes and lesions can introduce vessel-like artifacts. As a result, uniform feature transformations may miss fine capillaries or amplify false positives, ultimately degrading vessel segmentation and A/V separation.

To address this, we propose NS^{3}Net, a frequency-explicit framework for robust fundus vessel segmentation and A/V separation. NS^{3}Net is built from NS^{3}Blocks that combine state-space modeling with NS^{3}MoE, a region-wise frequency-routed module that decomposes features into spectral components and assigns frequency-plane regions to specialized experts. It refines high-frequency vessel-edge and capillary cues while preserving low-frequency appearance through a residual path, targeting boundary-sensitive failure cases under domain shift. By refining Fourier magnitudes with preserved phase, NS^{3}MoE decouples fine vessel evidence from acquisition-dependent appearance, reducing artifact-driven false positives without heavy computational overhead. Coupled with a state-space backbone, NS^{3}Net further captures long-range context along vessel trees.

The limited availability of annotated fundus data, especially for A/V segmentation compared with large-scale labeled datasets in other modalities such as chest X-ray [5], remains a critical challenge. In this low-data regime, high-capacity models may overfit to dataset-specific appearance cues rather than anatomical structure, degrading vessel connectivity and OOD performance. Generic spatial regularization, such as dropout or aggressive augmentation, is often insufficient because strong perturbations can corrupt the high-frequency signals that govern capillary boundaries and continuity. To improve generalization without unlabeled data, pseudo-labeling, or target-domain adaptation, we introduce supervised two-view routing consistency (RC), which aligns latent expert-routing distributions between an image and a photometrically perturbed view of the same labeled image.

The main contributions of this paper are as follows:

- We propose NS^{3}Net, a lightweight frequency-explicit segmentation network that integrates NS^{3}MoE and gated residual fusion to refine high-frequency vessel-edge cues, preserve low-frequency appearance, and improve robust capillary delineation under domain shift.
- We introduce supervised two-view routing-consistency training to regularize latent expert-routing distributions between original and photometrically perturbed labeled views.
- We evaluate NS^{3}Net on six public datasets (FIVES [11], CHASEDB1 [18], DRIVE [23], HRF [1], Fundus-AVSeg [4], and LES-AV [17]), showing consistent ID/OOD gains for vessel segmentation and A/V separation, particularly in boundary-sensitive metrics, while preserving an efficient accuracy–efficiency trade-off.

2 Methods

In this section, we present our fundus vessel segmentation framework, comprising the proposed Non-Stationary Spectral State-Space Network (NS^{3}Net) and a two-view routing-consistency training strategy designed to improve out-of-distribution (OOD) generalization.

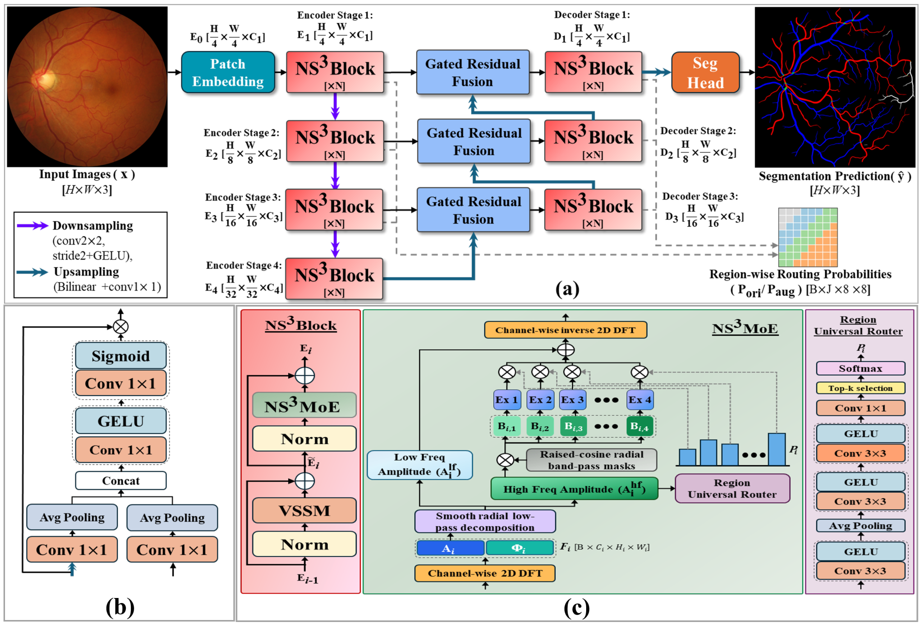

Fig. 1. NS^{3}Net architecture and components. (a) NS^{3}Net architecture. (b) Gated residual fusion. (c) NS^{3}Block with DFT-based NS^{3}MoE.

2.1 NS^{3}Net Architecture

NS^{3}Net is a hierarchical encoder–decoder architecture (Fig. 1(a)) designed to separate vessel anatomy from domain-specific imaging artifacts, with NS^{3}MoE as the core module of each NS^{3}Block. Given an RGB fundus image $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$, a 4×4 convolutional patch embedding with stride 4 produces $\mathbf{E}_0 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C_0}$. The encoder processes $\mathbf{E}_0$ through four NS^{3}Block stages with channel dimensions $(48, 96, 192, 384)$ and $\times 2$ downsampling, while the three-stage decoder progressively restores spatial resolution. At each decoder stage, the upsampled decoder feature and corresponding encoder skip feature are fused by the Gated Residual Fusion module (Fig. 1(b)) for adaptive skip-feature integration. The module projects both inputs to the decoder dimensionality, computes a channel-wise sigmoid gate from globally pooled descriptors, and residually adds the gated skip feature to the projected decoder feature. The fused feature is passed to the current decoder stage. Finally, a lightweight refinement head upsamples the last decoder feature by a factor of 4 and applies a 1×1 convolution to produce the segmentation logits.

2.2 NS^{3}Block

The NS^{3}Block is the core computational unit of NS^{3}Net, designed to preserve global vascular continuity while remaining sensitive to thin capillaries (Fig. 1(c)).

Each NS^{3}Block consists of two components: (1) a Vision Mamba-based state-space module (VSSM) [6, 15, 14] for efficient global dependency modeling, and (2) the proposed NS^{3}MoE, a frequency-domain feed-forward module that performs region-wise routing on the Fourier plane. Specifically, NS^{3}MoE selectively refines high-frequency amplitude responses through frequency-region routing and radial-band expert processing as follows:

$$\begin{aligned}\tilde{\mathbf{E}}_i &\leftarrow \mathbf{E}_{i-1} + \mathrm{VSSM}(\mathrm{Norm}(\mathbf{E}_{i-1}))\,,\\ \mathbf{E}_i &\leftarrow \tilde{\mathbf{E}}_i + \mathrm{NS^3MoE}\Big(\mathrm{Norm}(\tilde{\mathbf{E}}_i)\Big)\,.\end{aligned} \tag{1}$$

where $\mathbf{E}_{i-1}$ denotes the input feature to the i-th NS^{3}Block, $\tilde{\mathbf{E}}_i$ is the intermediate feature after the state-space module, $\mathrm{Norm}(\cdot)$ denotes a normalization layer, and $\mathbf{E}_i$ is the output feature of the block.

NS^{3}MoE: Region-Wise Frequency-Routed FFN Building on the MoE routing paradigm [22], NS^{3}MoE operates on the normalized intermediate feature from the state-space module. Given the intermediate feature $\tilde{\mathbf{E}}_i \in \mathbb{R}^{B\times C_i\times H_i\times W_i}$ from the NS^{3}Block, we define $\mathbf{U}_i = \mathrm{Norm}(\tilde{\mathbf{E}}_i)$ as the input to NS^{3}MoE. We first transform each channel into the frequency domain using a channel-wise 2D Discrete Fourier Transform (DFT):

$$\mathbf{F}_i = \mathcal{F}_{2D}(\mathbf{U}_i), \tag{2}$$

where $\mathbf{F}_i \in \mathbb{C}^{B\times C_i\times H_i\times W_i}$ denotes the complex-valued frequency coefficients. We represent $\mathbf{F}_i$ by its amplitude and phase:

$$\mathbf{A}_i = |\mathbf{F}_i|, \qquad \mathbf{\Phi}_i = \angle \mathbf{F}_i, \tag{3}$$

where $\mathbf{A}_i, \mathbf{\Phi}_i \in \mathbb{R}^{B\times C_i\times H_i\times W_i}$. A fixed smooth radial low-pass mask $\mathbf{M}_{\mathrm{lp}} \in [0,1]^{1\times 1\times H_i\times W_i}$ is then used to decompose the amplitude spectrum into low- and high-frequency components:

$$\mathbf{A}_i^{\mathrm{lf}} = \mathbf{A}_i \odot \mathbf{M}_{\mathrm{lp}}, \qquad \mathbf{A}_i^{\mathrm{hf}} = \mathbf{A}_i \odot (1 - \mathbf{M}_{\mathrm{lp}})\,, \tag{4}$$

where $\odot$ denotes element-wise multiplication. The mask is implemented as a raised-cosine transition with normalized cutoff 0.25 and width 0.10.

NS^{3}MoE performs region-wise expert routing on $\mathbf{A}_i^{\mathrm{hf}}$, where each routing cell corresponds to a local area of the high-frequency plane. The Region Universal Router pools $\mathbf{A}_i^{\mathrm{hf}}$ into an $h_g \times w_g = 8 \times 8$ frequency grid and predicts J expert logits per cell:

$$\mathbf{Z}_i = h_1\big(g_3\big(g_1\big(\mathrm{AvgPool}_{8\times 8}\big(g_1(\mathbf{A}_i^{\mathrm{hf}})\big)\big)\big)\big)\,, \tag{5}$$

where $\mathbf{Z}_i \in \mathbb{R}^{B\times J\times 8\times 8}$, $g_1(\mathbf{x}) = \mathrm{GELU}(\mathrm{Conv}_{1\times 1}(\mathbf{x}))$, $g_3(\mathbf{x}) = \mathrm{GELU}(\mathrm{Conv}_{3\times 3}(\mathbf{x}))$, and $h_1(\mathbf{x}) = \mathrm{Conv}_{1\times 1}(\mathbf{x})$. Top-$k$ masking followed by softmax over the expert dimension gives the region-wise routing probabilities:

$$\mathbf{P}_i = \mathrm{Softmax}_J(\mathrm{TopKMask}(\mathbf{Z}_i, k))\,, \tag{6}$$

where $k = 2$, $\mathbf{P}_i \in \mathbb{R}^{B\times J\times 8\times 8}$, and $J = 4$ in all experiments.

To encourage expert specialization across radial frequency ranges, we partition the high-frequency range $\rho \in [0.25, 1]$ into J equal-width bands with edges $a_j = 0.25 + \frac{j-1}{J}(1 - 0.25)$ and $b_j = 0.25 + \frac{j}{J}(1 - 0.25)$. For each expert j, a raised-cosine radial band-pass mask $\mathbf{M}_{\text{bp}}^{(j)}$ smoothly decays to zero outside $[a_j - 0.10,\, b_j + 0.10]$. The corresponding band-isolated response is obtained by

$$\mathbf{B}_i^{(j)} = \mathbf{A}_i^{\text{hf}} \odot \mathbf{M}_{\text{bp}}^{(j)}, \qquad j = 1, \dots, J, \tag{7}$$

where $\mathbf{B}_i^{(j)} \in \mathbb{R}^{B\times C_i\times H_i\times W_i}$. Each band response is refined by a lightweight spectral expert implemented as follows:

$$\tilde{\mathbf{B}}_i^{(j)} = \mathbf{B}_i^{(j)} + \text{GELU}\Big(\text{Conv}_{1\times1,2}^{(j)}\Big(\text{GELU}\Big(\text{Conv}_{1\times1,1}^{(j)}(\mathbf{B}_i^{(j)})\Big)\Big)\Big). \tag{8}$$

The routing probabilities are bilinearly upsampled to the frequency resolution and used to fuse the expert-refined band responses:

$$\hat{\mathbf{A}}_i^{\text{hf}} = \sum_{j=1}^{J} \text{Up}\Big(\mathbf{P}_i^{(j)}\Big) \odot \tilde{\mathbf{B}}_i^{(j)}, \tag{9}$$

where $\mathbf{P}_i^{(j)}$ denotes the j-th expert probability map and $\text{Up}(\mathbf{P}_i^{(j)}) \in \mathbb{R}^{B\times 1\times H_i\times W_i}$ is broadcast along the channel dimension. The refined amplitude is reconstructed by recombining the low-frequency amplitude and the refined high-frequency amplitude:

$$\mathbf{A}_i^{\text{rec}} = \mathbf{A}_i^{\text{lf}} + \hat{\mathbf{A}}_i^{\text{hf}}. \tag{10}$$

The complex spectrum is then reconstructed using the preserved phase and transformed back to the spatial domain:

$$\mathbf{F}_i^{\text{rec}} = \mathbf{A}_i^{\text{rec}} \odot \exp(\mathrm{i}\boldsymbol{\Phi}_i), \qquad \Delta\mathbf{E}_i = \mathcal{F}_{2D}^{-1}(\mathbf{F}_i^{\text{rec}}). \tag{11}$$

NS[3]MoE outputs the frequency-aware residual $\Delta\mathbf{E}_i$, which is added to $\tilde{\mathbf{E}}_i$ to produce the block output.

2.3 Two-View Routing-Consistency Training

To improve robustness to acquisition variability, we regularize the expert-routing behavior of NS[3]MoE under two photometric views of the same labeled image, as shown in Fig. 2. Given an input image $\mathbf{x}$ and label $\mathbf{y}$, we generate an augmented view $\mathbf{x}^{\text{aug}}$ using stochastic photometric perturbations. We restrict the perturbations to photometric transformations so that vessel geometry and routing-grid correspondence remain unchanged when computing routing consistency. Both views are processed by the same NS[3]Net, denoted by $f_\theta(\cdot)$, with shared parameters:

$$(\hat{\mathbf{y}},\, \{\mathbf{P}_{i,b}^{\text{ori}}\}) = f_\theta(\mathbf{x}), \qquad (\hat{\mathbf{y}}^{\text{aug}},\, \{\mathbf{P}_{i,b}^{\text{aug}}\}) = f_\theta(\mathbf{x}^{\text{aug}}), \tag{12}$$

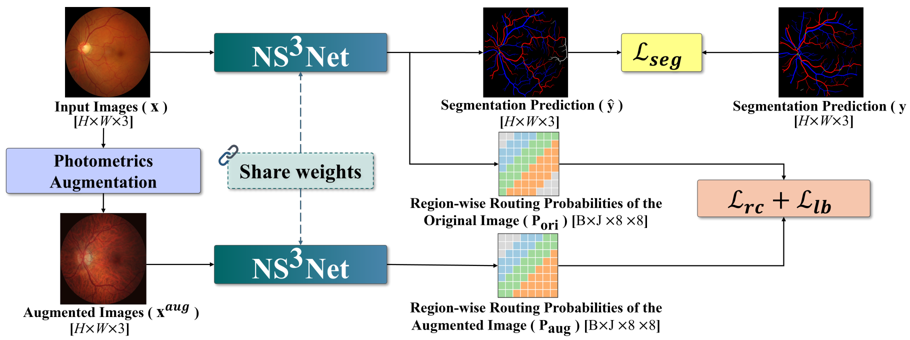

Fig. 2. Supervised two-view routing-consistency training.

where $\{\mathbf{P}_{i,b}^{\text{ori}}\}$ and $\{\mathbf{P}_{i,b}^{\text{aug}}\}$ denote the collections of region-wise routing probability tensors extracted from the NS^{3}MoE blocks for the original and augmented views, respectively. Specifically, $\mathbf{P}_{i,b}$ denotes the routing tensor from the b-th block at stage i.

Unlike semi-supervised prediction consistency, routing consistency uses only labeled images and regularizes latent expert routing without unlabeled data, pseudo-labels, or target-domain adaptation. To prevent both views from drifting together, we treat the original-view routing as a fixed reference by stopping gradients through $\mathbf{P}_{i,b}^{\text{ori}}$ and minimize the KL divergence from the augmented-view routing to this reference:

$$\mathcal{L}_{\text{rc}} = \frac{1}{|\mathcal{I}|\,|\Omega_g|} \sum_{(i,b)\in\mathcal{I}} \sum_{u\in\Omega_g} \sum_{j=1}^{J} \mathbf{P}_{i,b}^{\text{aug},(j)}(u) \log \frac{\mathbf{P}_{i,b}^{\text{aug},(j)}(u)}{\text{stopgrad}\Big(\mathbf{P}_{i,b}^{\text{ori},(j)}(u)\Big)}, \tag{13}$$

where $\mathcal{I}$ is the set of NS^{3}MoE blocks and $\Omega_g = \{1,\ldots,h_g\}\times\{1,\ldots,w_g\}$ denotes the frequency-region routing grid, with each $u \in \Omega_g$ indexing one routing cell. This direction encourages the routing distribution of the photometrically perturbed view to match the original-view reference, stabilizing expert utilization under acquisition shifts.

The training objective combines supervised segmentation with routing regularization:

$$\mathcal{L} = \mathcal{L}_{\text{seg}}(\hat{\mathbf{y}}, \mathbf{y}) + \lambda_{\text{rc}}\mathcal{L}_{\text{rc}} + \lambda_{\text{lb}}\mathcal{L}_{\text{lb}}, \qquad \lambda_{\text{rc}} = 0.1, \quad \lambda_{\text{lb}} = 0.01, \tag{14}$$

where $\mathcal{L}_{\text{seg}}$ is a combined Dice and cross-entropy loss with equal weight, and $\mathcal{L}_{\text{lb}}$ is a Switch-style load-balancing loss.

3 Experiments

3.1 Dataset

We evaluate two fundus segmentation tasks: binary retinal vessel segmentation and artery–vein (A/V) segmentation.

Table 1. Performance comparison on the binary retinal vessel segmentation task across in-distribution (FIVES) and out-of-distribution (CHASEDB1, DRIVE, HRF) test sets.

Method	ID Test			OOD Test								
	FIVES			CHASEDB1			DRIVE			HRF		
	DSC↑	IoU↑	HD95↓	DSC↑	IoU↑	HD95↓	DSC↑	IoU↑	HD95↓	DSC↑	IoU↑	HD95↓
nnUNet [10]	92.35±0.07	86.71±0.12	33.93±0.58	77.30±0.21	63.09±0.28	31.63±1.16	69.23±0.37	53.06±0.44	46.09±0.36	80.44±0.03	67.62±0.04	29.65±0.48
TransUNet [3]	88.25±2.19	80.19±3.42	33.72±3.20	74.14±1.27	59.02±1.60	48.80±6.84	62.65±2.83	45.25±2.16	53.59±5.09	78.31±0.94	64.72±1.27	37.80±3.91
Swin-UNet [2]	90.69±0.04	83.84±0.05	35.13±0.89	76.43±0.04	61.92±0.05	36.79±0.14	70.19±0.21	54.16±0.25	35.65±0.36	79.35±0.04	66.13±0.05	32.27±0.29
UNETR [8]	89.32±1.24	81.85±1.81	46.78±8.01	76.21±0.56	61.65±0.73	36.17±5.34	67.93±2.41	51.57±2.72	41.77±7.02	79.83±0.53	66.77±0.72	**29.17±1.08**
MISSFormer [9]	92.16±0.07	86.35±0.07	32.26±0.81	77.11±0.06	62.85±0.08	31.15±0.84	71.47±0.31	55.74±0.36	35.96±1.33	80.28±0.11	67.40±0.14	31.67±0.17
T-Mamba [7]	89.85±0.02	82.79±0.03	43.95±1.72	75.89±0.16	61.23±0.21	41.16±0.75	66.19±0.08	49.58±0.09	45.11±0.15	78.89±0.01	65.52±0.03	33.45±0.53
LightM-UNet [13]	87.96±0.18	79.57±0.26	44.03±0.50	75.36±0.11	60.55±0.14	39.54±0.94	67.28±1.46	51.56±0.36	43.23±1.37	79.27±0.16	64.31±3.16	31.29±0.58
NS^{3}Net	**92.56±0.02**	**86.97±0.01**	**32.03±0.19**	**77.43±0.15**	**63.28±0.20**	**30.29±1.37**	**72.60±0.08**	**56.29±0.16**	**33.51±0.48**	**80.68±0.09**	**68.04±0.03**	30.14±0.03

Retinal vessel segmentation. We use FIVEs [11] with the official split (600 development, 200 in-distribution (ID) test, 2048×2047). The development set is split into 480/120 for training/validation. For out-of-distribution (OOD) evaluation, we directly test the trained model on CHASEDB1 [18] (28, 1280×960), DRIVE [23] (40, 768×584), and HRF [1] (45, 3504×2336), using all images in each dataset.

A/V vessel segmentation. We use Fundus-AVSeg [4] (100 images; 2656×1992 or 1280×1280) with the official split (80 development, 20 ID test). For OOD generalization, we evaluate on LES-AV [17] (20 images; 1444×1620 or 1958×2196).

3.2 Implementation details

All models were implemented in PyTorch and trained from scratch for 300 epochs on an NVIDIA A100. NS^{3}Net used AdamW with lr 4×10^{-4}, and baseline learning rates were validation-tuned. All datasets used channel-wise z-score normalization, random 512×512 patches, batch size 8, and full-resolution sliding-window inference. We report mean $\pm$ std over three seeds for binary vessel segmentation and five fold-trained models for A/V segmentation. Metrics include DSC, IoU, and HD95. Each encoder/decoder stage uses one NS^{3}Block, with DFT/inverse DFT implemented by rFFT/irFFT for real-valued features.

3.3 Performance evaluation

We compare NS^{3}Net with prior methods and ablate each component. Best and second-best results are shown in **red bold** and blue underline, respectively.

Comparisons with representative baselines Tables 1 and 2 compare NS^{3}Net (4.68M parameters, 10.03G FLOPs) with representative CNN-, Transformer-, and Mamba-based baselines. On binary vessel segmentation (Table 1), NS^{3}Net achieves the best ID performance on FIVES (DSC/IoU/HD95: 92.56/86.97/32.03) and improves cross-dataset generalization on CHASEDB1, DRIVE, and HRF, with lower boundary error on CHASEDB1 and DRIVE (HD95: 30.29 and 33.51) and the highest DSC/IoU on both. For multiclass A/V segmentation (Table 2), NS^{3}Net attains the best in-distribution results on Fundus-AVSeg (DSC/IoU/HD95:

Table 2. Performance comparison on multiclass artery–vein (A/V) segmentation on Fundus-AVSeg (in-distribution) and LES-AV (out-of-distribution). clDSC denotes Dice on vessel centerlines (1-pixel skeletons).

Method	Params (M)	FLOPs (G)	Fundus-AVSeg (ID Test)				LES-AV (OOD Test)			
			DSC↑	IoU↑	HD95↓	clDSC↑	DSC↑	IoU↑	HD95↓	clDSC↑
nnUNet [10]	46.33	60.14	57.28±0.42	46.76±0.10	216.83±70.70	77.79±0.07	48.17±0.61	36.35±0.52	**268.85±52.64**	69.42±0.10
TransUNet [3]	109.15	42.26	60.47±0.52	51.04±0.43	238.06±16.47	81.75±0.07	50.36±0.24	38.56±0.23	311.47±12.83	71.51±0.12
Swin-UNet [2]	27.18	30.92	59.65±0.75	49.17±0.67	248.49±23.33	82.25±0.09	46.15±0.69	34.34±0.71	340.54±24.11	69.39±0.12
UNETR [8]	88.30	120.11	57.06±0.48	46.45±0.47	265.23±24.52	82.05±0.08	47.93±0.40	35.90±0.37	322.97±20.47	72.83±0.10
MISSFormer [9]	20.04	31.30	63.75±0.45	54.50±0.45	223.74±11.45	87.32±0.07	51.00±0.56	39.36±0.43	306.63±31.58	71.19±0.09
T-Mamba [7]	0.76	3.09	60.22±0.52	50.28±0.49	230.70±8.65	81.46±0.08	47.41±0.72	35.35±1.03	317.19±8.88	63.90±0.14
LightM-UNet [13]	6.01	31.54	55.07±0.38	44.57±0.35	276.23±4.54	77.31±0.11	40.89±0.66	30.08±0.51	342.75±23.37	56.92±0.13
NS^{3}Net	4.68	10.03	**64.92±0.38**	**55.99±0.14**	**193.16±21.14**	**88.21±0.07**	**51.72±0.24**	**39.63±0.06**	291.14±0.40	**74.20±0.10**

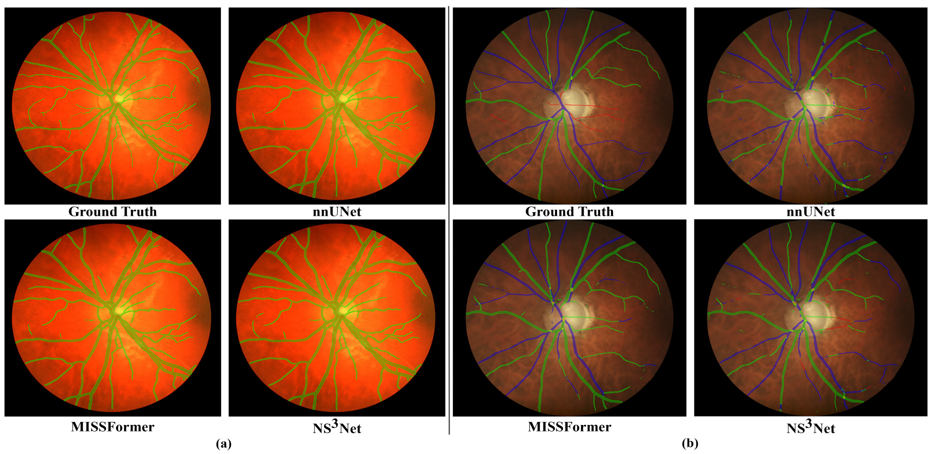

Fig. 3. Comparison of NS^{3}Net with prior methods on OOD test sets: (a) CHASEDB1 (retinal vessel segmentation) and (b) LES-AV (artery/vein segmentation).

64.92/55.99/193.16) and remains robust under domain shift on LES-AV (DSC/IoU: 51.72/39.63; HD95: 291.14). Additionally, qualitative segmentation comparisons are shown in Fig. 3.

Ablation study We ablate NS^{3}MoE and routing consistency (RC) on binary vessel segmentation. As shown in Table 3, replacing the MLP FFN in the VSSM baseline with NS^{3}MoE improves ID performance and all OOD results. Adding RC further enhances OOD robustness by stabilizing expert routing under photometric shifts, reducing HD95 by 3.30 pixels on DRIVE and 3.05 pixels on CHASEDB1 compared with the MoE-only variant.

4 Conclusion

We presented NS^{3}Net, a lightweight retinal vessel segmentation framework that combines state-space modeling, NS^{3}MoE, gated residual fusion, and supervised

Table 3. Ablation on binary vessel segmentation: the VSSM baseline uses an MLP FFN without routing-consistency (RC).

Components		ID (FIVES)			OOD								
					CHASEDB1			DRIVE			HRF		
NS^{3}MoE	RC	DSC↑	IoU↑	HD95↓	DSC↑	IoU↑	HD95↓	DSC↑	IoU↑	HD95↓	DSC↑	IoU↑	HD95↓
–	–	92.27±0.02	86.23±0.56	33.80±0.47	76.50±0.41	61.94±0.62	35.07±0.67	71.25±0.10	55.46±0.12	38.54±0.54	80.22±0.04	67.32±0.06	31.60±0.31
✓	–	92.46±0.03	86.62±0.39	33.46±1.66	77.07±0.05	62.71±0.48	33.34±0.34	72.05±0.03	55.93±0.02	36.81±0.23	80.27±0.03	67.33±0.17	30.86±0.06
✓	✓	**92.56±0.02**	**86.97±0.01**	**32.03±0.19**	**77.43±0.15**	**63.28±0.20**	**30.29±1.37**	**72.60±0.08**	**56.29±0.16**	**33.51±0.48**	**80.68±0.09**	**68.04±0.03**	**30.14±0.03**

two-view routing consistency for robust vessel and A/V segmentation. By refining high-frequency vessel-edge cues while preserving low-frequency appearance and phase, NS^{3}Net improves boundary-sensitive ID/OOD performance without unlabeled data, target-domain adaptation, or substantial computational overhead. Future work will explore adaptive frequency partitioning and broader routing consistency under acquisition shifts.

Acknowledgements This work was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (Nos. RS-2025-02215813, RS-2025-00520578, and RS-2025-02634603) and by an NRF grant funded by the Korean government (MSIT and MOE) (No. RS-2025-16063688).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Budai, A., Bock, R., Maier, A., Hornegger, J., Michelson, G.: Robust vessel segmentation in fundus images. International journal of biomedical imaging **2013**(1), 154860 (2013)
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Proceedings of the European Conference on Computer Vision Workshops(ECCVW) (2022)
3. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis p. 103280 (2024)
4. Deng, Z., Gao, W., Gong, Z., Gan, R., Chen, L., Zhang, S., Ma, L.: A fundus image dataset for ai-based artery-vein vessel segmentation. Scientific Data **12**(1), 1298 (2025)
5. Gaggion, N., Mosquera, C., Mansilla, L., Saidman, J.M., Aineseder, M., Milone, D.H., Ferrante, E.: Chexmask: a large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images. Scientific Data **11**(1), 511 (2024)
6. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. In: International Conference on Learning Representations (ICLR) (2022)
7. Hao, J., Zhu, Y., He, L., Liu, M., Tsoi, J.K.H., Hung, K.F.: T-mamba: a unified framework with long-range dependency in dual-domain for 2d & 3d tooth segmentation. IEEE Transactions on Multimedia (2026)

8. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
9. Huang, X., Deng, Z., Li, D., Yuan, X., Fu, Y.: Missformer: An effective transformer for 2d medical image segmentation. IEEE transactions on medical imaging **42**(5), 1484–1494 (2022)
10. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
11. Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., Zhang, Q., Wang, Y., Ye, J.: Fives: A fundus image dataset for artificial intelligence based vessel segmentation. Scientific data **9**(1), 475 (2022)
12. Kamran, S.A., Hossain, K.F., Tavakkoli, A., Zuckerbrod, S.L., Sanders, K.M., Baker, S.A.: Rv-gan: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In: international conference on medical image computing and computer-assisted intervention. pp. 34–44. Springer (2021)
13. Liao, W., Zhu, Y., Wang, X., Pan, C., Wang, Y., Ma, L.: Lightm-unet: Mamba assists in lightweight unet for medical image segmentation. arXiv preprint arXiv:2403.05246 (2024)
14. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. Advances in neural information processing systems **37**, 103031–103063 (2024)
15. Nguyen, E., Goel, K., Gu, A., Downs, G., Shah, P., Dao, T., Baccus, S., Ré, C.: S4nd: Modeling images and videos as multidimensional signals with state spaces. Advances in neural information processing systems **35**, 2846–2861 (2022)
16. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
17. Orlando, J.I., Barbosa Breda, J., Van Keer, K., Blaschko, M.B., Blanco, P.J., Bulant, C.A.: Towards a glaucoma risk index based on simulated hemodynamics from fundus images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 65–73. Springer (2018)
18. Owen, C.G., Rudnicka, A.R., Mullen, R., Barman, S.A., Monekosso, D., Whincup, P.H., Ng, J., Paterson, C.: Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. Investigative ophthalmology & visual science **50**(5), 2004–2010 (2009)
19. Poletti, E., Fiorin, D., Grisan, E., Ruggeri, A.: Automatic vessel segmentation in wide-field retina images of infants with retinopathy of prematurity. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 3954–3957. IEEE (2011)
20. Qi, Y., He, Y., Qi, X., Zhang, Y., Yang, G.: Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6070–6079 (October 2023)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

22. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538 (2017)
23. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE transactions on medical imaging **23**(4), 501–509 (2004)
24. Ting, D.S.W., Cheung, G.C.M., Wong, T.Y.: Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. Clinical & experimental ophthalmology **44**(4), 260–277 (2016)