Dual-Field Microvascular Segmentation: Hemodynamically-Consistent Attention Learning for Retinal Vasculature Mapping

Jingling Zhang[1,2], Xiangfei Liu[1,3], Shuting Zheng[1,2], Wen Zhang[4], Jia Gu [1*]

1 Institute of Data Science, City University of Macau, Macau, 999078, China;
2 New Engineering Industry College, Putian University, Putian, Fujian Province 351100, China;
3 Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, 518055, China;
4 Charles E. Schmidt College of Science, Florida Atlantic University, Boca Raton, FL 33431, USA;
D24092100270@cityu.edu.mo； xfliu0102@163.com; fgtear@ptu.edu.cn；
zhangw@fau.edu；jiagu@cityu.edu.mo

Abstract: Accurate segmentation of retinal microvasculature necessitates a dual-field synergy between anatomical fidelity and hemodynamic relevance, yet existing methods often fragment critical structures such as capillary junctions and bifurcations, undermining clinical utility. To address this, we propose DFMS-Net, a hemodynamically-inspired architecture integrating three key innovations: 1) **Geometric-Field Modeling** employs (a) **Spatial Pathway Extractor (SPE)** to model vascular trajectory continuity through preserving geometric-hemodynamical consistency, and (b) **Transformer-derived Topology Interaction module (TTI)** to resolve complex branching patterns through spatial connectivity analysis. 2) **Functional-Field Optimization** leverages channel-wise **Semantic Attention Amplification** (SAA) to enhance low-contrast capillaries while suppressing imaging artifacts, prioritizing clinically salient features like microaneurysms. 3) C Learning unifies these fields through a **Dual-Field Synergy Unit (DFSU)**, bridging geometric precision with hemodynamically-inspired semantic consistency. Validations on retinal (DRIVE, STARE, CHASE_DB1) and coronary angiography (DCA1, CHUAC) datasets, demonstrates that, our framework, DFMS-Net redefines microvascular phenotyping for precision ophthalmology and blood-related diseases risk prediction.

***Index Terms*—**Retinal Microvasculature; Segmentation; DFUS; Hemodynamically-Consistent; Attention

I. Introduction

Precise mapping of retinal vasculature networks, integrating both structural-spatial integrity and hemodynamically relevant mapping, forms a critical diagnostic cornerstone in the evaluation of small vessel disorders[1], [2]. The morphological characteristics of retinal vessels, such as vascular width, tortuosity, and branching patterns, act as sensitive biomarkers for ocular pathologies, and play a critical role in diagnosing and predicting the prognosis of common eye diseases like diabetic retinopathy, glaucoma, and age-related macular degeneration[1], [3]. **Beyond ophthalmology,** emerging evidence suggests that the retinal microvasculature also has far-reaching effects on the occurrence of circulatory system diseases. Retinal microvascular abnormalities can increase the risk of stroke by 2- to 3-fold and are closely connected to the generation of subclinical atherosclerosis[4]. Regarding hemodynamics, analyzing retinal vascular hemodynamics facilitates the early discovery of abnormalities in circulatory system-related diseases, including stroke, ischemic heart disease, heart failure, and renal dysfunction.

Thus, clinicians can take measures before overt morphological changes happen[2]. Considering the clinical significance of the structural spatial features and hemodynamic consistency of retinal vessels, developing automated retinal segmentation methods capable of tackling both geometric topology and blood flow consistency has become an urgent task.

However, automated retinal vessel segmentation still faces the following challenges. 1) Anatomical complexity: Significant variations in vessel size and interference from pathological factors make it arduous to distinguish overlapping vessel structures and detect subtle lesions related to diabetic retinopathy[3]. 2) Imaging limitations: The low contrast between capillaries and background noise, coupled with motion artifacts, makes accurate segmentation quite complex[5]. 3) Tree-like structure challenges: The typical tree-like structure, featuring slender and narrow branches, often has distal branches that are too fine to be precisely segmented, thereby intensifying computational challenges. Apart from the inherent properties of medical images, the small size of medical image datasets, resulting from the high cost and complexity of data acquisition and annotation, presents major challenges for effective model training[6], [7], [8].

Currently, deep learning techniques for retinal vessel segmentation primarily adhere to two patterns: U-Net-based and Transformer-based. Although performing well in hierarchical feature aggregation thanks to the encoder-decoder design, U-Net[9] and its variants are inherently short of the ability of hemodynamic-aware geometric modeling. The local receptive field disrupt the continuity of vessels, and the undifferentiated feature fusion warps key angles. In contrast, while capable of capturing long-range dependencies, transformer-based methods are limited by two factors: 1) the quadratic computational complexity that hampers clinical implementation, and 2) the patch-level processing that results in the degeneration of capillary delineation. In an attempt to combine these patterns, hybrid architectures have insufficiently addressed the dual-domain discrepancies—they fail to balance structural fidelity and hemodynamic relevance, and also fall short in resolving the critical gap for clinically operational segmentation outcomes, leading to unreliable vascular mapping in diagnostic and prognostic contexts.

Aiming to overcome these challenges, we present a Dual-Field Microvascular Segmentation framework, incorporating geometric field modeling and functional field optimization. Our work features the following significant contributions to retinal vessel segmentation, pushing forward the progress of retinal vessel segmentation:

- We propose the first framework via the Dual - Field Synergistic Unit (DFSU) model. It uses hemodynamics - guided attention learning to unify geometric field modeling (preserving structural continuity) and functional field optimization (enhancing clinical semantic significance), systematically solving the fragmentation of key anatomical structures.
- We introduce a Geometric-Field Modeling with hierarchical attention combing Spatial Pathway Extractor (SPE)-Employs directional coordinate attention to preserve vascular trajectory continuity, resolving spatial discontinuities in branching networks, with Transformer-derived Topology Interaction (TTI)-Leveraging transformer spatial reasoning to decode complex bifurcation patterns and ensuring anatomical validity.
- We explore a novel Functional-Field Optimization model, utilizing the Semantic Attention Amplifier (SAA) dynamically recalibrates channel-wise features to enhance low-contrast

capillaries and suppress imaging artifacts.

- We verify the cross-domain hemodynamic consistency across retinal and coronary angiography datasets, this mechanism enables robust generalization to diverse vascular imaging modalities while maintaining clinical interpretability.

II. Related Work

A. CNN-based Medical Image Segmentation

CNN-based architectures have established foundational paradigms in medical image analysis, with U-Net[9] remaining the cornerstone for its encoder-decoder design that balances localization and context. Its variants—including UNet++[10] for nested feature refinement, FR-UNet[11] for restoring the lost spatial information, DenseUNet[12] for multi-scale connectivity, I^2U-Net[13] for bridging the encoder and the decoder semantic gap, and MSVM-UNet[14] for simultaneous capturing of localized fine-grained patterns and comprehensive contextual awareness across vascular hierarchies. In addition to the aforementioned methodologies, several other approaches such as V-Net[15], nnU-Net[16], KiU-Net[17], and Point-Unet[18] have been developed for 3D medical image segmentation. Convolutional neural network (CNN)-driven approaches have revolutionized medical image segmentation, yet struggle with preserving vascular continuity across bifurcations—critical for hemodynamic simulation.

B. Transformer-based Medical Image Segmentation

Recent advances in Transformer-based medical image segmentation primarily follow four paradigms: **1)** hybrid architectures (e.g., TransUNet [20], MISSFormer [21]) that integrate CNNs and Transformers, **2)** efficient attention designs (e.g., Swin-trans [23], DAE-Former [22]) to mitigate computational overhead, **3)** generative models (e.g., MedSegDiff [25]) for noise-robust segmentation, and **4)** graph-based approaches (e.g., GFT-EFA [24]) for topological reasoning. However, these methods either neglect hemodynamic-semantic alignment or compromise geometric fidelity through heuristic approximations.

C. Attention Mechanisms

The development of Attention mechanisms has evolved through three key phases. Channel attention, initiated by SENet's channel-wise recalibration[19] and refined by ECANet's[20] lightweight dynamic convolution, optimizes feature discriminability for low-contrast regions. Spatial attention advanced with RA-Net's[21] residual-guided weighting, critical for handling anatomical variability in CT/MRI. Frequency attention (e.g., FcaNet's DCT-based spectral filtering[22]) enhanced edge preservation in ultrasound and retinal imaging. Hybrid architectures like CBAM[23] synergized multi-dimensional attention, while medical-specific designs—Attention UNet (gated skip connections)[24] and DCSAU-Net (cross-scale fusion)[25]—tailored attention for U-Net-based segmentation.

Ⅲ. METHODOLOGY

A. Overview of DFMS-Net Architecture

Fig. 1 presents the Dual-Field Microvascular Segmentation Network (DFMS-Net), a hemodynamically-consistent inspired architecture that revolutionizes retinal vasculature mapping through synergistic

geometric-functional learning. Building upon FR-UNet [11], our framework consists of four main components: the Encoder module, the Geometric-Field Modeling module, the Functional-Field Optimization module and the Decoder module. The encoder and decoder modules contain three stages of down-sampling and up-sampling layers, implemented by using the standard 2×2 convolution and a 2×2 fractionally-strided convolutions, respectively. Each stage of the encoder and decoder modules comprises multiple convolution blocks. Unlike the FR-Unet, we substitute the modified residual block with a newly designed standard convolutional block, which consists of two 3×3 convolution (Conv) layers, two batch normalization (BN) layers and a LeakyReLU (LR) activation layer. To generate dense predictions, we retain the skip-connection structure, which fuses the feature maps from the previous block with those obtained from the down-sampling and up-sampling layers of adjacent stages.

The Geometric-Field Modeling module as a core part of DFMS-Net, achieving precise vascular anatomical mapping, by combining two collaborating mechanisms——Spatial Path Extractor (SPE) and Transformer-based Topological Interaction (TTI). The SPE module embeds position-sensitive and orientation-specific information along horizontal and vertical axes, maintaining vessel trajectory continuity and modeling orientation priors via direction-aware coordinate attention. Leveraging the learnable attention matrix, the TTI module captures long-remote spatial dependencies between different pixels, enabling precise non-local connection analysis for branch alignment. To merge geometric accuracy with hemodynamic relevance, the SPE-optimized features are sent to the Semantic Attention Amplifier (SAA) in the functional domain optimization module. Here, channel-wise hemodynamic is designed to highlight key semantics and minimizes medical image artifacts. Lastly, we put forward two additional variants with excellent performance.

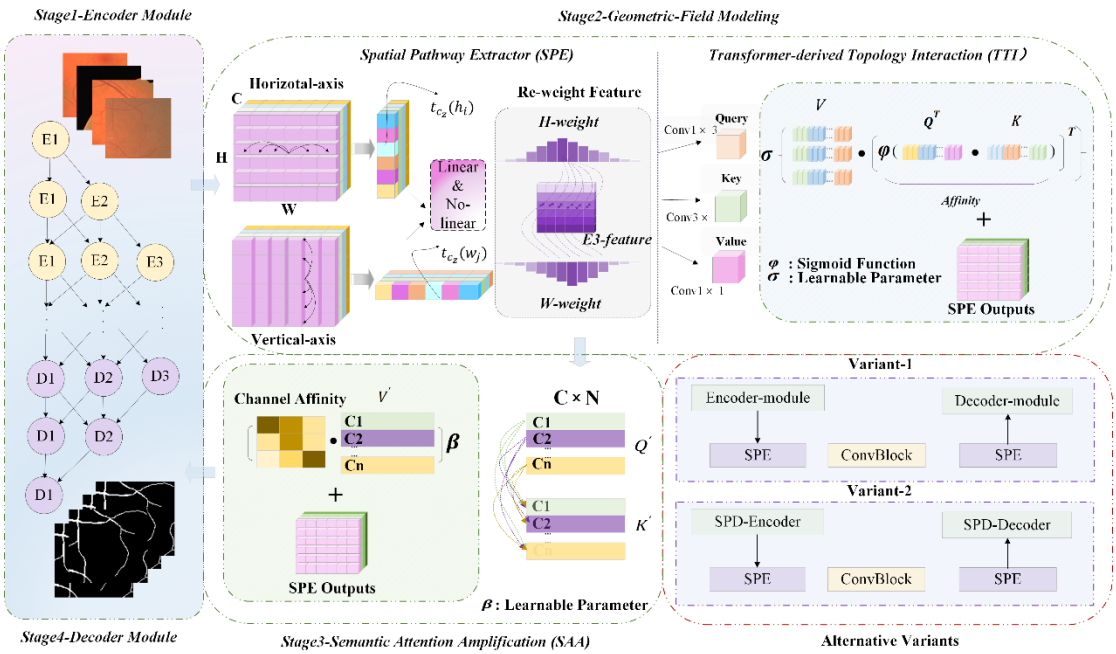

Fig 1. The overview of the proposed DFMS-Net. It consists of four components: the Encoder module, the Geometric-Field Modeling module, the Functional-Field Optimization module and the Decoder module. The SPE module, synergistically integrated with the No-Strided Convolution design, dynamically constitutes two lightweight enhancement variants.

B. Geometric-Field Modeling

Considering the intricate architecture of retinal vessel segmentation—characterized by delicate local capillary structures and highly variable global tree-like morphology—it is imperative to capture comprehensive global contextual features from multi-axis perspectives while preserving vascular trajectory continuity, ensuring both anatomical fidelity for geometric precision and hemodynamic consistency for clinical phenotyping. As demonstrated in many vision tasks[19], [23], [26], [27], spatial and channel information play crucial role to capture structure features and generate precise segmentation maps. Hence, we design a novel Geometric-Field Modeling module that adheres to three key principles. First, it can effectively acquire rich spatial information, enabling the generation of accurate segmentation maps. Second, it must capture long-range dependencies to accurately identify regions of interest while maintaining low memory consumption. Finally, the design should accommodate the slender and variable characteristics of retinal blood vessels. Our proposed geometric-field module is divided into two steps: Spatial Pathway Extractor and Transformer-derived Topology Interaction. These steps will be described in detail in the following sections.

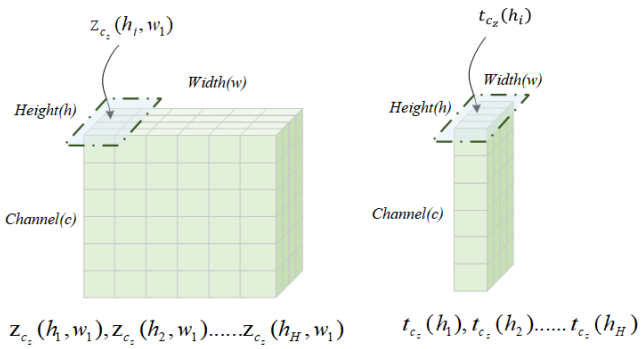

Fig 2. Direction-aware feature map generation in height direction.

1）Spatial Pathway Extractor : Retinal microvasculature exhibits anisotropic tubular geometry, where vessels predominantly align along horizontal or vertical orientations (e.g., major arcades and their bifurcating branches). To encode this anatomically consistent directional prior, the SPE module employs orthogonal spatial pooling along vessel trajectories. Given a feature map $Z \in R^{C\times H\times W}$, which is directly generated from the third layer of the encoder module. Here H, W and C denote the height, width and channel number of the feature map, respectively. We embed directional feature maps using two orthogonal average poling with kernel sizes of (H,1) and (W,1), dependently[27]. As shown in Fig. 2 (taking the vertical direction as an example). These pooling operations simulate the clinical practice of tracing vessel continuity along their dominant orientations. For a vertical vessel segment (Fig. 2), horizontal pooling (kernel (H,1)) aggregates intensity profiles across the vessel's width, suppressing noise while preserving its longitudinal continuity:

$$t_{c_z}(h_i) = \frac{1}{W}\sum_{j=1}^{W} z_c(h_i, w_j) \tag{1}$$

Conversely, vertical pooling (kernel (1,W)) captures cross-sectional variations critical for detecting bifurcation points:

$$t_{c_z}(w_j) = \frac{1}{H}\sum_{i=1}^{H} z_c(h_i, w_j) \tag{2}$$

Here, $h_i(i \in [1, H])$ denotes the height index, $w_j(j \in [1, W])$ denotes the width index, $c_z(z \in [1, C])$ denotes the channel index. The transformed features of the c_z channel at each spatial location index (h_i, w_j) fuse the long-rang global direction-sensitive relationships and spatial location information, which is essential to model complex structure details. The pooled 1D features encode global vascular

geometry but may overlook local texture details (e.g., capillary wall irregularities). To reconcile global continuity with local fidelity, we concatenate the orthogonal embeddings into a composite signal $p_c \in R^{2H\times 1}$, then apply a 1×1 convolutional filter $f^{1\times 1}$ acting as a differentiable vessel edge detector:

$$g = \delta(BN(f^{1\times 1}(p_c))) \tag{3}$$

Here, the hard-swish activation (δ) enhances non-linear separability between vessel pixels and background, while batch normalization (BN) stabilizes training under varying contrast conditions. The split features g_h and g_w represent orthogonal attention cues: g_h emphasizes horizontal vessel continuity (e.g., main arcades parallel to the retina's equatorial plane), g_w **highlights vertical branching patterns** (e.g., perpendicular capillaries extending towards the macula). These directional attentions are combined through element-wise multiplication to geometrically constrain vascular connectivity. Then they are subsequently fed into another two distinct convolutional layers to generate the attention maps. Lastly, the final re-weighted feature can be written as:

$$SPE_{out} = Z \odot S(C_h^{1\times 1}(g_h)) \odot \mathrm{S}(C_h^{1\times 1}(g_w)) \tag{4}$$

Here, SPE_{out} indicates the final output feature maps and $C_h^{1\times 1}$ and $C_w^{1\times 1}$ marks the distinct 1×1 convolutional kernel size. *S* represents the sigmoid gates, which **is used to adjust feature responses according to anatomical plausibility** and to suppress disconnected pseudo-vessels while boosting true bifurcations.

2）Transformer-derived Topology Interaction module：Although the SPE module effectively maintains the continuity of vessel trajectories, capillary connections and branch structures are still prone to fragmentation caused by minor intensity fluctuations and complicated topological interactions among remote vessel branches. This challenge is addressed by the TTI module via modeling non-local spatial dependencies. Specifically, it creates pixel-wise affinity map through interactions between different vascular segment regions, making it possible for the network to: (a) enhance potential bifurcation detection through reinforcing the structural association between main vessels and their branches, (b) reduce false discontinuities via long-distance contextual reasoning along physiologically consistent vascular paths. By emphasizing vessel continuity rather than local morphological appearances, thus the spatial connectivity restores the topology of hemodynamic consistency which is essential for microvascular segmentation and relevant disease analysis. Topological interaction modeling is realized via a differentiable pipeline, consisting of the following phases:

First, we use two standard 3×3 convolutional layers to further encode SPE_{out} to fuse more richer local information and generate the local features $X \in R^{C'\times H'\times W'}$. We then project X into three distinct representations—query Q, key K, and value V—via 1×3, 3×1, and 1×1 convolutional kernels respectively, capturing orientation-specific contextual dependencies. Here, The 1×3 and 3×1 convolutional kernels explicitly model horizontal and vertical vascular trajectory continuity, aligning with the anisotropic structure of retinal microvasculature. To amplify the internal features correlation, they are then immediately flattened into a sequence of size $C' \times N'$, where $N' = H' \times W'$. Then, a matrix multiplication is used to calculate the affinity score between Q and K. The affinity score QK^T quantifies pairwise vascular connectivity likelihood, where higher values indicate anatomically plausible branch linkages (e.g., gradual curvature transitions rather than abrupt termination). Additionally, the softmax function is performed to normalize the affinity score and thus derive the attention weight. Meanwhile, the value matric V is reshaped to $R^{C'\times N'}$ and is multiplied by attention weight using a matrix multiplication to calculate the affinity attention map. Finally, to avoid an overemphasis on specific

elements, the affinity attention map is scaled by a trainable scalar parameter. This adjusted feature maps are then combined with the original input X through a residual connection, thereby boosting the model's capacity to generalization. The calculation method is shown in the following formula:

$$TTI_{output} = \sigma(V \cdot \varphi(Q^T \cdot K)) + X \tag{5}$$

Where, φ denotes the softmax function which is used to compute the attention weight and $\cdot$ is the matrix multiplication. The learnable scalar σ dynamically balances global topology cues against local geometric fidelity, adapting to vessel density variations across retinal regions.

C. Functional-Field Optimization

In retinal images, pathologies such as microaneurysms and hemorrhages are characterized by localized intensity changes, whereas capillaries typically exhibit low-contrast patterns. Deep learning models inherently encode such variations into distinct feature channels, enabling channel attention mechanisms to prioritize clinically salient regions[28], [29]. This selective focus enhances the model's capability to effectively represent and emphasize essential features. Thus, in this section, we further explore the inter-relationships among channel maps using the computed coordinate attention maps. Unlike the SE attention mechanism [19], which relies on scalar operations, our approach employs matrix multiplication to explicitly model the interactions between different channels. To quantify channel-wise semantic synergy, we first project the geometric-refined features $X \in R^{C' \times H' \times W'}$ into query Q', Key K' and Value V' via reshape operations. To obtain the channel affinity score Aff, we perform a multiplication between Q' and the transposed K'. Higher affinity values indicate complementary channel pairs (e.g., microaneurysm channels and their surrounding capillary channels), whereas lower values suggest redundant or contradictory semantics. **To amplify clinically critical channel interactions**, we then apply contrast enhancement by subtracting each affinity value from the channel-wise maximum.

$$Aff = Q' \cdot (K')^T \tag{6}$$

$$Aff_{new} = expand(\max(Aff)) - Aff \tag{7}$$

Where *expand* presents the expand_as function, max presents the max function and $Aff_{new} \in R^{(C',C')}$ presents the resulting new affinity matrix. This operation suppresses dominant but less informative channel correlations (e.g., background-dominated channels), while highlighting subtle yet pathology-indicative relationships. Similar to above section, we then normalize the calculated affinity matrix with softmax function and then it is multiped with V' by a linear matrix multiplication to produce the attention-weighted value matrix. In order to dynamically adjust the weights of the new attention feature map and the geometric-refined feature map X, we perform the following calculation formula:

$$SAA_{out} = \beta(softmax(Aff_{new}) \cdot V') + X \tag{8}$$

Where, SAA_{out} represent the output of channel attention module, β represents the learnable parameter.

D. Dual-Field Synergy Unit

The Dual-Field Synergy Unit (DFSU) integrates geometric continuity with hemodynamic principles, which synthesizes three physiologically grounded feature fields—geometric continuity, topological integrity, and functional saliency—into a hemodynamically consistent representation through a unified fusion paradigm. Given geometric-consistent feature (SPE_{out}), topological (TTI_{output}), and functional (SAA_{out}) from SPE, TTI, and SAA respectively, this operation is mathematically computed via:

$$DFSU_{\text{out}} = SPE_{out} + TTI_{out} + SAA_{out} \tag{9}$$

E. Alternative Variants

While DFMS-Net achieves balanced performance, specialized applications may prioritize topological

fidelity or geometric precision. We thus propose two alternative variants targeting, as illustrated in Fig. 1. Variant 1aim to maximizing vascular network connectivity for flow simulation tasks. Variant 2 is designed for preserving capillary boundary geometry for micron-scale morphometry. Both variants inherit DFMS-Net's core SPE modules but diverge in architectural inductive biases—Variant 1 through symmetry, Variant 2 via SPD—to address orthogonal challenges in vascular structure recovery. Detailed structural descriptions of the two variants are provided below:

1) Variant 1: The architecture comprises encoder-decoder backbone which inherits the lightweight structure of DFMS-Net, ensuring compatibility with existing workflows. This variant symmetrically deploys SPE modules in both encoder and decoder stages, enforcing consistency in vessel trajectory learning. Encoder-SPE extracts horizontal/vertical vascular pathways via direction-aware pooling, encoding geometric continuity constraints, while Decoder-SPE reuses SPE to refine upsampled features, ensuring coherence between coarse and fine scales. The middle ConvBlock with two 3×3 convolutions and LeakyReLU, captures neighborhood-level vessel interactions, bridging global topology and local geometry. With 5.73M parameters (vs. 5.76M for DFMS-Net), Variant 1 achieves superior connectivity through its inductive bias toward vascular symmetry, demonstrating that performance gains need not come at the cost of increased model complexity.

2) Variant 2: To address the vanishing capillary problem in low-resolution fundus images, this variant introduces a non-destructive downsampling module (SPD-Conv) [30], replacing traditional strided convolutions and pooling layers. With 6.07M parameters (vs. 5.76M for DFMS-Net), Variant 2 prioritizes geometric fidelity over parameter efficiency, achieving sub-pixel boundary precision critical for capillary morphometry. As illustrated in Fig. 3, the architecture comprises:

Firstly, A Conv1×1 layer adjusts channel dimensions without spatial compression, preserving boundary voxels typically lost in strided convolution. Secondly, the feature map $Y \in R^{(D,S,S)}$ is decomposed into four sub-features via stride-2 slicing:

$$M_0 = Y[\ldots, :: 2, :: 2]$$
$$M_1 = Y[\ldots, 1 :: 2, :: 2]$$
$$M_2 = Y[\ldots, :: 2, 1 :: 2]$$
$$M_3 = Y[\ldots, 1 :: 2, 1 :: 2]$$

Then, sub-features are concatenated along the channel dimension, yielding a feature map of size $R^{(4D,\frac{S}{2},\frac{S}{2})}$. Finally, a conv3×3 layer is used to reduce channels while maintaining boundary granularity.

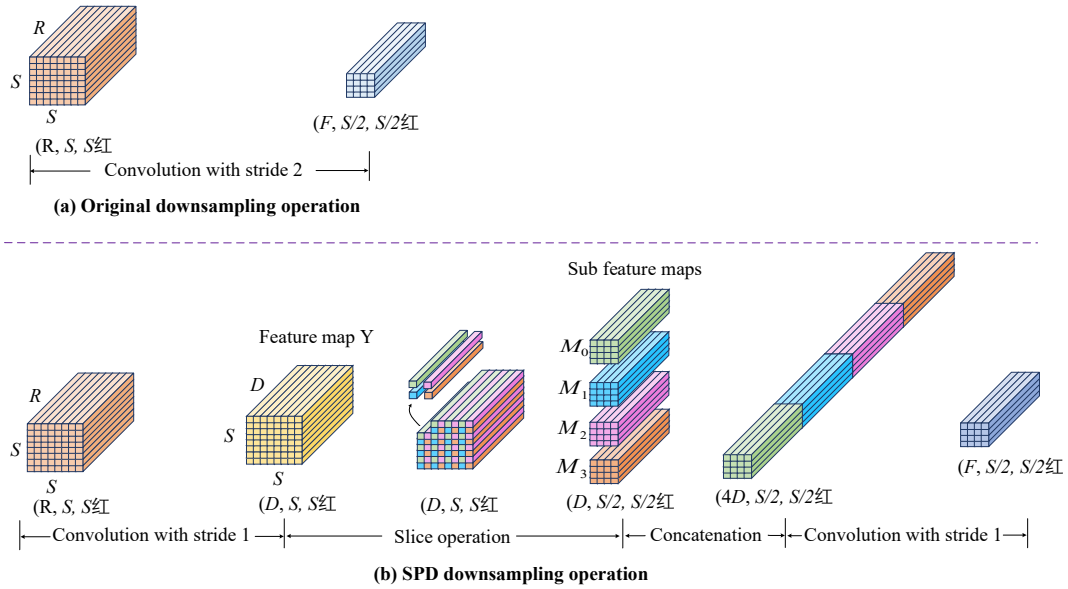

Fig. 3. SPD-down module. (a) is our original operation with stride 2. (b) is the new designed SPD downsampling operation using slicing strategy. We firstly use the convolution layer with stride 1 to change channel numbers, then utilize the slice operation to obtain four sub feature maps. Finally, the convolution layer is reused to decrease channel numbers.

Ⅳ. Experiments

A. DATASETS

We evaluate our method on five public datasets with varying image resolutions and sizes, including DRIVE [31], STARE [32], CHASE_DB1 [33], DCA1 [34] and CHUAC [33]. The details are as follows:

1) DRIVE Dataset: The DRIVE dataset contains 40 fundus images, each measuring 584 ×565 pixels, evenly divided into 20 training images and 20 testing images. These images were captured using Canon CR5 non-mydriatic 3CCD camera, including 7 with abnormal pathology and 33 from health subjects. Each image includes two manual annotations provided by two different observers, and we use the first annotation as the ground truth.
2) STARE Dataset: The STARE dataset consists of 20 colored retinal images, each with dimension of 700 × 605 pixels. In alignment with SCS-Net[35], we adopt a 10-fold cross-validation method to split the training and testing sets.
3) CHASE_DB1 Dataset: The CHASE_DB1 dataset includes 28 fundus images collected from 14 schoolchildren, each with a resolution of 999×960 pixels. Consistent with recent studies [3], [36], we designate the first 20 images for training and the remaining 8 for testing.
4) DCA1 Dataset: The DCA1 dataset contains 130 X-ray coronary angiograms obtained from the Mexican Social Security Institute. Each image, manually annotated by an expert cardiologist, has a resolution of 300 × 300 pixels in PGM grayscale format. The database is divided into 100 images for training and 30 for testing.
5) CHUAC Dataset: The CHUAC dataset consists of 30 coronary angiography images, each originally measuring 189 × 189 pixels, with ground truth masks of 512 × 512 pixel. These masks were manually created by doctors using an image editing tool. As in previous work, the first 20 images

are used for training and the remaining 10 for testing. To align the dimensions of the images with their ground truth, we resized the angiography images to 512 × 512 pixels using bilinear interpolation.

B. Implementation Details and Evaluation Metrics

The proposed framework is implemented in PyTorch 2.0.1 on an NVIDIA A100 GPU. We employ the Adam optimizer to train the networks, with the learning rate and weight decay set to 0.0001 and 0.00001, respectively. Moreover, we utilize binary cross-entropy (BCE) as the loss function and adopt a cosine annealing learning rate scheduler. Specifically, to conserve computational resources and augment the data, we extract image patches with a 48 × 48 sliding window using the unfold function. In this study, we evaluate the model's segmentation performance by comparing the segmentation results to the ground truths using several widely used metrics, including the intersection over union (IoU), accuracy (Acc), sensitivity (Sen), specificity (Spe), F1 score (F1) and area under the curve (AUC). These criteria are defined as follows:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad Sen = \frac{TP}{TP+FN} \quad Spe = \frac{TN}{TN+FP} \quad F1 = \frac{2TP}{2TP+FP+FN} \quad IoU = \frac{TP}{TP+FP+FN} \quad (10)$$

For medical segmentation maps, pixels are divided into two categories: vascular pixels (those with blood flow) and nonvascular pixels (those without blood flow). The classification of each image results in four potential outcomes: vascular pixels correctly identified as vessels (True Positive, TP), non-vascular pixels correctly identified as non-vessels (True Negative, TN), nonvascular pixels incorrectly predicted as vessels (False Positive, FP), vascular pixels incorrectly predicted as non-vessels (False Negative, FN).

C. Comparative Experimental Studies

To evaluate the performance of our designed DFMS-Net, we conduct series experiments against state-of-the-art methods, including TransUNet [12], DE-DCGCN-EE [10], PLVS-Net [12], DPF-Net [12], MedSegDiff [15], GT-DLA-dsHFF [11] and DPL-GFT-EFA [14]. Also, we compare our framework with classic methods, such as U-Net [1], UNet++ [2] and Attention U-Net [3]. To demonstrate the advantages of our approach more effectively, we conduct a comprehensive evaluation incorporating both qualitative and quantitative analyses across five public benchmarks. Quantitative comparisons on the DRIVE, STARE and CHASE_DB1 fundus datasets are summarized in Table I and II. Cross-domain generalization capabilities are assessed through the DCA1 and CHUAC coronary angiography benchmarks, with detailed metrics presented in Table III.

TABLE I

COMPARISON WITH THE STATE-OF-THE-ART METHODS ON DRIVE AND STARE

Methods	DRIVE						STARE					
	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)
U-Net[9]	96.78	80.57	98.33	98.25	81.41	68.64	97.30	78.50	98.84	98.75	81.18	68.56
UNet++ [10]	96.79	78.91	98.50	98.25	81.14	68.27	97.34	79.09	98.83	98.84	81.50	69.02
Attention U-Ne t[24]	96.62	79.06	98.31	97.74	80.39	67.21	97.30	78.04	98.87	98.78	81.06	68.39
TransUNet [37]	96.90	82.64	98.27	98.38	82.35	70.30	97.48	83.32	98.62	98.68	83.25	71.45
DE-DCGCN-EE [38]	97.05	83.59	98.26	98.66	82.88	70.77	97.51	84.05	98.61	98.99	83.63	71.87
PLVS-Net [39]	96.79	81.95	98.19	98.32	81.56	68.89	97.28	83.08	98.43	98.71	82.20	69.87
DPF-Net [40]	96.96	83.05	98.30	98.53	82.70	70.54	97.44	83.60	98.56	98.76	83.04	71.17
MedSegDiff [41]	95.61	78.00	97.31	94.32	75.65	60.86	91.22	70.56	92.83	88.42	66.43	50.80
GT-DLA-dsHFF [42]	97.00	**83.77**	98.27	98.66	82.99	70.96	97.60	84.88	98.65	99.13	84.32	72.89
DPL-GFT-EFA [43]	97.05	83.61	98.34	98.69	**83.22**	71.03	97.70	85.31	98.69	99.17	84.74	73.64
Baseline*	97.05	83.13	98.40	98.87	83.07	71.06	97.65	87.39	98.67	99.42	86.75	76.70

Variant-1	97.06	82.77	98.46	98.88	83.09	71.10	97.68	80.61	98.85	99.43	86.73	76.66
Variant-2	97.04	83.30	98.38	98.87	83.06	71.06	97.71	85.55	**98.92**	99.44	86.83	76.80
DFMS-Net	**97.08**	82.37	**98.52**	**98.89**	83.11	**71.12**	**97.72**	**88.01**	98.68	**99.47**	**87.18**	**77.35**

The bold fonts represent the best results and the * denotes the retrained baseline results.

1) Evaluation on the DRIVE Dataset: Our proposed DFMS-Net attains state-of-the-art performance on the DRIVE benchmark, showing a superior balance across key metrics. In TABLE **I**, our framework ranks first among all methods—accuracy (ACC) of 97.08%, specificity (Spe) of 98.52%, AUC of 98.89%, and IoU of 71.12%, exceling in suppressing false positives while keeping vascular continuity. Although GT-DLA-dsHFF [42] has a slightly higher sensitivity (Sen: 83.77% vs. 82.37%), the spatial redundancy introduced by hierarchical attention fusion, leads to a drop in specificity (98.27% vs. 98.52%) and IoU (70.96% vs. 71.12%). The problem is solved by DFMS-Net through the special designed **DFSU strategy, which** incorporates the SPE module to ensure vessel trajectory smoothness and decrease fragmentation in low-contrast regions (Spe↑ 0.25% vs. GT-DLA-dsHFF). To solve the forking ambiguity problem, the designed TTI leverages adaptive affinity learning, obtaining the highest IoU of 71.12% and AUC of 98.89%. It should be highlighted that DFMS-Net surpasses the transformer-based TransUNet[37] by 0.82% and the diffusion - driven MedSegDiff[41] by 10.26% in IoU, which verifies its robustness to architecture-bias. While DPL-GFT-EFA [43] shows a slightly higher F1 score than DFMS-Net (83.22% vs. 83.11%), its graph-based fusion has problems with capillary boundary stability (AUC: 98.69% vs. 98.89%).

Meanwhile, the designed Variant 1 reaches almost the same level of performance with minimal reduction in parameters (IoU: 71.10%, Spe: 98.46%), by prioritizing symmetric - driven coherence Variant 2, through boundary-centric SPD-Conv, enhances thin-vessel sensitivity (Sen: 83.30%) and outperforms PLVS-Net[39] by 1.71% in sensitivity, while keeping a robust IoU (71.06%). These outcomes validate DFMS-Net's effectiveness in geometric precision, hemodynamics - related correlation.

2) Evaluation on the STARE Dataset: On STARE, DFMS-Net surpasses current methods with a 0.62% higher sensitivity (88.01%) and a 0.65% higher IoU (77.35%), achieving a new top - level performance, with specificity (98.68%) and AUC (99.47%), which proves its diagnostic reliability. This breakthrough illustrates the framework's ability in solving two lasting challenges in retinal analysis challenges: (a) the detection of low-contrast vessels and (b) the maintenance of peripheral vascular continuity. Although DPL-GFT-EFA[43] shows near-perfect discrimination (AUC 99.17%), its probability-based fusion strategy cannot capture the inherent anisotropic geometric prior in retinal vessels, leading to fragmented output in the macular region (F1: 84.74%). GT-DLA-dsHFF[42] achieves a high Sen score of 84.88%. However, its hierarchical attention blocks over-smooth fine vessels, leading to a 4.46% lower IoU than DFMS-Net. TransUNet[37] shows worse performance in Spe (98.62% vs. 98.68%) and F1 (83.25% vs. 87.18%) due to the lack of directional priors. In comparison, variant 1 effectively balances IoU and Spe, while variant 2 shows promising performance in boundary-critical situations, realizing 85.55% Sen (↑1.5% over DE-DCGCN-EE) using SPD-Conv's spatial fidelity.

TABLE II

COMPARISON WITH THE STATE-OF-THE-ART METHODS ON CHASEC_DB1 DATASETS

Methods	CHASE_DB1					
	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)
U-Net [9]	97.43	76.50	98.84	98.36	78.98	65.26
UNet++ [10]	97.39	83.57	98.32	98.81	80.15	66.88
Attention U-Net [24]	97.30	83.84	98.20	98.48	79.64	66.17
TransUNet [37]	97.55	82.22	98.58	98.73	80.92	67.98
DE-DCGCN-EE [38]	**97.62**	84.00	98.56	98.98	**82.61**	**70.37**
PLVS-Net [39]	97.42	81.89	98.46	98.60	80.05	66.75
DPF-Net [40]	97.53	83.12	98.50	98.59	80.95	68.03
MedSegDiff [41]	95.98	78.57	97.15	93.60	70.93	55.01
GT-DLA-dsHFF [42]	97.57	83.39	98.52	98.83	81.20	68.35
DPL-GFT-EFA [43]	97.71	79.60	**98.93**	99.01	81.45	68.74
Baseline*	97.39	89.06	97.97	**99.12**	81.14	68.29
Variant-1	97.48	86.96	98.20	98.96	81.27	68.50
Variant-2	97.52	87.07	98.23	99.08	81.54	68.86
DFMS-Net	97.33	**89.29**	97.88	99.11	80.84	67.87

The bold fonts represent the best results and the * denotes the retrained baseline results.

3) Evaluation on the CHASE_DB1 Dataset: On CHASE_DB1, our proposed method also shows its prioritized performance in detecting tree-critical capillaries, a vital capability for early-stage ocular lesion regions. It reaches 89.29% Sen, 5.29% higher than DE-DCGCN-EE[38], while maintaining a competitive AUC of 99.11% in TABLE II. DE-DCGCN-EE[38] displays superior structural metrics with 70.37% IoU and 82.61% F1 using edge-optimized graph convolutions. Nonetheless, its Sen metric decreased by 5.29%, reflecting its inability to address low-signal vascular features. By contrast, DPL-GFT-EFA[43] focuses on non-vascular specificity (Spe: 98.93%) via its diffusion probabilistic learning module, but obtains lower Sen, indicating that excessive noise-related information suppresses the detection of real vessels. However, our Variant 1 provides a simplified choice for high-specificity scenarios, where suppressing background pixels is critical, achieving 98.20% Spe. It is worth mentioning that Variant 2 provides a balanced performance in boundary accuracy and sensitivity. Using the spatial fidelity of SPD-Conv, it outperforms DPL-GFT-EFA by 7.47% in Sen.

As shown in Fig. 4, our architecture demonstrates excellent segmentation capabilities in preserving topology-structural features in medical image segmentation, especially for complex scenarios. Here, the x-axis represents various methods from Tables I and II, which are encoded as 1, 2, etc. from top to bottom, while the y-axis denotes the AUC and IoU metrics. This chart intuitively presents our experimental results on three distinct datasets, revealing that the designed approach offers the most effective segmentation.

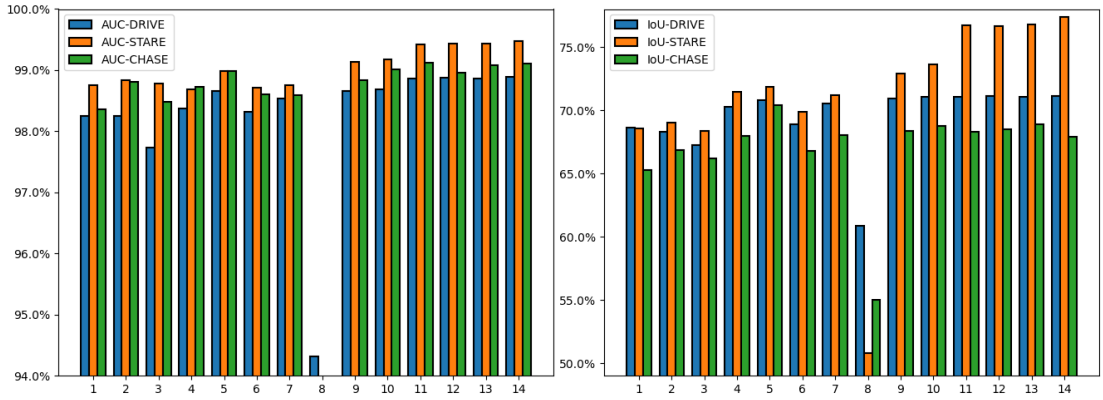

Fig. 4. Bar chart comparison results on AUC and IoU metrics across DRIVE, STARE and CHASE_DB1 datasets. The x-axis represents the methods in Table I-II and they are encoded as 1, 2, 3, etc. from top to bottom. The y-axis represents the AUC or IoU performance on different datasets.

4) Evaluation on the DCA1 and CHUAC Dataset: In the field of coronary angiography, the proposed DFMS-Net shows unprecedented consistency, which is evidenced by its remarkable performance in the dynamic coronary angiography (DCA1) and complex microvasculature (CHUAC) benchmarks in Table III. The dual-domain advantage confirms its capacity to tackle two widespread challenges in medical image segmentation—detection of artifacts vessels and preservation of continuity in low-signal objects objective pixel points. DFMS-Net reaches higher Spe in both situations: 98.91% for DCA1 and 99.60% for CHUAC. Its DFSU approach enables it to outperform all competitors by at least 0.15%. Moreover, DFMS-Net focuses on the retention of vascular topology and enhancement of critical-vessel intensity, through dynamically adjusting the weights of geometric and branched structural features, achieving 99.19% AUC. It effectively suppresses noise interference using DFSU's phase-aware attention mechanism, which ensures spatio-semantic synchronization (Sensitivity: 82.12% vs. UNet++'s 66.87% on CHUAC).

Variant 1 provides deployment efficiency: Maintains 97.81% accuracy (ACC) on DCA1, an increase of 0.18% compared to the baseline, with 5.73 million parameters, making it ideal for real-time interventions. Variant 2 excels in boundary-critical tasks: DCA1 achieves state-of-the-art Intersection over Union (IoU) of 66.53% using SPD-Conv's non-destructive downsampling. CHUAC attains the highest sensitivity (Sen) of 83.65% by preserving capillary endpoints in low-SNR regions.

TABLE III

COMPARISON WITH THE STATE-OF-THE-ART METHODS ON DCA1 AND CHUAC DATASETS

Methods	DCA1						CHUAC					
	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)
U-Net [9]	97.58	78.16	98.66	98.79	77.35	63.07	97.84	58.81	99.40	96.14	67.68	51.15
UNet++ [10]	97.61	79.54	98.62	98.84	77.86	63.75	**98.12**	66.87	99.37	97.07	73.23	57.77
Attention U-net[24]	97.55	79.86	98.53	98.55	77.48	63.24	98.00	65.26	99.13	95.06	71.54	55.69
CS-Net[44]	97.63	78.95	98.67	98.89	77.90	63.80	97.96	67.35	99.18	97.47	71.71	55.89
HRNet [45]	97.77	80.07	98.76	98.99	79.19	65.54	98.11	74.56	99.06	98.70	75.26	60.33
Baseline*	97.63	**86.19**	98.28	99.22	79.12	65.58	98.03	81.71	98.68	**99.08**	**76.01**	**61.51**
Variant-1	**97.81**	82.75	98.66	99.22	79.76	66.44	98.00	81.76	98.65	99.06	75.73	61.16
Variant-2	**97.81**	82.77	98.66	**99.23**	**79.81**	**66.53**	97.80	**83.65**	98.37	99.00	74.47	59.55
DFMS-Net	97.77	77.79	**98.91**	99.19	78.30	64.60	97.48	82.12	**99.60**	98.04	74.30	61.20

The bold fonts represent the best results.

5) **Visualization segmentation results on five datasets:** The segmentation results visualized in Fig. 5 shows the dual-field synergy of DFMS-Net in addressing cross-domain vascular complexity. Due to space limitations, we focus on comparing the baseline network with our proposed methods. In retinal microvasculature (DRIVE/CHASE_DB1/STARE), the baseline method fragments delicate capillary networks, particularly in low-contrast peripheral regions, as seen in discontinuous branching patterns marked by green indicator arrow located at the 2nd and 5th rows. This arises from its inability to model anatomic continuity under hemodynamic variations. DFMS-Net resolves this through its Spatial Pathway Extractor (SPE), which enforces trajectory-aware attention, bridging gaps between capillary junctions (red circle in line 1) while suppressing noise from non-vessel pixels (blue circle in line 2).

The visualization outcomes reaffirm DFMS-Net's hemodynamically-consistent design philosophy on cross-domain DCA1 and CHUAC datasets. In the context of coronary angiography, the Semantic Attention Amplification (SAA) module enhances true capillaries over extraneous background textures, as evidenced by clear segmentations around vessel branches (red circle, row 5). Our designed Dual-Field Synergy Unit (DFSU) effectively balances boundary sharpness and noise interference, enabling the differentiation between pseudo-vessels and true vessels-a critical capability where FR-Unet fails due to its static feature aggregation, often leads to the misclassification of background pixels as valid branches (blue circle, row 4).

In line with aforementioned results, the introduced variants show promising optimization. By utilizing the SPE method, Variant 1 achieves high-quality segmentation, preserving complex coronary detail detection and significantly improving computing efficiency. At the same time, Variant 2 excels in the detection of low-SNR angiographic sequences due to the application of SPD-Conv, as shown by the purple circle and green arrow in line 5.

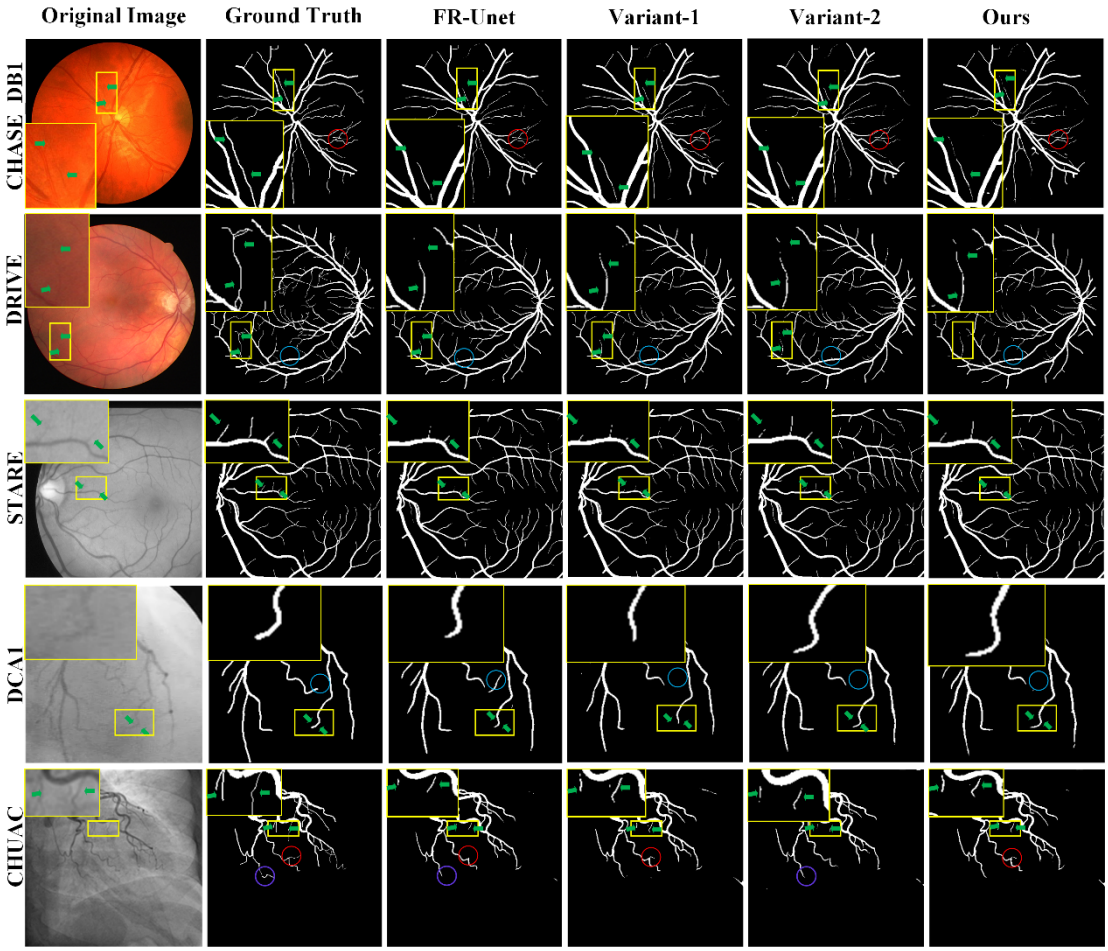

Fig. 5. Visualization segmentation results on five datasets. The yellow-boxed areas identify instances where the proposed methods excel over FR-Unet method, while arrows and circles indicate significant difference between different methods.

D. Ablation Study

To rigorously dissect the contributions of DFMS-Net's dual-field paradigm, we perform systematic ablation studies on DRIVE and STARE—datasets representing divergent vascular phenotypes (retinal microvasculature and complex bifurcations). We systematically disable individual modules (SPE/TTI/SAA), with configuration setting detailed in Table IV. Tables V-VI present comprehensive quantitative metrics, while Fig. 6-7 anatomically dissect performance gains through multi-type visualizations.

1) **Network Configurations:** To deconstruct the contributions of DFMS-Net's dual-field components, we engineer ten progressively enriched architectures atop the FR-U-Net baseline, systematically integrating modules from the geometric (SPE, TTI) and functional (SAA) fields. Configurations are codified in Table IV, with key variants categorized as:

(a) Baseline: FR-Unet network.

(b) Baseline + SPE×2 (Variant 1): The architecture includes two SPE modules based on baseline network.

(c) Baseline + SPE+ SAA: It means that the architecture is composed of one SPE module and one SAA module based on baseline network.

(d) Baseline + SPE×2 + SPD_down (Variant 2): It means that the architecture is composed of two SPE modules and one SPD_ down module based on baseline network.
(e) Baseline + TTI / SAA: It represents that the TTI and the SAA modules are paralleled designed based on the baseline network.
(f)Baseline + SPD_down _SPE + TTI/SAA: It represents that the network consists of SPD_down, SPE, TTI and SAA components based on the baseline network.
(g) Baseline + Down _SPE + TTI/ SAA: It presents that the network consists of Down _SPE, TTI and SAA components based on the baseline network.
(h) Baseline + Out_SPE+ TTI/ SAA: It represents that the network consists of Out_SPE, TTI and SAA components based on the baseline network.
(i) Baseline + SPE×2 + SPD_down + Out_weight: It represents that the network consists of two SPE, one SPD_ down and one Out_weight modules based on the baseline network.
(j) Baseline + SPE + TTI / SAA (DFMS-Net.): It represents that the network consists of SPE, TTI and SAA modules based on the baseline network.

TABLE IV**.** List of abbreviations that used in ablation studies

Notations	Explanation
/	Parallel connection
SPD_down	The strided convolution is replaced by non-strided convolution
SAA	The channel attention of stage 3
Down_SPE	The SPE module is concatenated to the normal down-sampling operation.
Out_weight	Outputs are summed with weights
SPD_ down _SPE	The SPE module is concatenated to the SPD_down down-sampling operation.
Out_SPE	The SPE module is concatenated to the output layer.

2) **Effect of the SPE module:** The SPE module uniquely integrates geometric accuracy and hemodynamics-relatedness through orientation-aware coordinate attention, as shown by the quantitative results in TABLE Ⅴ and visualization attention map (Fig. 7). On DRIVE, network (j) integrated SPE module boosts sensitivity by 0.29% and IoU by 0.04%, comparing to the baseline network (network (e)) using only TTI/SAA. This is a statistically meaningful improvement considering the special complexity of microvascular. These improvements are intuitively shown by the yellow box in Fig. 6, network (j) significantly reduces misclassified red pixels (false negatives), comparing to network (e), and recovers fine vessels usually lost from background noise. Visualization results in Fig. 7 further explains these improvements, network (j) not only focuses on thick vessels (column 5), but even accurately highlights surrounding capillaries (blue halos). Conversely, network (e) places too much attention on the edges of thick vessels (brighter activation), but neglects capillary branches, which explains its lower sensitivity.

The results on STARE dataset further evident the capability of SPE, since network (j) obtains an improvement of 2.04% over network (e) on Sen, validating its superior performance in identifying thin-vessels branches. Remarkably, the inherent limitations of TTI/SAA, which excel in pure global contextual aggregation yet fall short in maintaining hemodynamic-geometric consistency, are addressed by SPE. Moreover, experimental results indicate that SPE exhibits enhanced geometric preservation within the encoding layer relative to the decoding layer, achieving 83.11% on DRIVE with 87.18% F1 on STARE (network (h) vs. (j)).

TABLE V

ABLATION EXPERIMENTS ON THE DRIVE AND STARE DATASETS.

Methods	DRIVE						STARE					
	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)
(a) Baseline	97.05	83.13	98.40	98.87	83.07	71.06	97.65	87.39	98.67	99.42	86.75	76.70
(c) Baseline + SPE+ SAA	96.67	88.57	97.46	98.86	82.25	69.87	96.77	74.93	98.90	98.32	80.15	66.94
(e) Baseline + TTI/ SAA	97.09	82.08	98.55	98.87	83.07	71.08	97.70	85.97	98.87	99.43	86.80	76.77
(h) Baseline + Out_SPE+ TTI/ SAA	96.99	83.19	98.33	98.53	82.78	70.65	97.60	85.38	98.83	99.23	86.25	75.94
(j) Baseline + SPE+ TTI/ SAA	97.08	82.37	98.52	98.89	83.11	71.12	97.72	88.01	98.68	99.47	87.18	77.35

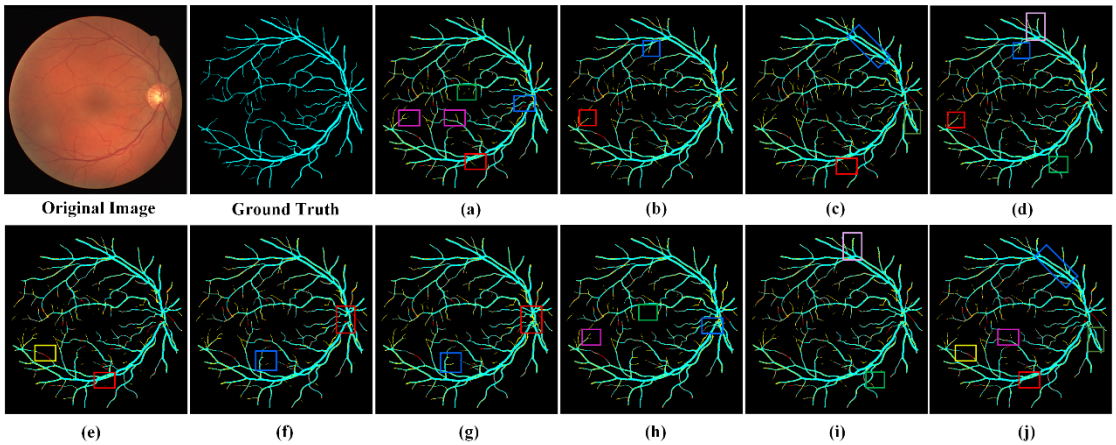

Fig. 6. Visualized colored segmentation results of ablation study on DRIVE dataset of network (a)- (j). The fluorescent blue pixels present the correctly segmented vessel pixels, the red and yellow pixels present the incorrectly vessel pixels and background pixels, respectively.

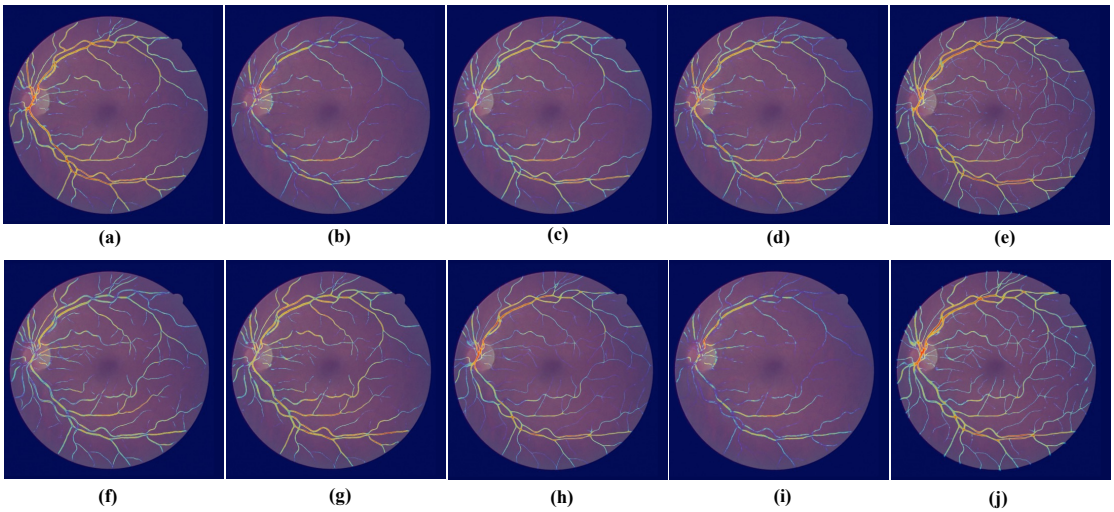

Fig. 7. Visualization of attention maps using Grad-CAM of ablation study on DRIVE dataset of network (a)- (j).

3) **Effect of the TTI module:** On the DRIVE dataset, the TTI module suppresses edge ambiguity in thick vessels (blue box, Fig. 6, network (j) vs. (c)) by modeling spatial connectivity priors, reducing overlapping boundaries (green box) and achieving a **1.25% IoU gain** (71.12% vs. 69.87%, TABLE **V**). Despite a marginal sensitivity trade-off (82.37% vs. 88.57%), validating its hemodynamically-driven balance between geometric precision and diagnostic fidelity. The visualization results of Fig.

7 reveals that network (c) overemphasizes thick vessel edges, indicated by bright peripheral activations, while losing capillary continuity, shown by fragmented halos, thus confirming its hemodynamic inconsistency.

On the STARE dataset, the TTI module's topology-aware affinity learning achieves an **88.01% sensitivity (Sen)—13.08% higher** than the SPE/SAA baseline (network (c))—by recovering **more discontinuous bifurcations** that conventional methods fragment due to noise interference or low contrast. This surge in Sen directly correlates with a **10.41% IoU gain** (77.35% vs. 66.94%), as TTI's spatial connectivity analysis enforces hemodynamic consistency across branching hierarchies, transforming scattered capillary activations into structural coherent vascular trees (network (j)).

4) **Effect of the SPD_down module:** Experiments across six topological configurations (b)-(d) and (g)-(f) in Table Ⅵ, reveals the SPD_down module's critical advantages. The IoU metric demonstrates stability on DRIVE (71.10% → 71.06%) (red box, Fig. 6, comparing network (b) with (d)), but shows an increase on STARE (76.66% → 76.80%), indicating improved adaptability to intricate vascular networks. Meanwhile, the designed module indicates enhanced detection of capillaries without increasing false positives on STARE (Spe: 98.85% → 98.92%, Sen: 80.61%→ 85.55%). Similarly, network (f) demonstrates a substantial improvement in sensitivity (Sen) over (g), showcasing superior detection capability for low-contrast lesions such as microaneurysms. As illustrated in Fig. 6, the SPD-integrated method notably enhances the focus on small vessel pixels while effectively reducing the misclassification of non-vessel pixels as blood vessels (red and blue boxes, second row).

TABLE VI

ABLATION EXPERIMENTS OF THE TTI MODULE and output strategy ON THE DRIVE AND STARE DATASETS

Methods	DRIVE						STARE					
	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)
(b) Baseline + SPE×2	97.06	82.77	98.46	98.88	83.09	71.10	97.68	80.61	98.85	99.43	86.73	76.66
(d) Baseline + SPE×2 + SPD_down	97.04	83.33	98.38	98.87	83.06	71.06	97.71	85.55	98.92	99.44	86.83	76.80
(f) Baseline + SPD_down _ SPE + TTI/ SAA	97.05	83.66	98.35	98.90	83.15	71.18	97.62	88.86	98.49	99.42	86.80	71.76
(g)Baseline + Down _SPE + TTI/ SAA	97.11	80.69	98.71	98.89	82.93	70.89	94.71	82.32	99.82	94.44	86.93	71.41
(i)Baseline + SPE×2 + SPD_down + Out_weight	96.58	89.32	97.29	99.87	81.99	69.50	96.79	72.94	99.11	98.39	79.82	66.50

5) **Effect of the output strategy module:** To address the inherent class imbalance in retinal vasculature segmentation (vessel-to-background pixel), we rigorously evaluate two hierarchical fusion strategies: Deep Supervision with weighted summation (network (i)) aggregates features from the last five layers using learnable weights, and average fusion (network (d)) computes element-wise averaging across multi-outputs. On DRIVE Dataset, Weighted summation achieves superior sensitivity (89.32% vs. 83.33% in network (d)) but compromises specificity (97.29% vs. 98.38%) and overall accuracy (96.58% vs. 97.04%), indicating overfitting to minority-class vessel pixels. On STARE Dataset, average fusion dominates across all metrics (ACC: 97.71% vs. 96.79%, Sen: 85.55% vs. 72.94%, AUC: 99.44% vs. 98.39%), demonstrating better generalization on complex vascular networks.

The limited capability of deep supervision stems from intrinsic gradient dissonance. During

backpropagation, the layer-wise parameter magnifies the slight feature differences in low-contrast regions. Conversely, average fusion method keeps angiographic coherence, which is highlighted by the higher intensity in the same spatial area (network (d) vs. (i), Fig. 6).

6) **Statistical Analysis:** In order to strictly quantify the segmentation performance in vessel classes, we employ six-number summary box plots (minimum, Q1, median, Q3, mean, maximum) to perform non-parametric analysis the distributions of TP, FP, and FN on the DRIVE dataset in Fig. 8. It should be noted that TN is excluded to ensure interpretable visualization of critical error types due to extreme class imbalance. Regarding the TP metric, DFMS-Net demonstrates highest statistics metric score. This is attributed to its remarkable capability in maintaining vascular continuity through geometric topology-aware consistency (SPE) and hemodynamic interaction (TTI). Meanwhile, the lowest FN score suggests that the SAA module considerably alleviates capillary omissions in low-contrast areas by suppressing non-hemodynamic artifacts.

In comparison to network (a), network (d) accomplishes lower FN and higher TP across nearly all statistical criteria, further validating the superiority of SPE in the accurate segmentation of vascular pixels, by resolving ambiguous pixel affiliations at capillary junctions, thereby reducing misclassification drift. Fig. 8 also demonstrates that network (b)/(e)/(g) achieve lower FP, which is corresponding to their overemphasis on shape specificity leading to disjointed blood vessels.

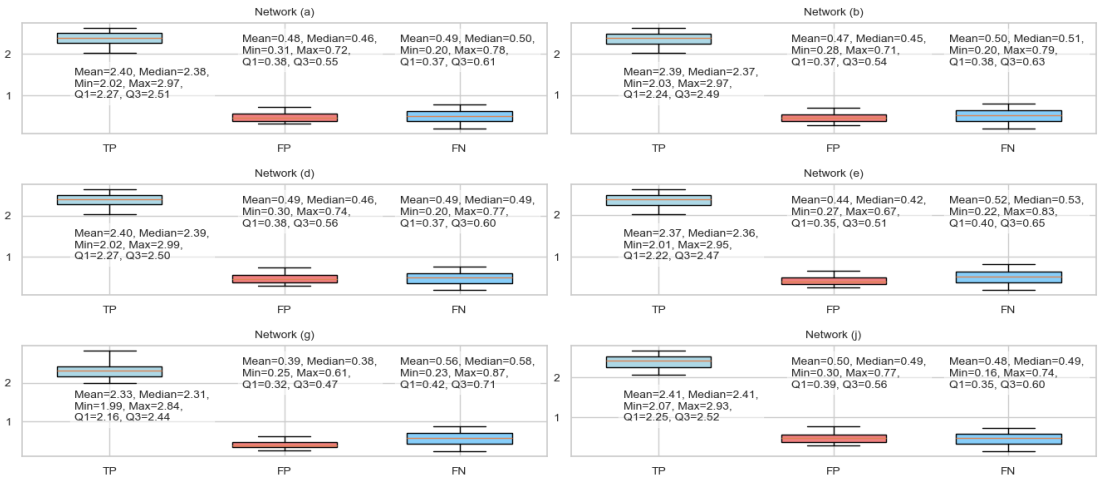

Fig. 8. Static analysis on DRIVE dataset of TP, FP, FN metrics for different methods. The y-axis represents the pixel number of the image (in ten thousand) and the x-axis denotes the evaluation metrics.

VI. DISCUSSIONS AND CONCLUSIONS

The DFMS-Net framework redefines retinal vessel segmentation by resolving the longstanding dilemmas: the geometric-functional disconnect in microvascular feature representation. Unlike conventional approaches that treat anatomical structures and hemodynamic patterns as separate entities, our DFMS-Net method based on **Dual-Field Synergy Unit (DFSU)** module, establishes a closed-loop interaction between vascular topology (via SPE/TTI) and functional saliency (via SAA). This biologically inspired design mirrors the ophthalmologist's diagnostic workflow—simultaneously tracing vessel morphology while assessing perfusion relevance—thereby achieving

unprecedented precision in capillary junction preservation.

Building upon the rapid evolution of foundation models like Segment Anything Model (SAM), our future research will pioneer two transformative directions: 1) **SAM-DFMS Synergy for Medical Vision**. Developing hybrid architectures integrating SAM's universal segmentation priors with DFMS-Net's hemodynamic awareness, enabling few-shot adaptation to rare vascular pathologies (e.g., retinal vein occlusion). Implementing cross-modal alignment strategies to jointly process fundus images, addressing the "modality gap" in microvascular analytics. 2) **Clinical Translation via Quantitative Phenotyping**: Deploying task-specific to predict cardiovascular risk indices directly from retinal vasculature segmentation maps. This dual-axis evolution—combining foundation model scalability with clinic-centric quantification—aims to bridge the current chasm between computational research and bedside microvascular diagnostics.

ACKNOWLEDGMENT

This work was supported by the Science and Technology Development Fund of Macao under Grant No. 0002/2024/RIA1, and by Fujian Province Young and Middle-aged Teachers' Educational Science Research Projects (Science and Technology Category) under Grant No. JAT232014, and by Research Projects of Putian University 2023036 and 2022047, and by Putian City Science and Technology Bureau under Grants 2022SZ3001ptxy03.

References

[1] M. M. Fraz 等, 《Blood vessel segmentation methodologies in retinal images – A survey》, *Comput. Methods Programs Biomed.*, 卷 108, 期 1, 页 407–433, 10 月 2012, doi: 10.1016/j.cmpb.2012.03.009.

[2] J. Malek, A. T. Azar, B. Nasralli, M. Tekari, H. Kamoun 和 R. Tourki, 《Computational analysis of blood flow in the retinal arteries and veins using fundus image》, *Comput. Math. Appl.*, 卷 69, 期 2, 页 101–116, 1 月 2015, doi: 10.1016/j.camwa.2014.11.017.

[3] W. Wang, J. Zhong, H. Wu, Z. Wen 和 J. Qin, 《RVSeg-Net: An Efficient Feature Pyramid Cascade Network for Retinal Vessel Segmentation》, 收入 *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 卷 12265, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu 和 L. Joskowicz, 编, 收入 Lecture Notes in Computer Science, vol. 12265. , Cham: Springer International Publishing, 2020, 页 796–805. doi: 10.1007/978-3-030-59722-1_77.

[4] R. Kawasaki 等, 《Retinal Microvascular Signs and Risk of Stroke: The Multi-Ethnic Study of Atherosclerosis (MESA)》, *Stroke*, 卷 43, 期 12, 页 3245–3251, 12 月 2012, doi: 10.1161/STROKEAHA.112.673335.

[5] G. Gojić 等, 《Deep Learning Methods for Retinal Blood Vessel Segmentation: Evaluation on Images with Retinopathy of Prematurity》, 收入 *2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY)*, 9 月 2020, 页 131–136. doi: 10.1109/SISY50555.2020.9217082.

[6] P. Liang 等, 《MambaSAM: A Visual Mamba-Adapted SAM Framework for Medical Image Segmentation》, *IEEE J. Biomed. Health Inform.*, 页 1–12, 2025, doi: 10.1109/JBHI.2025.3544548.

[7] Y. Lu 等, 《SKGC: A General Semantic-Level Knowledge Guided Classification Framework for

Fetal Congenital Heart Disease》, *IEEE J. Biomed. Health Inform.*, 卷 28, 期 10, 页 6105–6116, 10 月 2024, doi: 10.1109/JBHI.2024.3426068.
[8] B. Pu 等, 《HFSCCD: A Hybrid Neural Network for Fetal Standard Cardiac Cycle Detection in Ultrasound Videos》, *IEEE J. Biomed. Health Inform.*, 卷 28, 期 5, 页 2943–2954, 5 月 2024, doi: 10.1109/JBHI.2024.3370507.
[9] O. Ronneberger, P. Fischer 和 T. Brox, 《U-Net: Convolutional Networks for Biomedical Image Segmentation》, 2015 年 5 月 18 日, *arXiv*. arXiv:1505.04597. 见于: 2024 年 9 月 16 日. [在线]. 载于: http://arxiv.org/abs/1505.04597
[10] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh 和 J. Liang, 《UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation》, 2020 年 1 月 28 日, *arXiv*. arXiv:1912.05074. 见于: 2024 年 9 月 25 日. [在线]. 载于: http://arxiv.org/abs/1912.05074
[11] W. Liu 等, 《Full-Resolution Network and Dual-Threshold Iteration for Retinal Vessel and Coronary Angiograph Segmentation》, *IEEE J. Biomed. Health Inform.*, 卷 26, 期 9, 页 4623–4634, 9 月 2022, doi: 10.1109/JBHI.2022.3188710.
[12] Y. Zhou, H. Chang, X. Lu 和 Y. Lu, 《DenseUNet: Improved image classification method using standard convolution and dense transposed convolution》, *Knowl.-Based Syst.*, 卷 254, 页 109658, 10 月 2022, doi: 10.1016/j.knosys.2022.109658.
[13] D. Dai 等, 《I 2 U-Net: A dual-path U-Net with rich information interaction for medical image segmentation》, *Med. Image Anal.*, 卷 97, 页 103241, 10 月 2024, doi: 10.1016/j.media.2024.103241.
[14] C. Chen, L. Yu, S. Min 和 S. Wang, 《MSVM-UNet: Multi-Scale Vision Mamba UNet for Medical Image Segmentation》, 2024 年 8 月 25 日, *arXiv*. arXiv:2408.13735. 见于: 2024 年 9 月 16 日. [在线]. 载于: http://arxiv.org/abs/2408.13735
[15] F. Milletari, N. Navab 和 S.-A. Ahmadi, 《V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation》, arXiv.org. 见于: 2024 年 10 月 8 日. [在线]. 载于: https://arxiv.org/abs/1606.04797v1
[16] F. Isensee 等, 《nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation》, 2018 年 9 月 27 日, *arXiv*. arXiv:1809.10486. doi: 10.48550/arXiv.1809.10486.
[17] J. M. Jose, V. Sindagi, I. Hacihaliloglu 和 V. M. Patel, 《KiU-Net: Towards Accurate Segmentation of Biomedical Images using Over-complete Representations》, 2020 年 7 月 8 日, *arXiv*. arXiv:2006.04878. doi: 10.48550/arXiv.2006.04878.
[18] N.-V. Ho, T. Nguyen, G.-H. Diep, N. Le 和 B.-S. Hua, 《Point-Unet: A Context-aware Point-based Neural Network for Volumetric Segmentation》, 2024 年 2 月 28 日, *arXiv*. arXiv:2203.08964. doi: 10.48550/arXiv.2203.08964.
[19] J. Hu, L. Shen 和 G. Sun, 《Squeeze-and-Excitation Networks》, 收入 *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, 6 月 2018, 页 7132–7141. doi: 10.1109/CVPR.2018.00745.
[20] Q. Wang, O. Aramoon, P. Qiu 和 G. Qu, 《Efficient Transfer Learning on Modeling Physical Unclonable Functions》, 收入 *2020 21st International Symposium on Quality Electronic Design (ISQED)*, Santa Clara, CA, USA: IEEE, 3 月 2020, 页 1–6. doi: 10.1109/ISQED48828.2020.9137057.
[21] F. Wang 等, 《Residual Attention Network for Image Classification》, 收入 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA: IEEE, 7 月 2017, 页 6450–6458. doi: 10.1109/CVPR.2017.683.
[22] Z. Qin, P. Zhang, F. Wu 和 X. Li, 《FcaNet: Frequency Channel Attention Networks》, 2021 年

7 月 23 日, *arXiv*. arXiv:2012.11879. doi: 10.48550/arXiv.2012.11879.
[23] S. Woo, J. Park, J.-Y. Lee 和 I. S. Kweon, 《CBAM: Convolutional Block Attention Module》, 收入 *Computer Vision – ECCV 2018*, 卷 11211, V. Ferrari, M. Hebert, C. Sminchisescu 和 Y. Weiss, 编, 收入 Lecture Notes in Computer Science, vol. 11211. , Cham: Springer International Publishing, 2018, 页 3–19. doi: 10.1007/978-3-030-01234-2_1.
[24] O. Oktay 等, 《Attention U-Net: Learning Where to Look for the Pancreas》, arXiv.org. 见于: 2024 年 10 月 8 日. [在线]. 载于: https://arxiv.org/abs/1804.03999v3
[25] Q. Xu, Z. Ma, N. HE 和 W. Duan, 《DCSAU-Net: A Deeper and More Compact Split-Attention U-Net for Medical Image Segmentation》, 2022 年 9 月 24 日, *arXiv*. arXiv:2202.00972. doi: 10.48550/arXiv.2202.00972.
[26] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille 和 L.-C. Chen, 《Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation》, 收入 *Computer Vision – ECCV 2020*, 卷 12349, A. Vedaldi, H. Bischof, T. Brox 和 J.-M. Frahm, 编, 收入 Lecture Notes in Computer Science, vol. 12349. , Cham: Springer International Publishing, 2020, 页 108–126. doi: 10.1007/978-3-030-58548-8_7.
[27] Q. Hou, D. Zhou 和 J. Feng, 《Coordinate Attention for Efficient Mobile Network Design》, 收入 *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA: IEEE, 6 月 2021, 页 13708–13717. doi: 10.1109/CVPR46437.2021.01350.
[28] R. Gu 等, 《CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation》, *IEEE Trans. Med. Imaging*, 卷 40, 期 2, 页 699–711, 2 月 2021, doi: 10.1109/TMI.2020.3035253.
[29] H. Huang, Z. Chen, Y. Zou, M. Lu 和 C. Chen, 《Channel prior convolutional attention for medical image segmentation》, 2023 年 6 月 8 日, *arXiv*. arXiv:2306.05196. doi: 10.48550/arXiv.2306.05196.
[30] R. Sunkara 和 T. Luo, 《No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects》, 2022 年 8 月 7 日, *arXiv*. arXiv:2208.03641. 见于: 2024 年 9 月 22 日. [在线]. 载于: http://arxiv.org/abs/2208.03641
[31] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever 和 B. Van Ginneken, 《Ridge-Based Vessel Segmentation in Color Images of the Retina》, *IEEE Trans. Med. Imaging*, 卷 23, 期 4, 页 501–509, 4 月 2004, doi: 10.1109/TMI.2004.825627.
[32] A. D. Hoover, V. Kouznetsova 和 M. Goldbaum, 《Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response》, *IEEE Trans. Med. Imaging*, 卷 19, 期 3, 页 203–210, 3 月 2000, doi: 10.1109/42.845178.
[33] A. Carballal 等, 《Automatic multiscale vascular image segmentation algorithm for coronary angiography》, *Biomed. Signal Process. Control*, 卷 46, 页 1–9, 9 月 2018, doi: 10.1016/j.bspc.2018.06.007.
[34] F. Cervantes-Sanchez, I. Cruz-Aceves, A. Hernandez-Aguirre, M. A. Hernandez-Gonzalez 和 S. E. Solorio-Meza, 《Automatic Segmentation of Coronary Arteries in X-ray Angiograms using Multiscale Analysis and Artificial Neural Networks》, *Appl. Sci.*, 卷 9, 期 24, 页 5507, 12 月 2019, doi: 10.3390/app9245507.
[35] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen 和 J. Qin, 《SCS-Net: A Scale and Context Sensitive Network for Retinal Vessel Segmentation》, *Med. Image Anal.*, 卷 70, 页 102025, 5 月 2021, doi: 10.1016/j.media.2021.102025.
[36] S. Zhang 等, 《Attention Guided Network for Retinal Image Segmentation》, 收入 *Medical*

Image Computing and Computer Assisted Intervention – MICCAI 2019, 卷 11764, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap 和 A. Khan, 编, 收入 Lecture Notes in Computer Science, vol. 11764. , Cham: Springer International Publishing, 2019, 页 797–805. doi: 10.1007/978-3-030-32239-7_88.

[37] J. Chen 等, 《TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation》, 2021 年 2 月 8 日, *arXiv*. arXiv:2102.04306. doi: 10.48550/arXiv.2102.04306.

[38] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang 和 T. Zhang, 《Dual Encoder-Based Dynamic-Channel Graph Convolutional Network With Edge Enhancement for Retinal Vessel Segmentation》, *IEEE Trans. Med. Imaging*, 卷 41, 期 8, 页 1975–1989, 8 月 2022, doi: 10.1109/TMI.2022.3151666.

[39] M. Arsalan, T. M. Khan, S. S. Naqvi, M. Nawaz 和 I. Razzak, 《Prompt Deep Light-Weight Vessel Segmentation Network (PLVS-Net)》, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 卷 20, 期 2, 页 1363–1371, 3 月 2023, doi: 10.1109/TCBB.2022.3211936.

[40] J. Li, G. Gao, L. Yang, G. Bian 和 Y. Liu, 《DPF-Net: A Dual-Path Progressive Fusion Network for Retinal Vessel Segmentation》, *IEEE Trans. Instrum. Meas.*, 卷 72, 页 1–17, 2023, doi: 10.1109/TIM.2023.3277946.

[41] J. Wu, W. Ji, H. Fu, M. Xu, Y. Jin 和 Y. Xu, 《MedSegDiff-V2: Diffusion based Medical Image Segmentation with Transformer》, 2023 年 12 月 24 日, *arXiv*. arXiv:2301.11798. doi: 10.48550/arXiv.2301.11798.

[42] Y. Li 等, 《Global Transformer and Dual Local Attention Network via Deep-Shallow Hierarchical Feature Fusion for Retinal Vessel Segmentation》, *IEEE Trans. Cybern.*, 卷 53, 期 9, 页 5826–5839, 9 月 2023, doi: 10.1109/TCYB.2022.3194099.

[43] Y. Li 等, 《Diffusion Probabilistic Learning With Gate-Fusion Transformer and Edge-Frequency Attention for Retinal Vessel Segmentation》, *IEEE Trans. Instrum. Meas.*, 卷 73, 页 1–13, 2024, doi: 10.1109/TIM.2024.3420264.

[44] L. Mou 等, 《CS-Net: Channel and Spatial Attention Network for Curvilinear Structure Segmentation》, 收入 *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 卷 11764, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap 和 A. Khan, 编, 收入 Lecture Notes in Computer Science, vol. 11764. , Cham: Springer International Publishing, 2019, 页 721–730. doi: 10.1007/978-3-030-32239-7_80.

[45] K. Sun, B. Xiao, D. Liu 和 J. Wang, 《Deep High-Resolution Representation Learning for Human Pose Estimation》, 2019 年 2 月 25 日, *arXiv*. arXiv:1902.09212. doi: 10.48550/arXiv.1902.09212.

[46] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh 和 D. Batra, 《Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization》, 2019 年 12 月 3 日, *arXiv*. arXiv:1610.02391. doi: 10.48550/arXiv.1610.02391.