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

Jingling Zhang, Xiangfei Liu, Shuting Zheng, Wen Zhang and Jia Gu

***Abstract*— Accurate retinal microvasculature segmentation demands a balanced combination of anatomical fidelity and hemodynamic relevance. However, existing methods fall short in preserving critical structures such as capillary junctions and bifurcations, thus limiting clinical applications and causing fragmentation. To address these problems, we propose DFMS-Net, a new dual-domain architecture inspired by hemodynamically awareness by integrating geometric-field modeling and functional-field optimization. Specifically, the geometric-field modeling module is combined by two modules: the Spatial Pathway Extractor (SPE) and the Transformer-based Topology Interaction (TTI). The SPE module excels in preserving vascular trajectory continuity by modeling feature correlations using the axial awareness attention learning mechanism, while TTI is designed to analyze long-remote spatial dependencies using adjacency matrices to guide the connections and interaction between bifurcation points, thereby resolving intricate branching patterns. Meanwhile, we develop a new functional-field optimization module which utilizes a transformer-based semantic attention amplification (SAA) component to enhance local intensity changes of low-contrast capillaries and suppress background noise of medical images. Subsequently, the geometric-hemodynamical consistency and semantic relations are unified by a dual-field synergy unit (DFSU), ensuring awareness of hemodynamic properties and accurate representation of complex vascular structures. Considering the demand of distinct scenarios, we develop two special variants: Variant-1 for lightweight deployment with 5.73M parameters, and Variant-2 for managing significant variations in vessel thickness. Extensive experiments on retinal (DRIVE and STARE) and coronary angiography (DCA1 and CHUAC) datasets, demonstrate that the proposed DFMS-Net achieves state-of-the-art performance. Notably, the generalization capability on coronary angiography further provides a foundation for predicting and diagnosing retinal and blood-related diseases. The code will be available on https://github.com/699zjl/DFMS-Net-new**

***Index Terms*— Retinal microvasculature segmentation, Dual-field synergy unit, Hemodynamically-aware, Transformer, Attention**

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.

Jingling Zhang is with Institute of Data Science, City University of Macau, Macau, 999078, China, also with New Engineering Industry College, Putian University, Putian, Fujian Province 351100, China (e-mail: D24092100270@cityu.edu.mo).

Xiangfei Liu is with Institute of Data Science, City University of Macau, Macau, 999078, China, also with Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, 518055, China (e-mail: xfliu0102@163.com).

Shuting Zheng is with Institute of Data Science, City University of Macau, Macau, 999078, China, also with New Engineering Industry College, Putian University, Putian, Fujian Province 351100, China (e-mail: fgtear@ptu.edu.cn).

Wen Zhang is with Charles E. Schmidt College of Science, Florida Atlantic University, Boca Raton, FL 33431, USA (e-mail: zhangw@fau.edu).

Jia Gu is with Institute of Data Science, City University of Macau, Macau, 999078, China, corresponding author (e-mail: jiagu@cityu.edu.mo).

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 multi-branched vessel disorders [1], [2]. The morphological characteristics of retinal vessels, such as vascular width, tortuosity, and branching patterns, which act as sensitive biomarkers for ocular pathologies, 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 evidences suggest 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 implement preventive strategies prior to the manifestation of detectable morphological alterations [2]. Prior studies demonstrate that medical image segmentation is the fundamental task of hemodynamic simulations, and the quality of segmentation results (e.g., the geometry of the vessel) directly affects hemodynamic simulation outcomes [5], [6], [7]. Considering the clinical significance of the structural spatial features and hemodynamically-informed analysis for fundus and blood-related diseases, 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. (a) 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]. (b) Imaging limitations: The low-contrast between capillaries and background noise, combined with motion artifacts, makes accurate segmentation complex [8]. (c) Tree-like structure challenges: The tree-like architecture of retinal vessels, with its intricate network of delicate distal branches, poses significant segmentation difficulties due to sub-resolution vessel dimensions, substantially increasing computational demands. Apart from the inherent properties of medical images, the limited scale of medical image datasets, which results from the high cost and complexity of data acquisition and annotation, presents major challenges for effective model training [9], [10], [11].

Currently, deep learning techniques for retinal vessel segmentation primarily adhere to two patterns: U-Net-based and transformer-based. Although U-Net [12] and its variants perform well in hierarchical feature aggregation thanks to the encoder-decoder design, they 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: (a) the quadratic computational complexity that limits clinical implementation, and (b) 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 awareness, and also fall short in resolving the critical gap for clinically operational segmentation outcomes, leading to unreliable vascular mapping in diagnostic and prognostic contexts.

To address these challenges, a dual-field microvascular segmentation framework is proposed. By integrating geometric-field modeling with functional-field optimization modules, our work achieves the following significant advancements for retinal vessel segmentation:

- We propose the first dual-field synergistic unit (DFSU) module, which uses the hemodynamics-inspired attention learning method. This model combines geometric-field modeling (preserving structural continuity) with functional-field optimization (enhancing clinical semantic significance), focusing on addressing the fragmentation of key critical structures.
- We develop a new geometric-field modeling module via two hierarchical attention strategies, including spatial pathway extractor (SPE)—employs a directional coordinate attention to preserve vascular trajectory continuity by resolving spatial discontinuities in branch networks, and the transformer-derived topology interaction (TTI)—leverages transformer concept to interpret complex branching patterns while ensuring anatomical accuracy.
- We introduce a novel functional-field optimization module through the semantic attention amplifier (SAA) component. It can dynamically recalibrate channel-wise features while enhancing low-contrast capillaries and suppressing image artifacts.
- We conduct extensive cross-domain experiments across retinal and coronary angiography datasets, and develop two specialized variants for meeting different scenarios demands. 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 [12] remaining the cornerstone due to its encoder-decoder design that effectively balances local detail preservation and global contextual integration. Its variants—including UNet++ [13] for nested feature refinement, FR-UNet [14] for restoring the lost spatial information, I^2U-Net [15] for bridging the encoder and decoder semantic gap, and MSVM-UNet [16] for simultaneous capturing of localized fine-grained patterns and comprehensive contextual integration across vascular hierarchies. In addition to the aforementioned methodologies, several other approaches such as nnU-Net [17], KiU-Net [18], and Point-Unet [19] 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 have witnessed four dominant research trajectories: (a) hybrid architectures (e.g., TransUNet [20], MISSFormer [21]) that integrate CNNs and Transformers, (b) efficient attention designs (e.g., Swin-trans [22], DAE-Former [23]) to mitigate computational overhead, (c) generative models (e.g., MedSegDiff [24]) for noise-robust segmentation, and (d) graph-based approaches (e.g., DPL-GFT-EFA [25]) for topological reasoning. However, these methods either neglect hemodynamic-semantic alignment or compromise geometric fidelity through heuristic approximations.

C. Attention Mechanisms

Attention mechanisms has evolved through three key phases. Channel attention, initiated by SENet's channel-wise recalibration [26] and refined by ECANet's [27] lightweight dynamic convolution, optimizes feature discriminability for low-contrast regions. Spatial attention is advanced by RA-Net's [28] through its residual-guided weighting, which is critical for handling anatomical variability in CT/MRI. Frequency attention (e.g., FcaNet's DCT-based spectral filtering [29]) enhances edge preservation in ultrasound and retinal imaging. Hybrid architectures like CBAM [30] synergizes

multi-dimensional attention, whereas medical-specific frameworks—Attention UNet (gated skip connections) [31] and DCSAU-Net (cross-scale fusion) [32] tailor 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-aware inspired architecture that revolutionizes retinal vasculature mapping through synergistic geometric-functional learning. Building upon FR-UNet [14], our framework consists of four 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 the standard 2×2 convolution and the 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, a crucial component of DFMS-Net, accurately maps vascular anatomy by integrating two complementary mechanisms: spatial path extractor (SPE) and transformer-derived topology interaction (TTI). The SPE module embeds position-sensitive and orientation-specific information along horizontal and vertical axes, while maintaining vessel trajectory continuity and modeling orientation priors. The TTI module captures long-remote spatial dependencies between different pixels by leveraging the learnable attention matrix, thus it enables accurate non-local connection for branch alignment. To combine geometric accuracy with hemodynamic relevance, the SPE-optimized features are then fed into the semantic attention amplifier module (SAA) in the functional field optimization module. Here, the channel-wise attention approach is designed to emphasize key semantic features and minimize background artifacts in medical images. Considering the clinical lightweight requirements and significant variations in vessel diameter across cross-domain images, we introduce two additional specialized variants.

B. Geometric-Field Modeling

The intricate architecture of retinal vessel segmentation—characterized by delicate local capillary structures and highly variable global tree-like morphology—demands comprehensive contextual features while preserving vascular topological continuity. This dual focus ensures anatomical fidelity for geometric precision and maintains hemodynamic consistency that is essential for clinical phenotyping. As demonstrated by many vision tasks [26], [30], [33], [34], spatial and channel information play crucial role in capturing structure features and generating 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, ,thereby 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.

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 $Z \in \mathbb{R}^{C\times H\times W}$, which is directly generated from the third layer of the encoder module. Here H, W and C denotes the height, width and channel number of the feature map, respectively. The directional feature maps are then embeded by two orthogonal average poling with kernel sizes of (H,1) and (1,W), dependently [34]. These pooling operations simulate the clinical practice of tracing vessel continuity along their dominant orientations. As shown in Fig. 2 (taking the vertical direction as an example), 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, and 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 direction-sensitive relationships and spatial location information, which are critical to model complex strcucture 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, those two orthogonal embeddings are concatenated and then projected into a composite signal $p_c \in \mathbb{R}^{2H\times 1}$ through a 1×1 convolutional filter $f^{1\times 1}$, which acts as a differentiable vessel edge detector:

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

Here, the hard-swish activation function (δ) is used to enhance non-linear separability between vessel pixels and background, while batch normalization (BN) is utilized to stabilize training process under varying contrast conditions. The split features g_h and g_w represent orthogonal attention

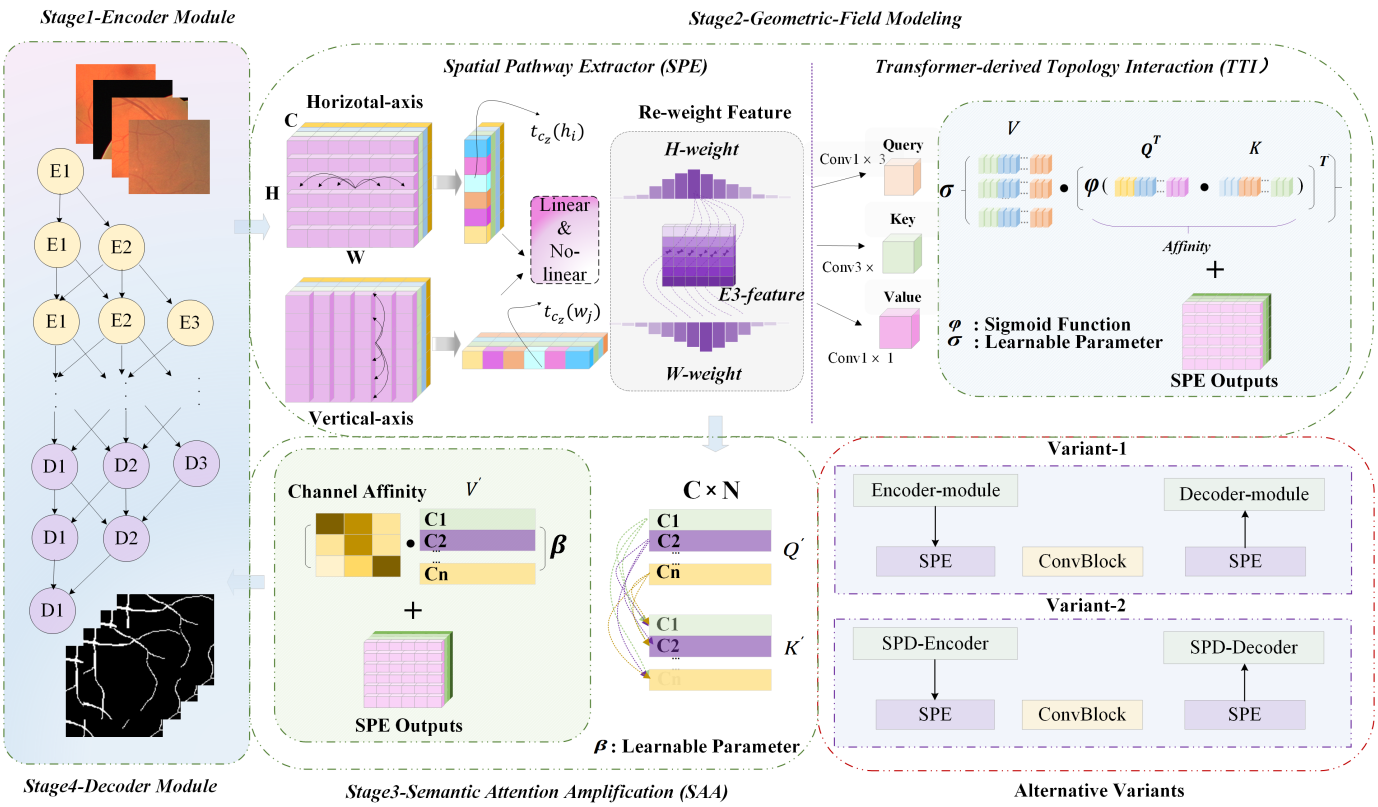

Fig. 1. The overview of the proposed DFMS-Net. It consists of four components: an encoder module, a geometric-field modeling module, a functional-field optimization module and a decoder module. The SPE module, is synergistically integrated with the no-strided convolution design, dynamically forming two lightweight enhanced variants

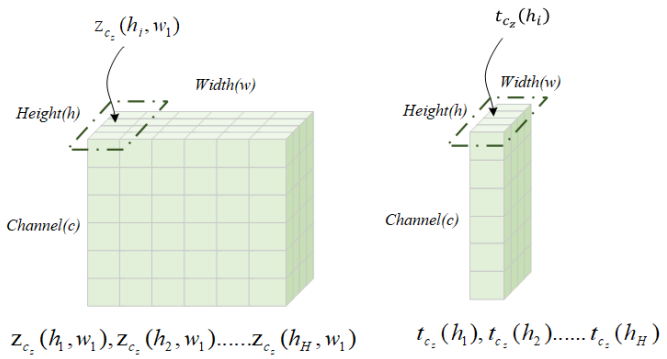

Fig. 2. Direction-aware feature map generation in height direction.

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 obtained by element-wise multiplication which is designed to geometrically constrain vascular connectivity. Then they are subsequently fed into another two distinct conv layers to generate the attention maps. Lastly, the final re-weighted feature can be written as:

$$SPE_{out} = Z \odot S(C_h^{1\times1}(g_h)) \odot S(C_w^{1\times1}(g_w)) \quad (4)$$

Here, SPE_{out} indicates the final output feature maps, while $C_h^{1\times1}$ and $C_w^{1\times1}$ correspond to two distinct conv kernels of size 1×1. S represents the sigmoid gate, which is used to adjust feature responses and suppress disconnected pseudo-vessels while boosting true bifurcations according to anatomical plausibility.

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 due to the minor intensity fluctuations and complicated topological interactions among remote vessel branches. This challenge is addressed by the TTI module by being aware of non-local spatial dependencies. Specifically, it creates pixel-wise affinity map through interactions of 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, the spatial connectivity restores the topology of hemodynamic consistency, which is essential not only for microvascular segmentation but also for relevant disease analysis. Topological interaction modeling is realized via a differentiable pipeline, consisting of the following phases:

Firstly, two standard 3×3 convolutional layers are used to

further encode SPE_{out} by fusing more richer local information. In order to capture orientation-specific contextual dependencies, the generated local feature $X \in \mathbb{R}^{C' \times H' \times W'}$ is then projected into three distinct representations—query Q, key K, and value V—throught the 1×3, 3×1, and 1×1 convolutional kernels, respectively. Here, these three convolutional kernels are explicitly used to model horizontal and vertical vascular trajectory continuity, aligning with the anisotropic structure of retinal structures. To enhance the inter-features relationships, they are then immediately flattened into a sequence of size $C' \times N'$, where $N' = H' \times W'$. A matrix multiplication is then performed to compute the affinity score between Q and K. The affinity score measures the probability of pairwise vascular connectivity, and the higher the value, the greater the likelihood of connection. Then, the softmax function is utilized to normalize the affinity score, resulting in the attention weights. Meanwhile, the value matric V is reshaped to $\mathbb{R}^{C' \times N'}$ and is multiplied by the attention weights via a matrix multiplication, aiming to calculate the affinity attention maps. Finally, to avoid overemphasizing specific objects, we scale the affinity attention maps using a trainable parameter. These adjusted feature maps are subsequently combined with the original input X through a residual connection, resulting in improved generalization capacity for the model. The calculation method is defined by the following formula:

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

Here, φ denotes the softmax function for computing the attention weights, and $\cdot$ represents the matrix multiplication. The learnable scalar σ is used to balance global topology iniformation against local geometric accuracy, adapting to variations in vessel density 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 channel features, enabling channel attention mechanisms to prioritize clinically salient regions [35], [36]. This selective focus enhances the model's capability to effectively represent and emphasize functional semantics of inter-ichannel features. Thus, in this section, we further explore the inter-relationships among channel maps with the computed coordinate attention maps. Unlike the SE attention mechanism [26], 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 firstly project the geometric-refined features $X \in \mathbb{R}^{C' \times H' \times W'}$ 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'. The 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 $Aff_{new} \in \mathbb{R}^{(C',C')}$ presents the resulting new affinity matrix. This operation suppresses dominant but less informative channel correlations (e.g., background-dominated channels), while highlighting subtle yet hemodynamic-indicative relationships. Similar to above section, we then normalize the calculated affinity matrix with softmax function φ 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:

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

Where, φ denotes the softmax function, SAA_{out} represent the output of SAA module, and β represents the learnable parameter.

D. Dual-Field Synergy Unit

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

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

E. Alternative Variants

Athough DFMS-Net achieves balanced performance, the complex demands of clinical scenarios have led us to design two additional specialized variants to better accommodate specific challenges. As illustrated in Fig. 1, Variant-1 is created to maximize vascular network connectivity for awareness of blood flow, leveraging a lightweight architectural design for efficiency. Variant-2 is designed to enhance capillary geometry information, specifically for scenarios characterized by marked changes in vessel size on across-domain datasets. 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 through direction-aware pooling, while encoding geometric continuity constraints. Meanwhile, Decoder-SPE leverages the same SPE mechanism to refine upsampled features, ensuring coherence across 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 (compared to 5.76M for DFMS-Net), Variant 1 achieves superior connectivity through its inductive bias toward vascular symmetry. This demonstrates that performance gains can be achieved without increasing model complexity.

2) Variant-2: To address the vanishing capillary problem and significant differences in vessel size of low-resolution fundus images, this variant introduces a non-destructive downsampling module (SPD-Conv) [37], replacing traditional strided convolutions and pooling layers. With 6.07M parameters, Variant-2 prioritizes geometric fidelity over parameter efficiency, achieving sub-pixel boundary precision, which is essential for accurate capillary morphometry. As illustrated in Fig. 3, the architecture comprises: Firstly, a 1×1 convolutional layer is used to adjust channel dimensions without spatial compression, thereby preserving boundary voxels that are typically lost in strided convolution. Secondly, the feature map $Y \in \mathbb{R}^{(D,S,S)}$ is decomposed into four sub-features via a stride-2 slicing:

$$\begin{aligned} M_0 &= Y[\ldots, :: 2, :: 2], \\ M_1 &= Y[\ldots, 1:2, :: 2], \\ M_2 &= Y[\ldots, :: 2, 1:2], \\ M_3 &= Y[\ldots, 1:2, 1:2]. \end{aligned} \tag{10}$$

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

IV. Experiments

A. DATASETS

We evaluate our method on four public datasets with varying image resolutions and sizes, including DRIVE [38], STARE [39], DCA1 [40] and CHUAC [41]. The details are as follows:

1) DRIVE Dataset: The DRIVE dataset contains 40 fundus images, each with dimensions of 584×565 pixels. The dataset is evenly divided into 20 training and 20 testing images. These images were captured by 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 measuring 700×605 pixels. In alignment with SCS-Net [42], we adopt a 10-fold cross-validation method to split the training and testing sets.

3) DCA1 Dataset: The DCA1 dataset consists of 130 X-ray coronary angiograms images, which are obtained from the Mexican Social Security Institute. Each image is manually annotated by an expert cardiologist, and has a resolution of 300×300 pixels in pgm grayscale format. These images are divided into 100 images for training and 30 for testing.

4) 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. Same to 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 data size, 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:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\text{Sen} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{Spe} = \frac{TN}{TN + FP} \tag{13}$$

$$\text{F1} = \frac{2TP}{2TP + FP + FN} \tag{14}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{15}$$

For medical segmentation results, pixels are divided into two categories: vascular pixels (those with blood flow) and non-vascular pixels (those without blood flow). The classification of each image produces four potential results: 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 [20], DE-DCGCN-EE [43], PLVS-Net [44], DPF-Net [25], MedSegDiff [24], GT-DLA-dsHFF [45] and DPL-GFT-EFA [25]. Also, we compare our framework with classic methods, such as U-Net [12], UNet++ [13] and Attention U-Net [31]. To demonstrate the advantages of our

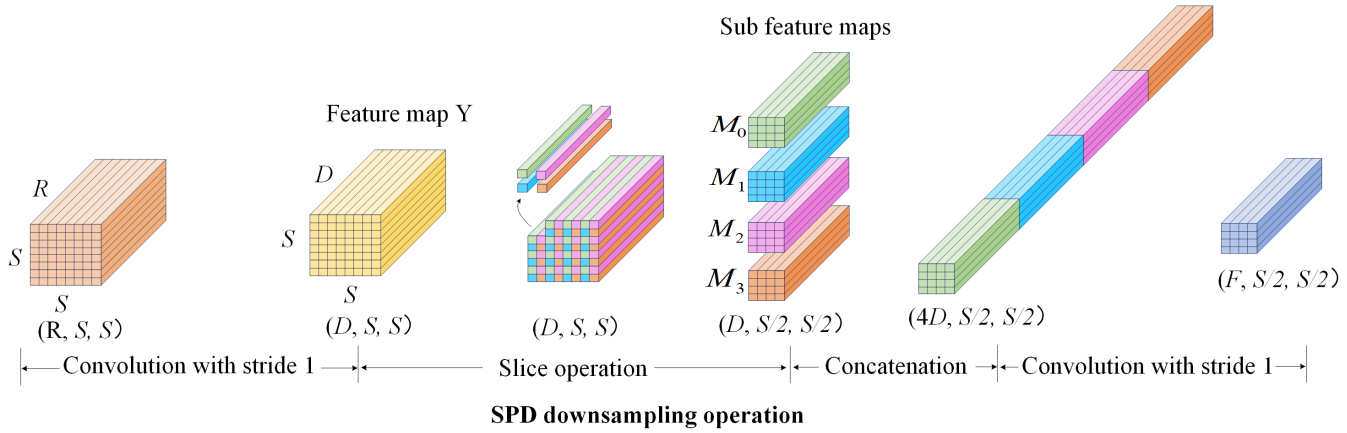

Fig. 3. SPD-down module. The newly designed SPD downsampling operation is based on slicing strategy. We firstly use the convolution layer with stride 1 to change channel numbers, then the slice operation is utilized to generate four sub feature maps. Finally, the convolution layer is reused to decrease channel numbers.

approach more effectively, we conduct a comprehensive evaluation incorporating both qualitative and qouantitative analyses across four public benchmarks. Quantitative comparisons on the DRIVE and STARE fundus datasets are summarized in Table I. Cross-domain generalization capabilities are assessed through the DCA1 and CHUAC coronary angiography benchmarks, with detailed metrics presented in Table II.

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 [45] 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 and IoU. 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 [20] by 0.82% and the diffusion-driven MedSegDiff [24] by 10.26% in IoU, verifying the promissing robustness to architecture-bias. While DPL-GFT-EFA [25] 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. Through boundary-centric SPD [37], Variant-2 is improved by 0.93% in sensitivity compared to DFMS-Net, while keeping a robust IoU (71.06%). These outcomes validate DFMS-Net's effectiveness in geometric precision and 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 [25] 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 [45] 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 [20] 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.

Fig. 4 demonstrates that our method has a strong ability to effectively preserve topology-structural features in medical image segmentation, particularly in complex scenarios. Here, the x-axis denotes various methods in Table I and II, and they are encoded as 1, 2, etc. from top to bottom. The y-axis is the AUC or IoU metric. This chart intuitively displays our experimental results across two different datasets, revealing that the designed approach maintains the best segmentation.

3) Evaluation on the DCA1 and CHUAC Dataset: In coronary angiography, the proposed DFMS-Net shows significant breakthrough in consistency, which is evidenced by its remarkable performance on DCA1 and CHUAC benchmarks in Table II. The dual-domain advantage verifies its practical ability to tackle two significant challenges in medical image segmentation: the detection of artifacts vessels and the preservation

TABLE I
COMPARISON WITH THE STATE-OF-THE-ART METHODS ON DRIVE AND STARE. THE BEST RESULTS ARE IN BOLD.

Methods	DRIVE						STARE					
	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)
U-Net [12]	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++ [13]	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-Net [31]	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 [20]	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 [43]	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 [44]	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 [46]	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 [24]	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 [45]	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 [25]	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**

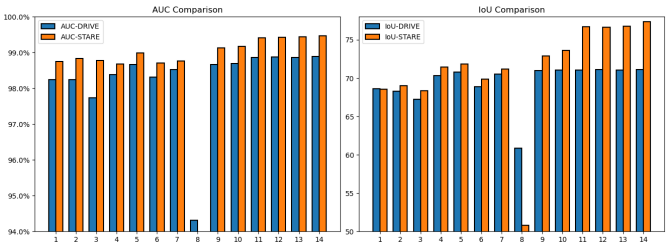

Fig. 4. Bar chart comparison results on AUC and IoU metrics across DRIVE and STARE datasets. The x-axis represents the methods in Table I-II and they are encoded as 1, 2, 3, etc. from top to bottom. The y-axis represents the AUC or IoU performance on different datasets.

of continuity in low-signal pixel points. The proposed DFMS-Net reaches higher Spe on both datasets: 98.91% for DCA1 and 99.60% for CHUAC. Moreover, DFMS-Net demonstrates superior performance by focusing on the retention of vascular topology and enhancement of critical-vessel intensity, which is enabled by dynamic adjustment of the weights of geometric and branched structural features, thereby achieving a high AUC of 99.19%. It effectively suppresses noise interference with the DFSU's phase-aware attention mechanism, obtaining a Sen of 82.12% on CHUAC, compared to UNet++'s 66.87%.

Variant-1 maintains a high ACC of 97.81% on DCA1, an increase of 0.18% compared to the baseline. With only 5.73 million parameters, it ensures efficient deployment, making a lightweight architecture for real-time interventions. Variant-2 achieves state-of-the-art IoU of 66.53% via the SPD-Conv's non-destructive downsampling on DCA1, and obtains the highest Sen of 83.65% by preserving capillary endpoints in low-SNR regions on CHUAC, demonstrating strong advantage of recovering blurred boundaries and handling scenarios with significant vessel size variations.

4) Visualization Segmentation Results on Four Datasets: As visualized in Fig. 5, the segmentation results show a clear demonstration of the dual-field synergy of DFMS-Net to address cross-domain vascular complexity. Due to space limitations, we focus on comparing the baseline network with our proposed framework. In retinal microvasculature domain (DRIVE and STARE), the baseline method fragments delicate capillary networks, particularly in low-contrast peripheral regions. As shown by discontinuous branching patterns marked by green indicator arrow in the 1st and 4th rows, arising from its inability to model anatomic continuity under hemodynamic variations. DFMS-Net solves this problem via its spatial pathway extractor (SPE), which enforces trajectory-aware attention to bridges gaps between capillary junctions while suppressing noise from non-vessel pixels (blue circle, line 1).

The visualization outcomes reaffirm DFMS-Net's hemodynamically-aware 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 4). Our designed Dual-Field Synergy Unit (DFSU) effectively balances boundary sharpness and noise interference, enabling the differentiation between pseudo-vessels and true vessels. That is a critical capability, where FR-Unet fails due to its static feature aggregation, leading to the misclassification of background pixels as valid branches (blue circle, row 3).

In line with aforementioned results, the introduced variants also show promising optimization ability. By utilizing the SPE method, Variant-1 achieves high-quality segmentation results. It preserves complex coronary detail detection and significantly improves 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 row 4.

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, which represent divergent vascular phenotypes (retinal microvasculature and complex bifurcations). We systematically disable individual modules (SPE/TTI/SAA), with configuration setting detailed in Table

TABLE II
COMPARISON WITH THE STATE-OF-THE-ART METHODS ON DCA1 AND CHUAC. THE BEST RESULTS ARE IN BOLD.

Methods	DCA1						CHUAC					
	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)	ACC (%)	Sen (%)	Spe (%)	AUC (%)	F1 (%)	IoU (%)
U-Net [12]	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++ [13]	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 [31]	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 [47]	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 [48]	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

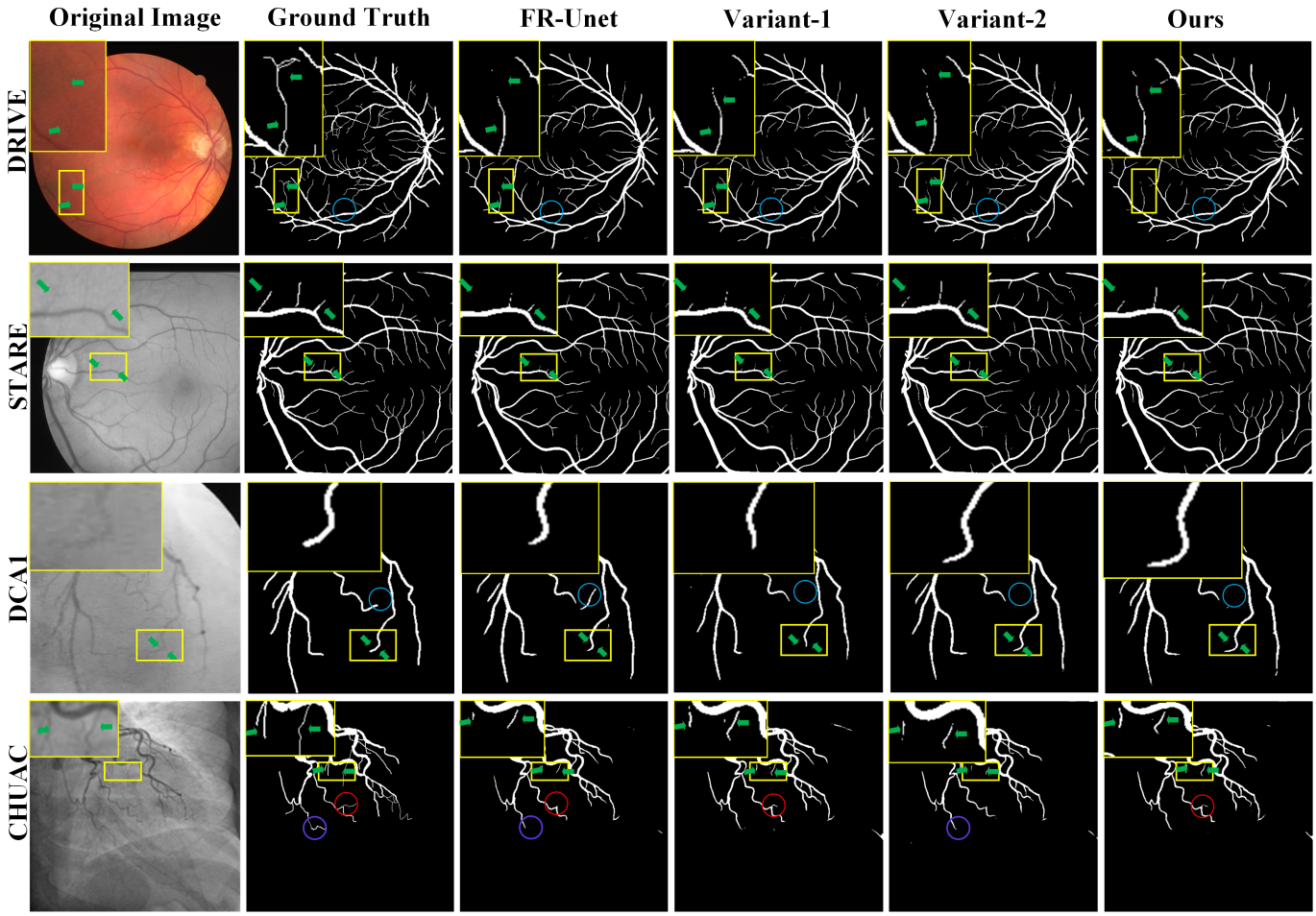

Fig. 5. Visualization segmentation results on four datasets. The yellow-boxed areas identify instances where the proposed methods excel over FR-Unet method, while arrows and circles indicate significant difference between different methods.

III. Tables IV-V 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-UNet baseline, by systematically integrating modules from the geometric (SPE, TTI) and functional (SAA) fields. Configurations are codified in Table III, 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 III
ABBREVIATIONS USED IN ABLATION STUDIES

Abbreviation	Explanation
/	Parallel connection
SPD_down	Replacing strided convolution with non-strided convolution
Down_SPE	SPE module concatenated to standard down-sampling
Out_weight	Weighted sum of outputs
SPD_down_SPE	SPE module concatenated to SPD_down operation
Out_SPE	SPE module concatenated to 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 IV and visualization attention maps in Fig. 7. On DRIVE, network (j) integrated SPE module boosts sensitivity by 0.29% and IoU by 0.04%, comparing to the network (e) which uses 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 that are usually lost due to background noise. Visualization attention maps in Fig. 7 further explain these improvements, demonstrating that 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. Specifically, network (j) obtains an improvement of 2.04% over network (e) on Sen, validating its superior performance in identifying thin-vessels branches. Remarkably, SPE overcomes the key limitation of TTI/SAA—despite their strength in global context aggregation, they fall short of maintaining hemodynamic-geometric relationships. Moreover, experimental results show that SPE preserves geometric features more effectively in the encoding layer compared to the decoding layer. It achieves 83.11% F1 on DRIVE and 87.18% on STARE (network (h) vs. (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. This approach reduces overlapping boundaries (green box) and achieves a 1.25% IoU gain (71.12% vs. 69.87%, TABLE IV). Despite the marginal improvement on sensitivity (88.57% vs. 82.37%), network (c) is difficult to balance between geometric precision and diagnostic fidelity. This is further supported by Fig. 7, revealing that it exhibits excessive focus on thick vessel edges (bright peripheral activations), and fails to preserve capillary continuity (fragmented halos).

When evaluated on the STARE dataset, the TTI module achieves a 13.08% improvement in Sen compared to network (c) integrated by SPE/SAA. This improvement is achieved by recovering more discontinuous bifurcations, which are typically fragmented by conventional methods due to noise interference or low contrast. Moreover, this great improvement is directly correlated with a 10.41% gain in IoU metric (77.35% vs. 66.94%, network(j) vs. (c)).

4) Effect of the SPD_down module: Experiments across topological configurations (b)-(d) and (g)-(f) in Table V, reveals the SPD_down's critical advantages. The IoU metric demonstrates its stability on DRIVE (71.10% $\rightarrow$ 71.06%), comparing network (b) with (d) (red box, Fig. 6). Yet it shows an increase on STARE (76.66% $\rightarrow$ 76.80%), indicating improved adaptability to complex vascular structures. Meanwhile, the designed module illustrates its enhanced detection of capillaries without increasing false positives on STARE (Spe: 98.85% $\rightarrow$ 98.92%, Sen: 80.61%$\rightarrow$85.55%). Similarly, network (f) demonstrates a substantial improvement in sensitivity (Sen) over (g), showcasing superior detection capability for low-contrast lesions. As illustrated in Fig. 6, the SPD-integrated method (network (f)) notably improves small vessel detection accuracy. Concurrently, it effectively reduces the misclassification of non-vessel pixels(red and blue boxes, row 2).

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 in Table V: (a) Deep Supervision with weighted summation (network (i)), which aggregates features from the last five layers using learnable weights, and (b) Average fusion (network (d)), where the multi-outputs are combined through element-wise averaging. On DRIVE, weighted summation approach (network (i) achieves higher Sen and AUC compared to network (d), but sacrifices ACC($\downarrow$0.46%), Spe ($\downarrow$1.09%), F1($\downarrow$1.07%) and IoU($\downarrow$1.56%). This trade-off suggests potential overfitting to vessel pixels in class-imbalanced scenarios. On STARE, the average fusion strategy dominates across all metrics (ACC: $\uparrow$0.92%, Sen: $\uparrow$12.61, AUC: $\uparrow$1.05%, F1: $\uparrow$7.01%, IoU: $\uparrow$10.30%), demonstrating better generalization on complex vascular networks.

The limited capability of deep supervision can be attributed to inherent gradient inconsistency. During backpropagation, each layer parameter has the effect of increasing the slight feature differences in low-contrast regions. Conversely, average fusion method keeps angiographic coherence, which is

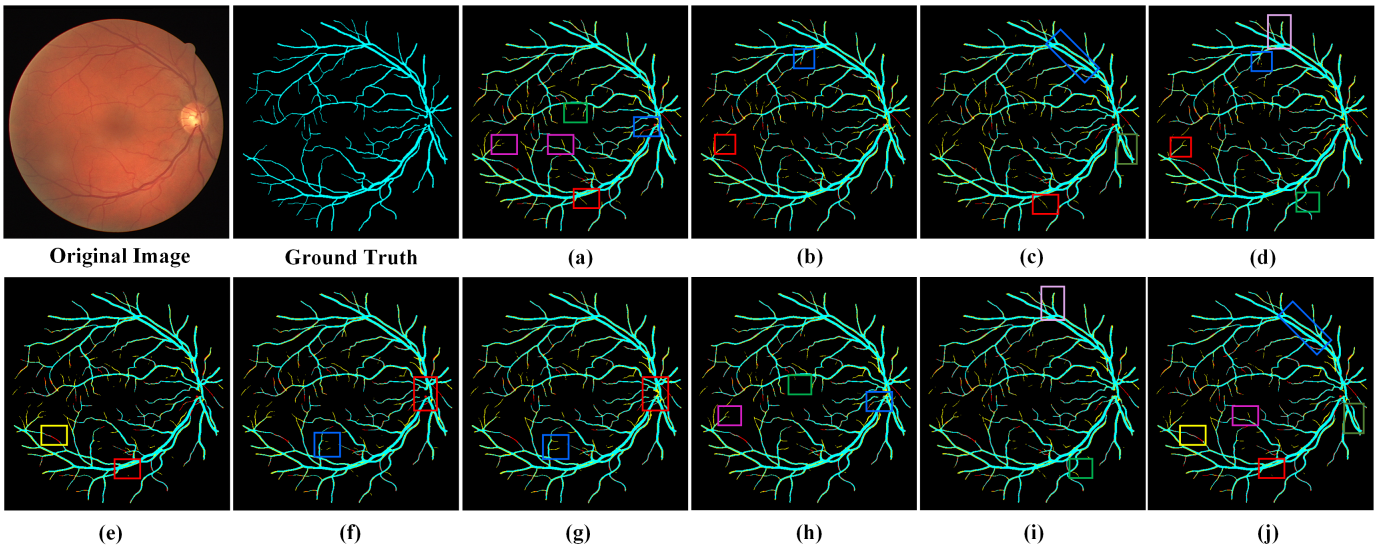

Fig. 6. Visualized colored segmentation results of ablation study on DRIVE dataset of network (a)-(j). The fluorescent blue pixels present the correctly segmented vessel pixels, the red and yellow pixels present the incorrectly vessel pixels and background pixels, respectively.

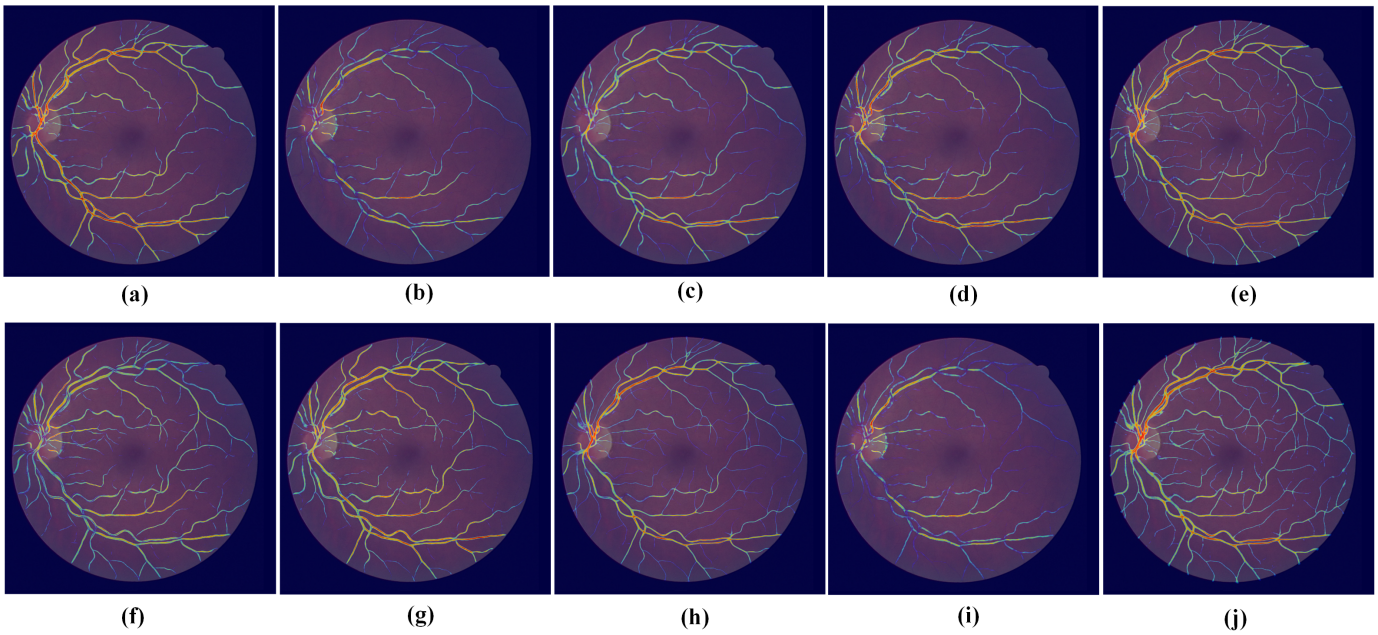

Fig. 7. Visualization of attention maps using Grad-CAM for ablation study on DRIVE dataset.

TABLE IV
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

prominently illustrated by the increased intensity in the same spatial area Fig. 6 (network (d) vs. (i)).

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 (%)
(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

6) Statistical Analysis: To further clearly quantify the segmentation performance, we use the six-number summary box plots to conduct non-parametric analysis of the distributions of TP, FP, and FN on the DRIVE dataset in Fig. 8, focusing on minimum, Q1, median, Q3, mean and maximum metrics. Notably, the TN metric is excluded to ensure interpretable visualization of critical error types, given the extreme class imbalance. Regarding the TP metric, DFMS-Net demonstrates the highest metric score, which is attributes to its remarkable ability of maintaining vascular continuity through SPE and TTI. Meanwhile, the lowest FN score suggests that the SAA module effectively reduces capillary omissions in low-contrast areas by suppressing the fake vessel pixels.

In comparison to network (a), network (d) achieves lower FN and higher TP across nearly all statistical metrics. This performance improvement validates the superiority of SPE in the vascular segmentation. By resolving pixel-level ambiguities at capillary junctions, SPE significantly reduces misclassification errors caused by complex vascular topology. Fig. 8 also demonstrates that network (b), (e) and (g) exhibit reduced FP, which is corresponding to their overemphasis on shape specificity. However, their over-reliance on morphological specificity leads to fragmented vessel segments, as rigid shape matching fails to capture continuous vascular.

V. DISCUSSIONS

The accurate segmentation of retinal microvasculature is important for blood-related disease detection and diagnosis, but existing methods struggle to preserves capillary connections and bifurcations. This limitation greatly impacts its clinical applicability, as fragmented vascular maps can lead to errors in hemodynamic analysis and early disease detection. Our DFMS-Net framework effectively combines anatomical accuracy with hemodynamic awareness, transforming retinal microvasculature segmentation into a practical clinical tool.

Our proposed DFMS-Net method employs the dual-field mechanism, integrating geometric-field modeling with functional-field optimization modules, which are specially designed to ensure structural coherence and preserve functional exactness. These techniques are vital to addressing the fragmentation issue of medical image segmentation. The geometric-field modeling utilizes the Spatial Pathway Extractor (SPE) which learns real axial perceptual attention features from real clinical datasets. This ensures the structural continuity that aligns with vascular scenarios. Through analyzing long-remote spatial dependencies, our method can explore special feature patterns that respect hemodynamic consistency via the Transformer-based Topology Interaction (TTI) module. The effectiveness of this technology is effectively validated by the ablation experiments on the two retinal datasets in Table 4. Compared to networks (j) and (e), we obtained impressive results with the improvement of sensitivity by 0.29%, AUC by 0.02%, F1 by 0.04%, and IoU by 0.04% on the DRIVE dataset by introducing SPE module. We also acquired enhanced ACC, Sen, AUC, F1, and IoU, especially increasing sensitivity by 2.04% on STARE dataset. Our developed TTI module demonstrated significantly performance as visualized by the blue boxes in Figure 6, showing its strong ability of suppressing edge ambiguity in thick vessels. On the other hand, the introduced functional-field optimization technique highlights subtle intensity fluctuations in weak capillaries by using a transformer-based semantic attention amplifier module. This method reflects the global relationship of vascular pixels and enhances the low-contrast capillary signals while reducing imaging artifacts. Furthermore, our designed variants can effectively meet the needs of different clinical scenarios. Variant-1 is designed for deployment in resource-constrained environments without considerable performance loss with less parameters, while Variant-2 is for scenarios that has significant differences in vessel size.

The evaluation of DFMS-Net on diver datasets such as retinal vessel datasets (DRIVE and STARE) and coronary angiography datasets (DCA1 and CHUAC), demonstrates its strong generalization ability on cross-domain images. This method obtains the highest ACC (97.08% & 97.72%), AUC(98.89% & 99.47%), and IoU (71.12% & 7.35%) on DRIVE and STARE, respectively, while archives the highest Spe on both DCA1 and CHUAC. In contrast, DPL-GFT-EFA's probability-based fusion strategy fails to capture the inherent anisotropic geometric priors of retinal vessels. Similarly, TransUNet method lacks directional priors, while the hierarchical attention-based GT-DLAdsHFF method struggles to focus on changes in finer vessels. These data cover the common characteristics of most data, indicating its strong adaptability to tackle different images from various clinical scenarios. Although it does not provide the evaluations for other types of datasets,such as CT scans and X-rays etc., its promising performance suggests that it can be effectively generalized to other imaging modalities, even those characterized by multi-resolutions and structural variations.

Although DFMS-Net has seen improvements, it still has certain limitations. Although our method is inspired by blood flow

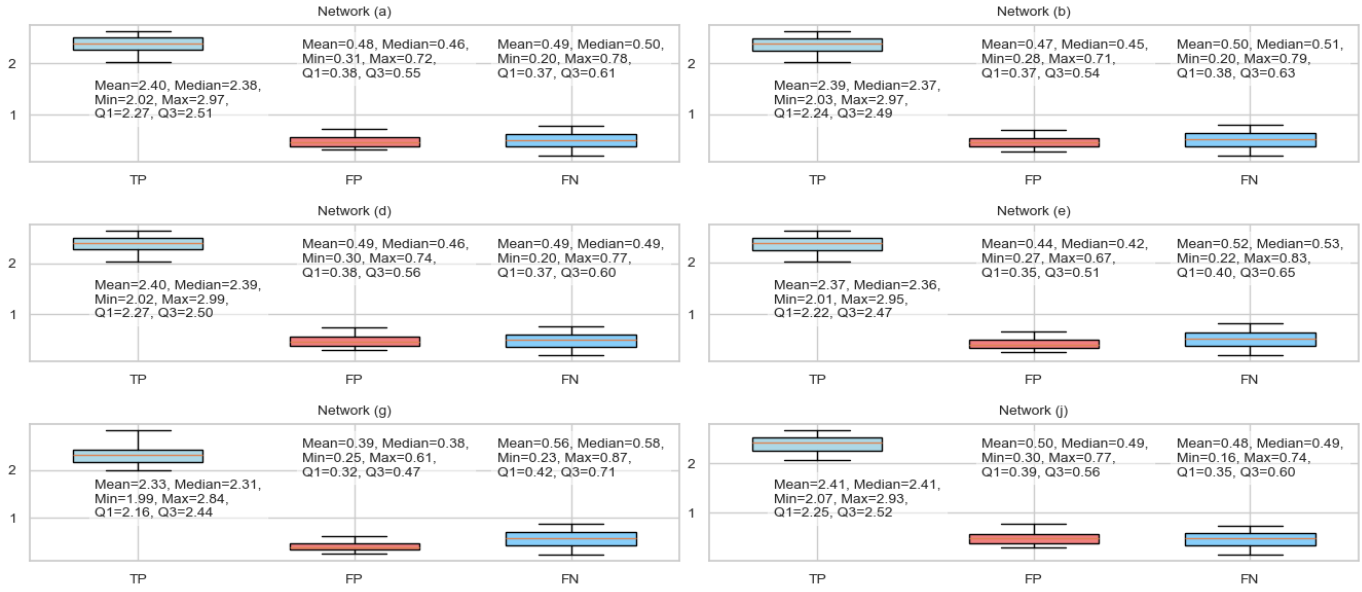

Fig. 8. Static analysis on DRIVE dataset of TP, FP, FN metrics for different methods. The y-axis represents the pixel number of the image (in ten thousand) and the x-axis denotes the evaluation metrics.

consistency, enhancing the relations between spatial structure and trajectory alignment, it does not explicitly integrate blood flow parameters into the segmentation algorithm. On the other hand, its segmentation ability in extreme lesion regions may be less effective due to the lack of specific disease-targeted evaluation. In future researches, we will further enhance physiological accuracy by incorporating hemodynamic simulation parameters, such as blood flow velocity, and evaluate the model's effectiveness in tracking disease progression in longitudinal studies.

VI. CONCLUSIONS

This study proposes a new dual-domain framework, called DFMS-Net, which improves microvascular segmentation performance by using accurate anatomical structures and consistent hemodynamics. It is designed by integrating axial-aware spatial modeling and transformer-based semantic optimization concepts, setting a new standard for preserving vascular structures, especially in low-contrast and complex branching situations. DFMS-Net shows markedly performance over SOTA DPL-GFT-EFA, demonstrating robust generalization across four retinal and coronary datasets. In the future, we will embeds dynamic hemodynamic parameters into segmentation algorithm to boost segmentation performance and maximiz its potential as a predictive tool for systemic vascular diseases.

VII. 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.

VIII. REFERENCES

[1] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images – a survey," *Computer Methods & Programs in Biomedicine*, vol. 108, no. 1, pp. 407–433, 2012.

[2] J. Malek, A. T. Azar, B. Nasralli, M. Tekari, H. Kamoun, and R. Tourki, "Computational analysis of blood flow in the retinal arteries and veins using fundus image," *Computers & Mathematics with Applications*, vol. 69, no. 2, pp. 101–116, 2015.

[3] W. Wang, J. Zhong, H. Wu, Z. Wen, and J. Qin, "Rvseg-net: An efficient feature pyramid cascade network for retinal vessel segmentation," 2020.

[4] R. Kawasaki, J. Xie, N. Cheung, E. Lamoureux, R. Klein, B. E. Klein, M. F. Cotch, A. R. Sharrett, S. Shea, T. Y. Wong *et al.*, "Retinal microvascular signs and risk of stroke: the multi-ethnic study of atherosclerosis (mesa)," *Stroke*, vol. 43, no. 12, pp. 3245–3251, 2012.

[5] Y. Sen, Y. Qian, A. Avolio, and M. Morgan, "Image segmentation methods for intracranial aneurysm haemodynamic research," *Journal of Biomechanics*, vol. 47, no. 5, pp. 1014–1019, 2014.

[6] T. Deschamps, P. Schwartz, D. Trebotich, P. Colella, D. Saloner, and R. Malladi, "Vessel segmentation and blood flow simulation using level-sets and embedded boundary methods," in *International Congress Series*, vol. 1268. Elsevier, 2004, pp. 75–80.

[7] T. Dunås, M. Holmgren, A. Wåhlin, J. Malm, and A. Eklund, "Accuracy of blood flow assessment in cerebral arteries with 4d flow mri: evaluation with three segmentation methods," *Journal of Magnetic Resonance Imaging*, vol. 50, no. 2, pp. 511–518, 2019.

[8] G. Gojić, V. Petrović, R. Turović, D. Dragan, A. Oros, D. Gajić, and N. Horvat, "Deep learning methods for retinal blood vessel segmentation: evaluation on images with retinopathy of prematurity," in *2020 IEEE 18th international symposium on intelligent systems and informatics (SISY)*. IEEE, 2020, pp. 131–136.

[9] P. Liang, L. Shi, B. Pu, R. Wu, J. Chen, L. Zhou, L. Xu, Z. Chen, Q. Chang, and Y. Li, "Mambasam: A visual mamba-adapted sam framework for medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, 2025.

[10] Y. Lu, G. Tan, B. Pu, H. Wang, B. Liang, K. Li, and J. C. Rajapakse, "Skgc: A general semantic-level knowledge guided classification framework for fetal congenital heart disease," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[11] B. Pu, K. Li, J. Chen, Y. Lu, Q. Zeng, J. Yang, and S. Li, "Hfsccd: a hybrid neural network for fetal standard cardiac cycle detection in ultrasound videos," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[13] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.

[14] W. Liu, H. Yang, T. Tian, Z. Cao, X. Pan, W. Xu, Y. Jin, and F. Gao, "Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation," *IEEE journal of biomedical and health informatics*, vol. 26, no. 9, pp. 4623–4634, 2022.

[15] D. Dai, C. Dong, Q. Yan, Y. Sun, C. Zhang, Z. Li, and S. Xu, "I2u-net: A dual-path u-net with rich information interaction for medical image segmentation," *Medical Image Analysis*, vol. 97, p. 103241, 2024.

[16] C. Chen, L. Yu, S. Min, and S. Wang, "Msvm-unet: Multi-scale vision mamba unet for medical image segmentation," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2024, pp. 3111–3114.

[17] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.

[18] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, "Kiu-net: Towards accurate segmentation of biomedical images using over-complete representations," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2020, pp. 363–373.

[19] N.-V. Ho, T. Nguyen, G.-H. Diep, N. Le, and B.-S. Hua, "Point-unet: A context-aware point-based neural network for volumetric segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer, 2021, pp. 644–655.

[20] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[21] X. Huang, Z. Deng, D. Li, and X. Yuan, "Missformer: An effective medical image segmentation transformer," *arXiv preprint arXiv:2109.07162*, 2021.

[22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[23] R. Azad, R. Arimond, E. K. Aghdam, A. Kazerouni, and D. Merhof, "Dae-former: Dual attention-guided efficient transformer for medical image segmentation," in *International workshop on predictive intelligence in medicine*. Springer, 2023, pp. 83–95.

[24] J. Wu, R. Fu, H. Fang, Y. Zhang, Y. Yang, H. Xiong, H. Liu, and Y. Xu, "Medsegdiff: Medical image segmentation with diffusion probabilistic model," in *Medical Imaging with Deep Learning*. PMLR, 2024, pp. 1623–1639.

[25] Y. Li, L. Xu, Y. Jin, X. Kuang, Y. Zhang, W. Cui, and T. Zhang, "Diffusion probabilistic learning with gate-fusion transformer and edge-frequency attention for retinal vessel segmentation," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[27] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 534–11 542.

[28] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.

[29] Z. Qin, P. Zhang, F. Wu, and X. Li, "Fcanet: Frequency channel attention networks," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 783–792.

[30] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[31] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[32] Q. Xu, Z. Ma, W. Duan *et al.*, "Dcsau-net: A deeper and more compact split-attention u-net for medical image segmentation," *Computers in Biology and Medicine*, vol. 154, p. 106626, 2023.

[33] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation," in *European conference on computer vision*. Springer, 2020, pp. 108–126.

[34] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 713–13 722.

[35] R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang, "Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation," *IEEE transactions on medical imaging*, vol. 40, no. 2, pp. 699–711, 2020.

[36] H. Huang, Z. Chen, Y. Zou, M. Lu, C. Chen, Y. Song, H. Zhang, and F. Yan, "Channel prior convolutional attention for medical image segmentation," *Computers in Biology and Medicine*, vol. 178, p. 108784, 2024.

[37] R. Sunkara and T. Luo, "No more strided convolutions or pooling: A new cnn building block for low-resolution images and small objects," in *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 2022, pp. 443–459.

[38] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE transactions on medical imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[39] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[40] F. Cervantes-Sanchez, I. Cruz-Aceves, A. Hernandez-Aguirre, M. A. Hernandez-Gonzalez, and S. E. Solorio-Meza, "Automatic segmentation of coronary arteries in x-ray angiograms using multiscale analysis and artificial neural networks," *Applied Sciences*, vol. 9, no. 24, p. 5507, 2019.

[41] A. Carballal, F. J. Novoa, C. Fernandez-Lozano, M. García-Guimaraes, G. Aldama-López, R. Calviño-Santos, J. M. Vazquez-Rodriguez, and A. Pazos, "Automatic multiscale vascular image segmentation algorithm for coronary angiography," *Biomedical Signal Processing and Control*, vol. 46, pp. 1–9, 2018.

[42] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "Scs-net: A scale and context sensitive network for retinal vessel segmentation," *Medical Image Analysis*, vol. 70, p. 102025, 2021.

[43] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang, and T. Zhang, "Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 8, pp. 1975–1989, 2022.

[44] M. Arsalan, T. M. Khan, S. S. Naqvi, M. Nawaz, and I. Razzak, "Prompt deep light-weight vessel segmentation network (plvs-net)," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 2, pp. 1363–1371, 2022.

[45] Y. Li, Y. Zhang, J.-Y. Liu, K. Wang, K. Zhang, G.-S. Zhang, X.-F. Liao, and G. Yang, "Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation," *IEEE Transactions on Cybernetics*, vol. 53, no. 9, pp. 5826–5839, 2022.

[46] J. Li, G. Gao, L. Yang, G. Bian, and Y. Liu, "Dpf-net: A dual-path progressive fusion network for retinal vessel segmentation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–17, 2023.

[47] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "Cs-net: Channel and spatial attention network for curvilinear structure segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*. Springer, 2019, pp. 721–730.

[48] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5693–5703.