A Flexible Structure-Guided Feature Aggregation Paradigm for Efficient Vascular Representation

Yaolei Qi[1], Wenbo Peng[1], Tong Wang[1], Chenwei Xu[1], Yuan Zhang[1], and Guanyu Yang[1,2(✉)]

[1] Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China
[2]Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, China
yang.list@seu.edu.cn

Abstract. Accurate representation of vascular structures is fundamental for a wide range of clinical applications and remains challenging due to the slender geometry, complex branching topology, and long-range continuity inherent to tree-like vasculature. Although structural priors have been widely explored in vascular image analysis, existing approaches typically integrate such priors implicitly or isotropically during feature aggregation, limiting their ability to faithfully capture directional and structural dependencies along the vessels. In this study, we propose a flexible, structure-guided feature aggregation paradigm that explicitly incorporates anatomical structural cues into the feature aggregation process for efficient vascular representation. The core idea is to aggregate features along directionally guided paths that follow local vascular structures, enabling consistent propagation of structural information while maintaining computational efficiency. To avoid indiscriminate prior enforcement, the proposed paradigm further introduces an adaptive mechanism that selectively injects structural guidance where representation is most challenging. This aggregation paradigm is lightweight, modular, and can be seamlessly integrated into multi-stage feature hierarchies without altering the backbone architecture. Extensive experiments on vascular imaging tasks show that our approach consistently improves the quality of vascular representation. Our code is available at: https://github.com/YaoleiQi/VSP.

Keywords: Feature aggregation · Vessel segmentation · Topology-aware

1 Introduction

The accurate and structurally faithful representation of vascular anatomy constitutes a fundamental problem in medical image analysis [1,2]. Across a broad spectrum of vascular imaging tasks, the central challenge lies in modeling a highly organized tree-structured anatomical system while preserving its intrinsic continuity and topological properties [1]. Vascular networks are not merely

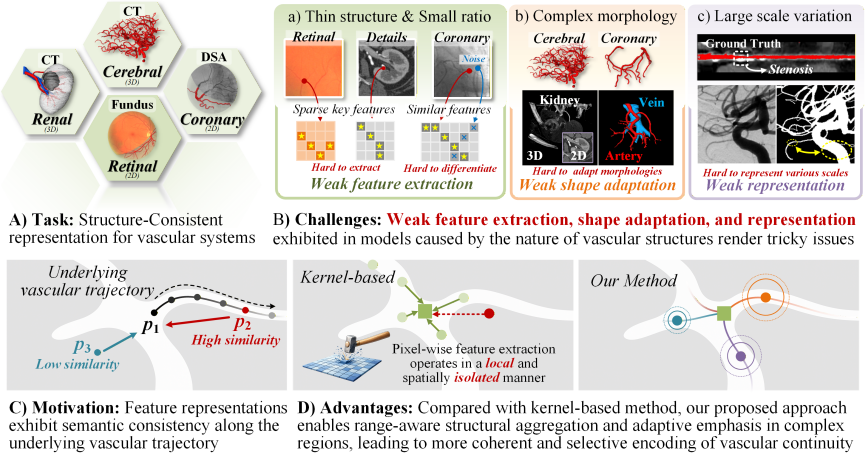

Fig. 1. The figure illustrates (A) the task of structure-consistent vascular representation, (B) key challenges arising from thin geometry, complex morphology, and scale variation, (C) the motivation that semantic features remain consistent along vascular trajectories, and (D) the advantage of our method over kernel-based approaches by enabling structure-aware feature propagation instead of isolated pixel-wise estimation.

collections of tubular pixels or voxels, but hierarchically connected and directionally coherent structures whose quantitative analysis and clinical interpretation critically depend on structural integrity. Therefore, reliable vascular modeling requires not only local segmentation accuracy, but also anatomically consistent representations that respect the organizational principles of the vascular tree.

Vascular representation is inherently challenging because the anatomical and morphological properties of vascular systems impose stringent requirements on structural preservation. As shown in Fig. 1 B, these challenges arise from three factors. First, the extremely small vessel-to-background ratio and thin branch geometry produce weak discriminative cues and unstable feature responses, thereby increasing the sensitivity to noise and local appearance variation. Second, the complex tree-like morphology, characterized by tortuous trajectories and frequent bifurcations, requires robust shape adaptation to preserve structural continuity and reduce ambiguity in branching regions. Third, substantial intra- and inter-branch scale variation, from major trunks to distal microvasculature, complicates consistent representation across hierarchical feature levels.

To mitigate these difficulties, existing studies generally follow two methodological directions. One direction focuses on improving vascular modeling by redesigning operators or architectural components to capture tubular geometry and structural continuity [3,4,5,6]. The other line introduces topology- or centerline-sensitive objectives to regularize predictions and discourage structural fragmentation during optimization [7]. Although these approaches have achieved promising results in various settings, structural priors are often in-

corporated through tailored operator or architectural designs or through additional objective-level regularization. In contrast, existing frameworks often lack a backbone-agnostic feature-level mechanism that can explicitly and progressively inject anatomical cues into hierarchical feature fusion while enabling selective reinforcement according to local representational difficulty. Consequently, structural continuity and branching organization are still largely encoded implicitly in intermediate representations, which may limit robustness in challenging regions.

Inspired by intrinsic structural coherence along vascular trajectories, it is important to examine how such consistency can be reflected in intermediate representations. The feature along the same vascular branch tends to remain semantically aligned despite local intensity or texture variation. This suggests that vascular modeling should preserve interactions along anatomical paths rather than treat each spatial location as an independent estimation unit (Fig. 1 C). Some structure-guided convolutional designs [3] incorporate such cues by adapting kernel shapes or sampling positions to align with tubular geometry. Although they improve local shape adaptivity, they remain fundamentally kernel-based. Specifically, the feature at a target location is computed from a finite set of discretely sampled neighbors and aggregated in a local operation. Thus, each sampled point contributes only pixel-level information, and aggregation is performed independently at each spatial location (Fig. 1 D). Thus, structural information is integrated in a localized one-step manner, without explicit accumulation of long-range dependencies across hierarchical feature levels. This limitation is particularly pronounced in branches and complex bifurcation regions, where structural coherence extends beyond the range of discrete local sampling.

To address the limitations, we propose a flexible structure-guided feature aggregation paradigm for vascular representation. The main contributions are summarized as follows: 1) **Structure-guided feature aggregation at the representation level**. We formulate vascular modeling as hierarchical representation learning and introduce a structure-guided aggregation branch that explicitly models interactions along anatomically consistent paths, enabling progressive accumulation of structural dependencies beyond localized kernel-based estimation. 2) **Adaptive structural modulation**. We design a dynamic modulation mechanism that adjusts the strength of structural guidance according to regional complexity, allowing efficient reinforcement in difficult vascular regions while maintaining representational flexibility elsewhere. 3) **Flexible architecture-agnostic integration**. The proposed aggregation paradigm is lightweight and modular, allowing seamless integration into a variety of backbone architectures, including convolutional, transformer-based, and state-space models, thus providing a general and efficient framework for vascular representation.

2 Methodology

2.1 Problem Formulation and Structural Aggregation Perspective

Let $X \in \mathbb{R}^{C_{in} \times H \times W}$ denote a 2D vascular image, or $X \in \mathbb{R}^{C_{in} \times D \times H \times W}$ denote a 3D vascular volume. For clarity, we first present the 2D formulation on the dis-

crete image domain $\Omega \subset \mathbb{R}^2$, and then describe its 3D extension in Sec. 2.2. The same residual aggregation principle is shared by both cases. A backbone network produces hierarchical feature maps $\{\mathbf{F}^{(l)}\}_{l=1}^{L}$, where $l \in \{1, \cdots, L\}$ indexes the network stage, $\mathbf{F}^{(l)} \in \mathbb{R}^{C_l \times H_l \times W_l}$, where C_l is the number of channels at stage l, and H_l, W_l are the spatial resolutions at that stage. The segmentation head predicts a vascular mask Y from these features, where $Y \in [0,1]^{H \times W}$.

Our objective is to improve the intermediate representation $\mathbf{F}^{(l)}$ so that structural continuity and branching organization are encoded with greater robustness. To this end, we introduce an auxiliary Vascular Structural Prior Branch (VSP-Branch) that operates on a chosen feature map $\mathbf{F}$ (we omit the stage superscript l when the context is clear). The branch generates a structure-aware residual update $\Delta\mathbf{F} \in \mathbb{R}^{C \times H \times W}$, and a Difficulty Gate $D \in [0,1]^{1 \times H \times W}$ that modulates the spatial injection strength. The enhanced feature map is defined as follows:

$$\mathbf{F}_{out}(\mathbf{p}) = \mathbf{F}_{in}(\mathbf{p}) + D(\mathbf{p})\Delta\mathbf{F}(\mathbf{p}), \tag{1}$$

where $\mathbf{p} \in \Omega$ denotes a spatial location (pixel), $\mathbf{F}_{in}(\mathbf{p}) \in \mathbb{R}^C$ is the input feature vector at $\mathbf{p}$, $\mathbf{F}_{out}(\mathbf{p}) \in \mathbb{R}^C$ is the output feature vector and $D(\mathbf{p})$ is a scalar gate value shared across channels. This formulation keeps the backbone unchanged and injects structural evidence through a dedicated fusion pathway. The gate is learned only through the final segmentation objective and does not require manually defined difficulty annotations.

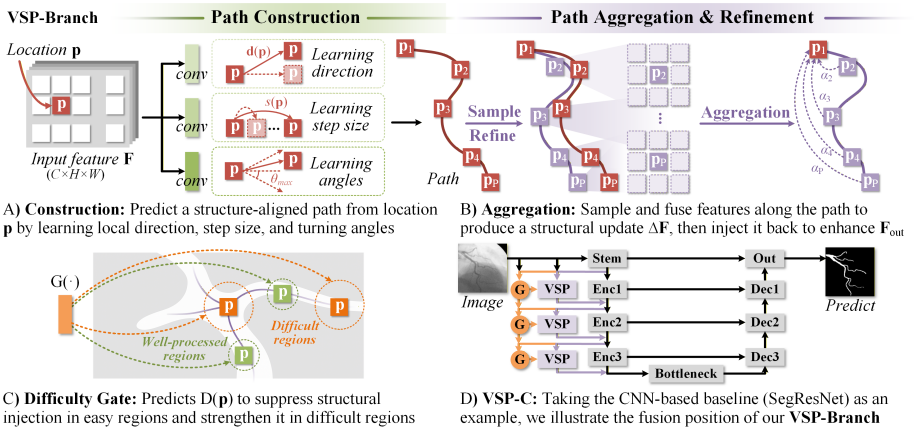

Fig. 2. Our method. The VSP-Branch includes two components: (A) path construction, which learns a structure-aligned sampling trajectory; and (B) path-based feature sampling and fusion, which generates a structural update for feature enhancement. (C) A difficulty gate adaptively modulates the injection strength in difficult regions. (D) We illustrate integration into a CNN-based baseline (SegResNet), and the backbone can be directly replaced with Transformer- or Mamba-based architectures.

2.2 VSP-Branch: Path Construction and Differentiable Path Fusion

Given an input feature map $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$, the VSP-Branch constructs, for each location $\mathbf{p} \in \Omega$, an ordered sequence of sampling points $\{\mathbf{p}_i\}_{i=0}^{P-1}$ that follows regionally inferred vascular geometry. Here, P is the number of steps along the path. The branch predicts three spatially varying quantities from $\mathbf{F}$: direction, step size, and turning angles, using lightweight convolutional predictors.

Direction. The branch predicts a direction vector $\mathbf{d}(\mathbf{p}) = (d_x(\mathbf{p}), d_y(\mathbf{p})) \in \mathbb{R}^2$. To ensure stable geometry, it is normalized as follows: $\mathbf{v}_0(\mathbf{p}) = \frac{\mathbf{d}(\mathbf{p})}{||\mathbf{d}(\mathbf{p})||_2+\varepsilon}$, where $||\cdot||_2$ denotes the Euclidean norm, and ε is a small constant for stability.

Step size. The branch predicts a scalar step size $s(\mathbf{p}) \in \mathbb{R}_+$ and bounds it to a finite interval by a sigmoid mapping: $s(\mathbf{p}) \in [s_{min}, s_{max}]$, where $s_{min} > 0$ and $s_{max} > s_{min}$ are implementation constants (in pixel units). Bounding the step size prevents unstable long jumps during training.

Turning angles. To model curvature while maintaining a controlled path geometry, the branch predicts incremental turning angles $\{\theta_i(\mathbf{p})\}_{i=1}^{P-1}$, each bounded within $[-\theta_{max}, \theta_{max}]$, using a tanh-parameterized approach, where $\theta_{max} > 0$ (in radians) corresponds to the maximum turning angle.

Starting from $\mathbf{o}_0(\mathbf{p}) = \mathbf{0} \in \mathbb{R}^2$, the branch recursively generates offsets $\{\mathbf{o}_i(\mathbf{p})\}_{i=1}^{P-1}$ in spatial coordinates. For each step $i \geq 1$, the direction is rotated and the offset is advanced: $\mathbf{v}_i(\mathbf{p}) = \mathbf{R}(\theta_i(\mathbf{p}))\mathbf{v}_{i-1}(\mathbf{p}), \mathbf{o}_i(\mathbf{p}) = \mathbf{o}_{i-1}(\mathbf{p}) + s(\mathbf{p})\mathbf{v}_i(\mathbf{p})$, where $\mathbf{R}(\theta) \in \mathbb{R}^{2\times 2}$ is the standard 2D rotation matrix for angle θ. The resulting sampling points are $\mathbf{p}_i = \mathbf{p} + \mathbf{o}_i(\mathbf{p})$. This recursion produces a curvature-limited path ideal for propagating information along trajectories.

We fuse evidence along the path by sampling the feature map at each point using differentiable bilinear interpolation: $\hat{\mathbf{F}}_i(\mathbf{p}) = \text{Sample}(\mathbf{F}, \mathbf{p}_i)$, where $\hat{\mathbf{F}}_i(\mathbf{p}) \in \mathbb{R}^C$ is the sampled feature vector and Sample$(\cdot)$ denotes bilinear sampling.

The branch also predicts point-wise fusion weights $\alpha_i(\mathbf{p})$ over steps and normalizes them across i, so that $\Sigma_{i=0}^{P-1}\alpha_i(\mathbf{p}) = 1$. In implementation, this is implemented by a P-channel predictor followed by softmax normalization or sigmoid-plus-renormalization. The path-aggregated feature is then:

$$\mathbf{F}_{path}(\mathbf{p}) = \Sigma_{i=0}^{P-1}\alpha_i(\mathbf{p})\hat{\mathbf{F}}_i(\mathbf{p}). \tag{2}$$

Finally, we map this aggregated evidence back to the feature space using a 1×1 mixing convolution M and scale it by a learnable scalar $\gamma \in \mathbb{R}$ initialized with a small positive value for stable training: $\Delta\mathbf{F}(\mathbf{p}) = \gamma\text{M}(\mathbf{F}_{\text{path}}(\mathbf{p}))$. This design implements structure-aligned fusion along a learned path while keeping the modification lightweight and compatible with standard backbones.

Extension to 3D. For volumetric data, VSP-Branch follows the same stepwise path evolution but uses directional vector updates instead of scalar turning-angle increments, since a single angle is insufficient to describe 3D direction changes. Specifically, for each voxel $p \in \Omega \subset \mathbb{R}^3$, the branch predicts an initial direction $v_0(p) \in \mathbb{R}^3$, a bounded step size $s(p)$, and per-step directional updates $\delta_i(p) \in \mathbb{R}^3$. The sampling locations are $p_i = p + o_i(p)$. Features are sampled using trilinear interpolation and fused with normalized step weights as in Eq.2.

2.3 Difficulty Gate for Selective Structural Infusion

Since vascular complexity is spatially heterogeneous, applying the same structural infusion everywhere can be suboptimal and may over-bias easy regions. We therefore use a Difficulty Gate to learn where the structure-guided residual is useful. We therefore use a Difficulty Gate to modulate the residual update spatially. Given the feature map $\mathbf{F}$, the gate predicts: $D(\mathbf{p}) = \sigma(\mathrm{G}(\mathbf{F})(\mathbf{p}))$, where $\mathrm{G}(\cdot)$ is a lightweight convolutional predictor producing a scalar map, $\mathrm{G}(\mathbf{F})(\mathbf{p}) \in \mathbb{R}$, and $\sigma(\cdot)$ denotes the sigmoid function, ensuring $D(\mathbf{p}) \in [0, 1]$. The final fusion at each location follows the residual definition in Sec. 2.1: $\mathbf{F}_{out}(\mathbf{p}) = \mathbf{F}_{in}(\mathbf{p}) + D(\mathbf{p})\Delta\mathbf{F}(\mathbf{p})$. Importantly, D is not supervised by explicit difficulty labels. It is optimized end-to-end through the segmentation loss, acting as a utility-driven gate that increases the contribution of structural residuals when they help improve the final prediction. As a consequence, the structural branch can emphasize continuity cues more strongly in regions that are difficult to represent (e.g. thin distal branches or bifurcations) while preserving the backbone feature in simpler regions, thereby avoiding unnecessary bias.

3 Experiments

Public datasets. To validate the effectiveness of the proposed method, we performed experiments on four tasks: 2D coronary arteries (DCA1 [8]), 3D cerebrovascular structures (COSTA [9]), 2D retinal vessels (ARIA [10], CHASEDB1 [11], DRIVE [12], HRF [13], STARE [14] merged into a unified benchmark-DeepRetina) and 3D renal vessels (KIPA [15]). These datasets span 2D and 3D imaging, different resolutions, and a wide range of structural complexities, providing a large testbed to assess the effectiveness of our VSP-Branch as a plug-and-play module across architectures.

Evaluation Metrics. To quantitatively evaluate the performance of different models, several widely adopted metrics are used, including the Dice coefficient (Dice), Relative Dice (RDice [16]), CenterlineDice (clDice [7]), Accuracy (ACC), Area Under the Curve (AUC) and Hausdorff Distance (HD).

Compared Methods. We compare our method against a large panel of representative methods, covering both classical architectures and very recent state-of-the-art models up to 2025. For clarity, the baselines are organized into 3 families: CNN-based models (U-Net [18], SegResNet [17], DSCNet[3], CMUNeXt [20], MEGANet [21], and ResU-KAN [4]), ViT-based models (SwinUNETR [22], CS2Net [19], SCUNet++ [23], MSLAU-Net [24], PA-Net [26], and CSWin-UNet [25]), and Mamba-style sequence models (U-Mamba [27], Swin-UMamba [28], CFM-UNet [29], MedVKAN [30], and H-vmunet [31]).

Quantitative evaluation. Tab. 1 summarizes the results. VSP-Branch consistently improves accuracy across various backbone families, including CNN, ViT and Mamba, showing that the gains stem from the proposed path-guided structural fusion, not specific network architecture. For example, on DCA1, the gains are clear even when starting from a strong Transformer baseline. For example, when applied to Swin-UNETR, adding VSP-Branch improves Dice from

Table 1. Results on COSTA, DCA1 (top), KIPA, and DeepRetina (bottom). Methods are grouped by backbone type (CNN, ViT, and Mamba), where "(base)" denotes the corresponding baseline backbone and "VSP-* (ours)" denotes the same backbone augmented with the proposed VSP-Branch. Best and second-best results are highlighted.

	Dataset	**COSTA**						**DCA1**					
Type	**Method**	**Volume** (%)					**Distance**	**Volume** (%)					**Distance**
		Dice ↑	RDice ↑	clDice ↑	ACC ↑	AUC ↑	HD ↓	Dice ↑	RDice ↑	clDice ↑	ACC ↑	AUC ↑	HD ↓
CNN	SegResNet [17] (base)	78.31	83.84	84.20	99.75	86.15	7.12	77.08	87.70	83.07	97.56	88.58	**4.45**
	U-Net [18]	78.01	84.96	83.77	99.74	86.93	8.56	76.18	87.47	82.19	97.44	88.35	4.49
	CS2-Net [19]	88.43	94.38	88.89	99.85	94.70	9.10	76.59	86.33	82.73	97.56	87.80	4.53
	DSCNet [3]	87.79	91.55	88.08	99.86	92.33	10.40	77.24	87.88	83.13	97.56	88.72	4.61
	CMUNeXt [20]	86.88	92.88	89.02	99.84	93.34	8.02	72.18	83.65	77.00	97.10	85.92	4.58
	MEGANet [21]	74.70	82.59	76.37	99.70	85.30	15.30	76.23	87.71	82.63	97.43	88.59	4.49
	ResU-KAN [4]	88.17	**94.78**	87.19	99.85	**95.08**	12.79	75.92	87.13	82.07	97.42	88.13	4.62
	VSP-C (ours)	**91.23**	94.74	**91.32**	**99.89**	95.02	**6.52**	**77.83**	**89.21**	**83.61**	**97.59**	**89.70**	4.67
ViT	SwinUNETR [22] (base)	88.55	**94.46**	86.15	99.86	94.32	15.90	74.33	85.76	79.66	97.27	87.08	**4.54**
	SCUNet++ [23]	71.62	82.73	74.13	99.65	85.35	16.44	74.79	87.03	80.01	97.26	88.06	4.59
	MSLAU-Net [24]	79.89	86.81	81.08	99.76	88.46	12.65	69.15	80.41	73.57	96.84	83.36	4.84
	CSWin-UNet [25]	79.02	85.53	81.06	99.75	87.40	11.88	73.84	84.39	78.39	97.28	86.20	4.62
	PA-Net [26]	88.41	94.04	**88.87**	99.86	94.38	**9.54**	75.23	86.99	81.12	97.34	88.06	4.62
	VSP-V (ours)	**89.12**	94.11	88.35	**99.87**	**94.46**	10.00	**76.05**	**87.55**	**81.52**	**97.42**	**88.42**	4.56
Mamba	U-Mamba [27] (base)	90.82	95.20	90.76	99.89	95.40	**7.09**	77.94	88.63	83.62	**97.62**	89.31	4.54
	Swin-UMamba [28]	91.27	95.61	89.86	99.89	95.79	10.17	77.01	88.54	83.19	97.51	89.18	4.54
	CFM-UNet [29]	85.29	89.53	84.96	99.83	90.68	13.00	69.76	83.45	75.44	96.76	85.38	4.71
	MedVKAN [30]	74.45	83.52	76.27	99.65	86.49	18.55	76.28	87.02	82.22	97.47	88.13	**4.52**
	H-vmunet [31]	85.97	92.18	86.74	99.82	92.73	12.54	70.24	81.08	74.88	96.95	83.78	4.70
	VSP-M (ours)	**91.54**	**95.92**	**91.19**	**99.90**	**96.06**	7.76	**78.58**	**88.74**	**84.05**	**97.70**	**89.38**	4.54

	Dataset	**KIPA**						**DeepRetina**					
Type	**Method**	**Volume** (%)					**Distance**	**Volume** (%)					**Distance**
		Dice ↑	RDice ↑	clDice ↑	ACC ↑	AUC ↑	HD ↓	Dice ↑	RDice ↑	clDice ↑	ACC ↑	AUC ↑	HD ↓
CNN	SegResNet [17] (base)	84.17	91.28	85.93	99.80	92.66	20.56	77.08	86.77	79.02	96.41	87.52	7.01
	U-Net [18]	84.78	89.82	85.35	99.81	91.33	22.25	76.82	86.62	78.59	96.37	87.37	6.98
	CS2-Net [19]	82.37	89.94	84.35	99.79	91.99	21.38	73.74	80.53	74.34	96.22	83.26	6.96
	DSCNet [3]	86.86	92.26	88.13	99.84	93.51	12.40	77.34	86.42	79.05	**96.50**	87.33	6.94
	CMUNeXt [20]	80.30	84.61	82.21	99.78	88.21	14.06	71.07	75.93	70.64	96.09	80.40	6.94
	MEGANet [21]	84.56	89.30	85.60	99.82	90.83	**10.87**	74.64	82.94	75.52	96.21	84.79	6.97
	ResU-KAN [4]	84.41	88.99	83.91	99.81	91.07	23.70	76.08	84.01	77.21	96.40	85.60	6.97
	VSP-C (ours)	**88.11**	**93.39**	**88.53**	**99.85**	**94.09**	12.87	**77.49**	**86.91**	**79.16**	96.48	**87.65**	**6.91**
ViT	SwinUNETR [22] (base)	82.53	**90.70**	80.47	99.78	92.14	28.58	74.26	83.18	74.70	96.13	84.90	7.00
	SCUNet++ [23]	59.50	65.67	56.03	99.56	75.68	36.92	75.18	84.90	76.56	96.18	86.13	7.00
	MSLAU-Net [24]	75.73	82.73	72.93	99.71	86.37	31.14	72.44	83.21	72.89	95.75	84.79	7.06
	CSWin-UNet [25]	76.61	83.51	71.23	99.72	86.69	29.07	74.91	84.34	75.73	96.17	85.70	7.01
	PA-Net [26]	84.06	89.49	84.27	**99.81**	91.15	**15.04**	71.10	77.76	71.51	95.94	81.38	6.99
	VSP-V (ours)	**84.62**	90.62	**84.65**	**99.81**	**92.44**	22.21	**75.85**	**85.26**	**76.94**	**96.29**	**86.43**	**6.96**
Mamba	U-Mamba [27] (base)	83.12	91.61	83.32	99.79	92.72	26.70	77.20	86.90	78.70	96.42	87.57	**6.87**
	Swin-UMamba [28]	82.48	90.34	79.41	99.79	91.97	34.11	76.94	86.16	78.36	96.43	87.08	6.94
	CFM-UNet [29]	79.12	89.48	78.14	99.73	91.03	36.36	73.16	86.12	75.11	95.63	86.65	7.10
	MedVKAN [30]	74.13	80.20	76.29	99.71	84.87	**19.48**	72.38	78.43	72.28	96.12	81.94	7.00
	H-vmunet [31]	68.00	76.88	66.83	99.64	82.52	32.20	74.90	86.71	76.62	95.97	87.24	7.05
	VSP-M (ours)	**84.89**	**91.79**	**85.13**	**99.81**	**92.91**	21.51	**78.15**	**88.23**	**80.19**	**96.53**	**88.61**	6.96

Table 2. Ablation study on DSA1. The impact of the Difficulty Gate (DG), the number of path sampling steps, and the step-size range ($s_{min/max}$) is evaluated.

Method	DG	Steps	$s_{min/max}$	Dice ↑	RDice ↑	clDice ↑	DG	Steps	$s_{min/max}$	Dice ↑	RDice ↑	clDice ↑
SegResNet [17] (base)	–	–	–	77.08	87.70	83.07	–	–	–	77.08	87.70	83.07
VSP-C (ours)	*w/o*	7	1/5	77.56	88.83	83.18	*w/*	9	1/5	77.31	88.69	83.23
VSP-C (ours)	*w/*	**7**	**1/5**	**77.83**	89.21	**83.61**	*w/*	7	1/3	77.47	88.84	83.43
VSP-C (ours)	*w/*	5	1/5	77.41	**90.62**	83.58	*w/*	7	1/7	77.67	89.93	83.54

74.33% to 76.05%. Notably, this improvement comes with negligible computational overhead: the parameter count increases only from **6.287M** to **6.288M**, and FLOPs from **4.859G** to **4.884G**, demonstrating that VSP-Branch is a lightweight plug-in rather than a heavy architectural redesign.

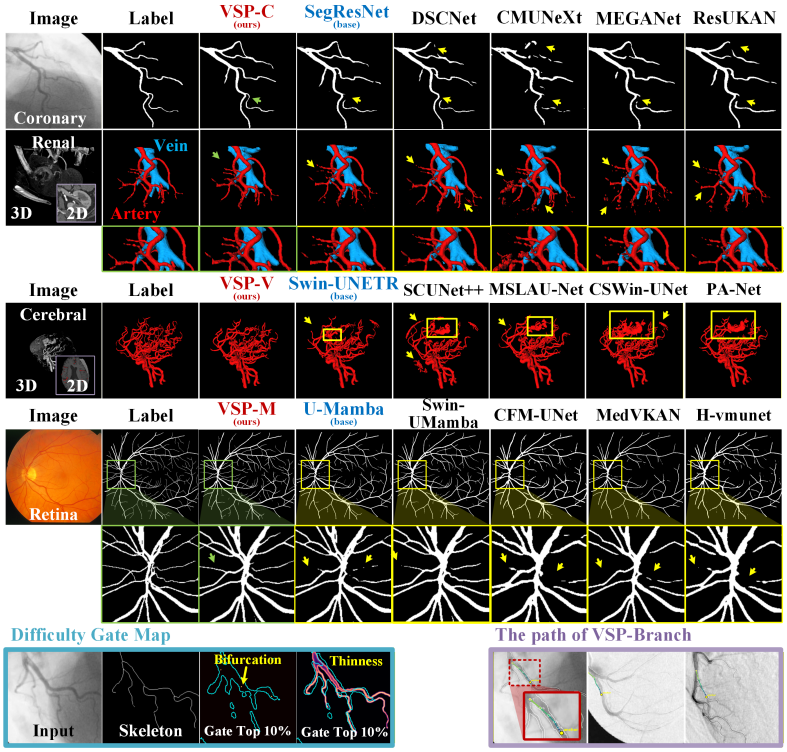

Fig. 3. Qualitative comparison on multi-dataset vascular segmentation

Ablation study. Tab. 2 presents an ablation study on DCA1 that examines both the module design and key hyper-parameters of VSP-C, including the Difficulty Gate (DG), the number of path sampling steps, and the step-size range (s_{min}/s_{max}). The ablation results demonstrate VSP-C consistently improves the baseline, with the Difficulty Gate further enhancing structural preservation through selective reinforcement in challenging regions. **Qualitative evaluation**. Fig. 3 shows how VSP-Branch improves vascular representation. Beyond segmentation outputs, blue overlays visualize Difficulty Gate responses, which place greater emphasis on structurally ambiguous regions. The gate consistently focuses on bifurcation points and thin branches, where conventional models often produce discontinuities. This selective modulation helps preserve fragile vessel boundaries and maintain continuity across branching junctions, reducing broken

segments in challenging regions. Meanwhile, purple traces show the learned path-guided sampling trajectories used by VSP-Branch during training. These trajectories align with the vascular direction and follow tubular structures, rather than dispersing isotropically, indicating that VSP-Branch learns structure-consistent paths for feature propagation.

4 Conclusion

In this work, we introduce VSP-Branch, a structure-guided feature aggregation method for vascular representation. By incorporating path-guided fusion, VSP-Branch enables feature interactions that surpass local sampling. The 'Difficulty Gate' module adjusts the guidance by reinforcing challenging regions and avoiding unnecessary enhancement elsewhere. Experiments on four datasets show consistent improvements with minimal computational overhead, supporting structure-guided fusion as an efficient solution for vascular representation.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (82441021), National Key Research and Development Program of China (2024YFC2417803), Jiangsu Province Cutting-edge Technology R&D Program (BF2025628), the Postgraduate Research & Practice Innovation Program of Jiangsu Province, the Fundamental Research Funds for the Central Universities (KYCX22_0239), and the Big Data Computing Center of Southeast University. This work was partially supported by SEU Kunpeng & Ascend Center of Cultivation.

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

References

1. Hao Li, Zeyu Tang, Yang Nan, and Guang Yang. Human treelike tubular structure segmentation: A comprehensive review and future perspectives. *Computers in biology and medicine*, 151:106241, 2022.
2. Chengyang Du, Jie Zhuang, and Xinglu Huang. Deep learning technology in vascular image segmentation and disease diagnosis. *Journal of Intelligent Medicine*, 1(1):6–41, 2024.
3. Yaolei Qi, Yuting He, Xiaoming Qi, Yuan Zhang, and Guanyu Yang. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6070–6079, October 2023.
4. Haibin Wang, Zhenfeng Zhao, Qi Liu, and Shenwen Wang. Resu-kan: a medical image segmentation model integrating residual convolutional attention and atrous spatial pyramid pooling. *Applied Intelligence*, 55(7):568, 2025.
5. Pengsheng Song, Huanhuan Zhang, Junfeng Jing, Pengfei Li, and Quan Pan. Dpanet: curve-like structure segmentation based on dual-path attention network. *Expert Systems with Applications*, page 129030, 2025.
6. Xiangxin Wang, Zhan Wu, Yujia Zhou, et al. Topology-oriented foreground focusing network for semi-supervised coronary artery segmentation. *Medical Image Analysis*, 101:103465, 2025.

7. Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, et al. cldice - a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16560–16569, June 2021.
8. Fernando Cervantes-Sanchez, Ivan Cruz-Aceves, Arturo Hernandez-Aguirre, et al. Automatic segmentation of coronary arteries in x-ray angiograms using multiscale analysis and artificial neural networks. *Applied Sciences*, 9(24):5507, 2019.
9. Lei Mou, Jinghui Lin, Yifan Zhao, et al. Costa: A multi-center tof-mra dataset and a style self-consistency network for cerebrovascular segmentation. *IEEE transactions on medical imaging*, 43(12):4442–4456, 2024.
10. Damian JJ Farnell, Fraser N Hatfield, Paul Knox, et al. Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *Journal of the Franklin institute*, 345(7):748–765, 2008.
11. Christopher G Owen, Alicja R Rudnicka, Robert Mullen, et al. 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.
12. Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
13. Jan Odstrcilik, Radim Kolar, Attila Budai, et al. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Processing*, 7(4):373–383, 2013.
14. AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.
15. Yuting He, Guanyu Yang, Jian Yang, et al. Meta grayscale adaptive network for 3d integrated renal structures segmentation. *Medical image analysis*, 71:102055, 2021.
16. Yaolei Qi, Han Xu, Yuting He, et al. Examinee-examiner network: Weakly supervised accurate coronary lumen segmentation using centerline constraint. *IEEE Transactions on Image Processing*, 30:9429–9441, 2021.
17. Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI brainlesion workshop*, pages 311–320. Springer, 2018.
18. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, et al. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.
19. Lei Mou, Yitian Zhao, Huazhu Fu, et al. Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical image analysis*, 67:101874, 2021.
20. Fenghe Tang, Jianrui Ding, Quan Quan, et al. Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2024.
21. Nhat-Tan Bui, Dinh-Hieu Hoang, Quang-Thuc Nguyen, et al. Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7985–7994, 2024.

22. Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, et al. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop*, pages 272–284. Springer, 2021.
23. Yifei Chen, Binfeng Zou, Zhaoxin Guo, Yiyu Huang, Yifan Huang, Feiwei Qin, Qinhai Li, and Changmiao Wang. Scunet++: Swin-unet and cnn bottleneck hybrid architecture with multi-fusion dense skip connection for pulmonary embolism ct image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 7759–7767, 2024.
24. Libin Lan, Yanxin Li, Xiaojuan Liu, Juan Zhou, Jianxun Zhang, Nannan Huang, and Yudong Zhang. Mslau-net: A hybird cnn-transformer network for medical image segmentation. *arXiv preprint arXiv:2505.18823*, 2025.
25. Xiao Liu, Peng Gao, Tao Yu, et al. Cswin-unet: Transformer unet with cross-shaped windows for medical image segmentation. *Information Fusion*, 113:102634, 2025.
26. Xuebing Luo, Lingxi Peng, Ziyan Ke, Jinhui Lin, and Zhiwen Yu. Pa-net: A hybrid architecture for retinal vessel segmentation. *Pattern Recognition*, 161:111254, 2025.
27. Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.
28. Jiarun Liu, Hao Yang, Hong-Yu Zhou, et al. Swin-umamba: Mamba-based unet with imagenet-based pretraining. In *International conference on medical image computing and computer-assisted intervention*, pages 615–625. Springer, 2024.
29. Ke Niu, Jiacheng Han, and Jiuyun Cai. Cfm-unet: coupling local and global feature extraction networks for medical image segmentation. *Scientific Reports*, 15(1):22236, 2025.
30. Hancan Zhu, Jinhao Chen, and Guanghua He. Medvkan: Efficient feature extraction with mamba and kan for medical image segmentation. *arXiv preprint arXiv:2505.11797*, 2025.
31. Renkai Wu, Yinghao Liu, Pengchen Liang, and Qing Chang. H-vmunet: High-order vision mamba unet for medical image segmentation. *Neurocomputing*, 624:129447, 2025.