DIMAN: A Dynamic Iterative Multi-scale Attention Network for Retinal Artery/Vein Segmentation

Yuanyuan Li, Jiansheng Wu, Yang Liu*

***Abstract*—To address critical challenges in medical image analysis, such as complex structure recognition, multi-scale feature fusion, and the need for computational efficiency, this study proposes an innovative Dynamic Iterative Multi-scale Attention Network (DIMAN). DIMAN incorporates two key modules: the Dynamic Mixed Channel Attention (DMCA) mechanism and Enhanced Partial Convolution (EPConv). DMCA adaptively integrates local and global features to capture multi-scale information effectively, improving the model's perception of intricate structures. EPConv enhances feature representation through selective channel processing combined with batch normalization and non-linear activation, while maintaining computational efficiency. Additionally, DIMAN adopts a multi-iteration refinement strategy that progressively enhances segmentation accuracy. Extensive experiments on two public retinal artery and vein segmentation datasets, RITE and LES-AV, demonstrate its effectiveness. On the RITE dataset, DIMAN achieves 96.04% sensitivity, 97.40% specificity, and 96.50% accuracy, outperforming current mainstream methods across all metrics. On the LES-AV dataset, it obtains 95.33% sensitivity, 94.81% specificity, and 93.75% accuracy, confirming its strong generalization and stability. Further analysis shows that DIMAN excels in segmenting fine vessels and complex bifurcation regions, highlighting its superior structural perception. Overall, the results validate the contributions of DMCA and EPConv to vascular feature discrimination. DIMAN achieves high segmentation accuracy with efficient computation, demonstrating excellent robustness and promising potential for clinical application.**

***Index Terms*—Artery-Vein segmentation, Fundus retinal images, Enhanced Partial Convolution, Iterative refinement**

I. Introduction

Alterations in retinal vascular morphology have significant clinical value for the early detection of multiple systemic and ocular pathologies. As the only noninvasively observable microvascular system in humans, the retinal vasculature serves as a critical biomarker, reflecting not only the progression of ocular diseases such as diabetic retinopathy, hypertensive retinopathy, and glaucoma, but also providing a basis for the early diagnosis of systemic conditions, including cardiovascular and cerebrovascular disorders. The structural and functional characteristics of retinal vessels are therefore essential indicators for early detection of severe pathologies[1]. Consequently, automated analysis of retinal vascular features has attracted considerable attention in medical imaging diagnostics. Clinically, the morphological distinctions between retinal arteries and veins indicate disease progression, and accurate arteriovenous segmentation enhances diagnostic accuracy while supporting personalized therapeutic strategies. Fueled by advances in computer vision and deep learning, automated segmentation of retinal vasculature has become a key research area in intelligent healthcare. In particular, arteriovenous segmentation poses substantial challenges due to the need for precise discrimination of subtle morphological variations, while offering critical clinical value for informed diagnostic decision-making.

Although deep learning techniques have achieved remarkable progress in image segmentation in recent years[2], directly applying these methods to retinal artery and vein segmentation remains highly challenging. First, the anatomical structure of retinal vessels is extremely intricate, with arteries and veins exhibiting subtle differences in color, width, and orientation. These distinctions become particularly ambiguous in small vessel branches, where boundaries are often blurred and vessels may intersect or merge with the background, thereby significantly increasing segmentation difficulty. Second, the quality of retinal images is highly susceptible to variations in acquisition devices and environmental conditions, often resulting in blur, noise, or low contrast, which impedes the representation of vascular features and compromises model robustness. Moreover, AV segmentation requires not only precise extraction of local features, such as color and texture, but also effective modeling of the global vascular structure and topological relationships. However, existing methods still encounter limitations in multi-scale feature fusion, making it difficult to preserve fine details while simultaneously capturing global semantic context. Additionally, AV segmentation often depends on prior vessel segmentation results, and errors in initial segmentation can propagate through subsequent stages, leading to accumulated inaccuracies and degraded final performance. Therefore, developing a segmentation framework capable of robust multi-scale feature perception, effective error mitigation, and strong generalization remains a critical and unresolved challenge.

To address the aforementioned challenges, this paper introduces a novel Dynamic Iterative Multi-scale Attention Network (DIMAN) that is specifically designed for efficient and accurate segmentation of retinal arteries and veins. The main contributions of this work are summarized below:

1) A DIMAN network is introduced, integrating U-Net with novel attention and convolution modules to build a deep neural network capable of multi-scale feature perception

Manuscript received June 17, 2025; revised August 29, 2025.

Yuanyuan Li is a postgraduate student at School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China (e-mail: 2248639409@qq.com).

Jiansheng Wu is a professor of University of Science and Technology Liaoning, Anshan, 114051, China (e-mail: ssewu@163.com).

Yang Liu is an associate professor of University of Science and Technology Liaoning, Anshan, 114051, China (corresponding author, e-mail: liuyang_lnas@163.com).

and iterative refinement, thereby enhancing the accuracy and robustness of retinal arteriovenous segmentation.

2) The designed DMCA module introduces a novel learnable mechanism for dynamically adjusting local and global weights, effectively enhancing fine-grained feature representation and improving the model's adaptability to complex vascular topologies.

3) The proposed EPConv module integrates partial convolution with batch normalization and ReLU activation, significantly reducing computational overhead while preserving high performance, thereby providing an efficient solution for high-resolution medical image analysis.

II. Related Work

A. Traditional Methods for Retinal Arteriovenous Segmentation

In the study of retinal arteriovenous vessel segmentation, traditional methods primarily relied on the color, morphology, topology, and contextual information to construct features, which were then integrated with graph models or classifiers for vessel recognition. Sathananthavathi et al.[3] proposed a segmentation method that used Bat Algorithm-optimized feature selection combined with a Random Forest classifier, significantly reducing dimensionality and improving segmentation accuracy by fusing color space and luminance models to construct efficient feature vectors. Xu et al.[4], by contrast, enhanced segmentation accuracy through intra-image regularization, inter-subject normalization, and multi-texture feature extraction in combination with a kNN classifier. Dashtbozorg et al.[5] introduced a graph-structure-based segmentation method analyzing node types in the vascular network, such as bifurcations and intersections, and combined this analysis with intensity features to automatically segment vascular segments. Vázquez et al.[6] developed a minimal path-based technique using local color clustering and cross-radius vascular segment tracking, incorporating a voting mechanism to enhance segmentation robustness. Zhao et al.[7] employed an explicit set clustering approach to reconstruct and segment vascular networks by building a vascular graph and clustering based on similarities in node features such as intensity, curvature, and diameter. Yan et al.[8], on the other hand, proposed a context-dependent method combining the B-COSFIRE filter with morphological and contextual features to finely segment blood vessels using JointBoost and Conditional Random Fields. Pellegrini et al.[9] extended their approach to Ultra Wide Field of View Scanning Laser Ophthalmoscope (UWFoV SLO) images, proposing a graph model that integrates local feature segmentation with global graph cut optimization to achieve fully automated arterial and venous segmentation on this image type for the first time.

B. Deep Learning Methods for Retinal Arteriovenous Segmentation

In deep learning research on retinal arteriovenous vessel segmentation, a variety of novel approaches have emerged in recent years. Yi et al.[10] proposed MMF-Net, which enhanced vascular structural features via multi-channel frequency filtering, combined with multi-scale transformation and a feature fusion module, significantly improving segmentation accuracy. Zhang et al.[11] introduced a generative adversarial network based on an improved U-Net architecture, integrating the ASPP module to expand the receptive field and optimizing vessel pixel weights through an attention mechanism for end-to-end automatic segmentation. To address computational efficiency, Diwakar et al.[12] designed a cascaded supervised mini-U-Net that utilized multispectral information from bimodal images, combining color and monochrome wavelength inputs to segment vessels and simultaneously distinguish arteries and veins through a cascaded three-phase network. Containing only one eighty-seventh of the parameters of traditional U-Net, it achieved a balance between high accuracy and low resource consumption. Cui et al.[13] developed MGA-Net to improve segmentation robustness by enhancing local features through a Feature Augmented Channel Attention module and capturing long-range dependencies via a Global Feature Aggregation module. In the domain of data augmentation, Liu et al.[14] proposed VSG-GAN, which decoupled vascular topology from background style using a hierarchical variational autoencoder, combined with a SPADE module to generate high-fidelity synthetic images and enhance model generalization. For retinal vessel segmentation, Zhang et al.[15] utilized U-Net variants, including a spatial attention-enhanced architecture, to optimize feature extraction and vascular boundary delineation via adaptive receptive fields and channel refinement. Chen et al.[16] introduced EAV-Net, which fused multi-scale convolutional blocks with dense color-invariant feature learning to effectively address mis-segmentation caused by illumination and contrast variations. Additionally, Patil et al.[17] combined CNNs with graph optimization, using the minimum spanning tree algorithm to optimize label propagation within vascular networks and reduce topological errors, while Bhimavarapu et al.[18] integrated multi-scale filtering, a parallel channel attention mechanism, and CNNs to jointly optimize segmentation. Collectively, these methods advanced retinal arteriovenous segmentation by enhancing feature representation, fusing modalities, reducing computational cost, and refining vascular boundaries, offering diverse technical solutions for clinical applications.

III. Method

A. Network architecture

The Dynamic Iterative Multi-scale Attention Network (DIMAN) is a novel deep learning architecture, whose overall structure is presented in Fig. 1. Its core idea is to initially extract features using a basic U-shaped network and subsequently apply a dynamic attention mechanism alongside an iterative refinement process to progressively capture multi-scale contextual information, thereby improving segmentation accuracy and robustness. DIMAN mainly consists of two key components: a basic U-shaped subnetwork and a dynamic multiscale refinement subnetwork.

The foundational U-shaped subnetwork module utilizes a standard U-Net architecture, which processes input images to generate initial segmentation outputs. As illustrated in Fig. 2, the U-Net comprises an encoder and a decoder: the encoder extracts hierarchical features through convolution and down-sampling operations, while the decoder reconstructs spatial

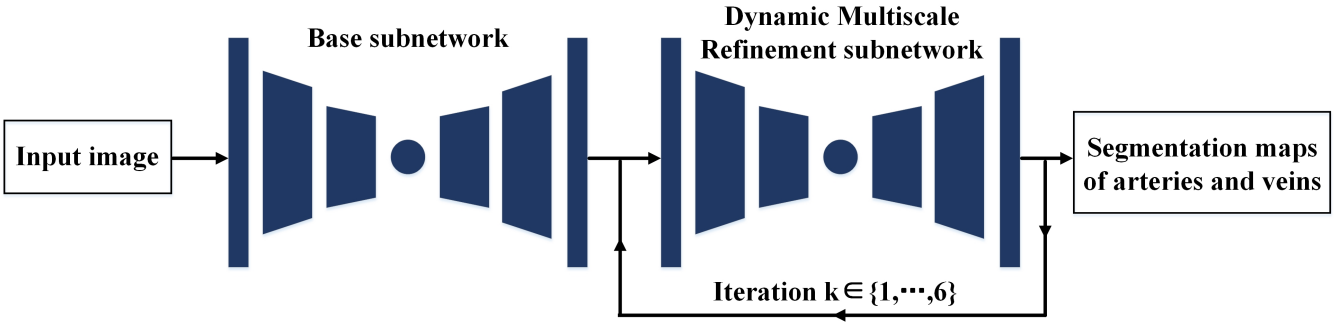


Fig. 1. Overall Architecture of the Proposed Network

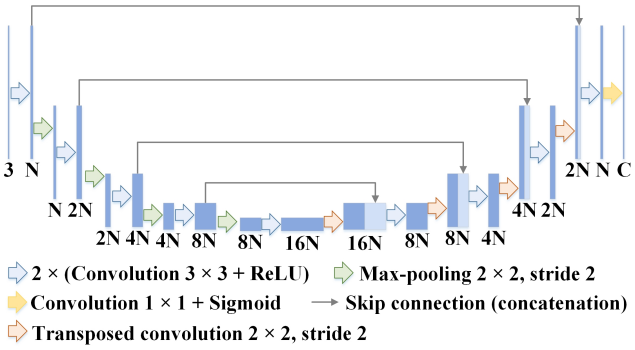


Fig. 2. Basic U-Net module

resolution using upsampling and convolution layers. This base module produces initial feature representations and segmentation maps that serve as the foundation for subsequent iterative refinement, ultimately enhancing the network's overall performance.

The Dynamic Multiscale Refinement (DMR) module constitutes the core innovation of DIMAN, with its structure illustrated in Fig. 3. Built upon the U-Net architecture, this module introduces three key enhancements: the Dynamic Mixed Channel Attention (DMCA) mechanism, Enhanced Partial Convolution (EPConv), and a multi-iteration strategy. The DMCA mechanism is integrated into the skip connections of the DMR module, enabling adaptive fusion of local and global features to improve the perception of multi-scale vascular structures. EPConv is applied within both the encoder and decoder, effectively balancing computational efficiency and representational capacity by selectively processing a subset of channels while preserving the information from the remaining channels. The DMR module is iteratively applied up to six times, allowing the network to progressively refine segmentation outputs, particularly in challenging regions containing fine vessels and complex intersections. Through repeated refinement, the network gradually improves its structural understanding, leading to more accurate segmentation results.

The overall workflow of DIMAN is structured as follows. Initially, the input image is processed by the base U-Net module to generate a coarse vessel segmentation map. This preliminary output is subsequently fed into the Dynamic Multiscale Refinement (DMR) subnetwork for iterative refinement. In each iteration, the output from the previous stage is combined with the original segmentation map to construct the input for the next iteration. Through this recurrent refinement process, the network progressively improves segmentation quality, particularly in regions characterized by complex vascular structures.

Overall, DIMAN provides a robust and efficient solution for retinal arteriovenous vessel segmentation, significantly improving segmentation accuracy and serving as a reliable tool for clinical diagnosis. Its architectural design and innovative modules also demonstrate considerable potential for broader applications in other medical image analysis tasks.

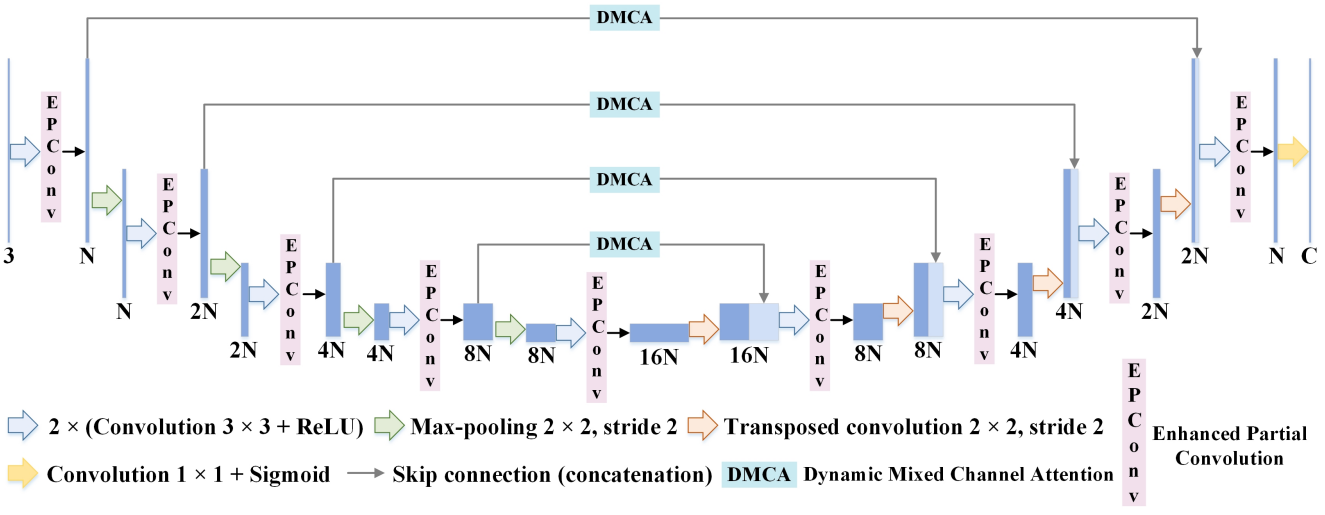


Fig. 3. Dynamic Multiscale Refinement module

B. Dynamic Mixed Channel Attention (DMCA) Mechanism

In this paper, we introduce a novel Dynamic Mixed Channel Attention (DMCA) mechanism, whose architecture is illustrated in Fig. 4. The DMCA employs a dual-branch design to effectively integrate local spatial features and global channel information. It incorporates learnable dynamic weight parameters, enabling adaptive modulation of feature responses via a sigmoid activation function. Unlike conventional channel attention mechanisms limited by single-scale modeling, DMCA addresses the challenges posed by the complex topology, multi-scale coexistence, and low-contrast characteristics commonly found in medical images. By dynamically balancing local detail enhancement and global semantic understanding, the proposed method significantly improves the model's sensitivity to subtle variations. This architecture establishes a stable and generalizable feature extraction paradigm, making it well-suited for the analysis of diverse medical imaging modalities.

The core idea of DMCA is to enhance the model's ability to recognize vascular structures at different scales by considering both local and global information and dynamically balancing their importance. Specifically, the operation of DMCA is as follows.

First, given an input feature map X, which belongs to $R^{C\times H\times W}$, the DMCA applies a Local Average Pooling (LAP) operation to capture local spatial information. As shown in Fig. 5, LAP pools the input feature map using a $k \times k$ sliding window (in this case, $k = 2$) to generate a local feature representation L, which belongs to $R^{C\times ks\times ks}$, where ks is the preset number of region divisions (usually set to 5). The LAP operation effectively preserves local structural information, which is critical for detecting fine vessels and local patterns.

Next, DMCA applies Global Average Pooling (GAP) to extract global context information. As shown in Fig. 5, GAP computes the mean of the entire feature map across the spatial dimensions, producing a single value for each channel, thereby generating a global semantic representation. The GAP operation effectively captures the global context of the feature map, facilitating a better understanding of the overall vessel distribution and large-scale structures.

Local and global features are first reshaped and then processed with 1D convolution operations, respectively, to capture inter-channel dependencies. The convolution kernel size k for the 1D convolution is dynamically determined[19] based on the number of input channels C, as shown in Equation (1):

$$k = \Phi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{\text{odd}} \tag{1}$$

where C is the number of channels, k is the size of the convolutional kernel, and γ and b are hyperparameters, both set to a default value of 2. To ensure that the kernel size is odd, 1 is added to k if it is initially even. This adaptive kernel-size design allows the model to efficiently capture inter-channel dependencies across different network layers and input resolutions.

In the design of the DMCA, the Unpooling of Average Pooling (UNAP) operation plays a crucial role, as it is not only used to resize the attention map but also facilitates the fusion of local and global features, as shown in Fig. 5. To achieve effective integration of information at different scales, DMCA introduces a dynamic weighting mechanism

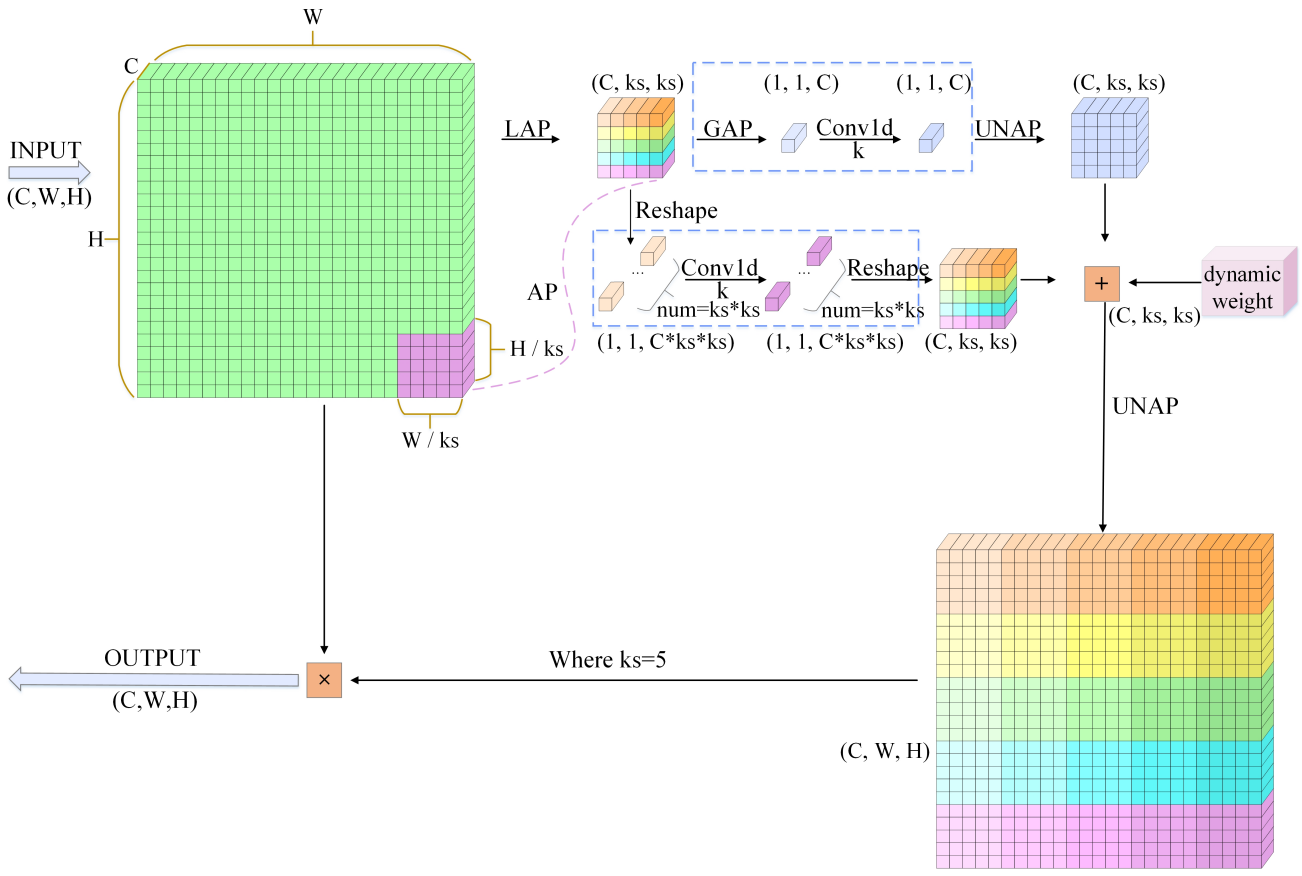


Fig. 4. Overview of the DMCA mechanism

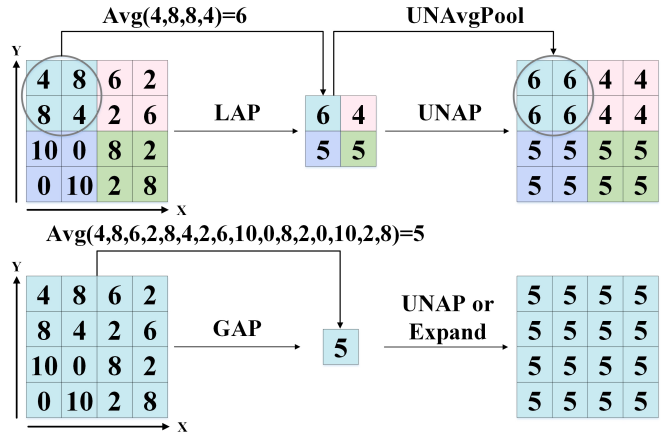


Fig. 5. Schematic diagram of LAP, GAP, UNAP

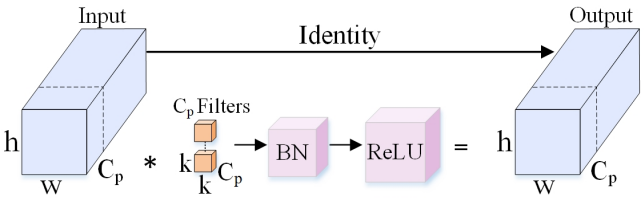


Fig. 6. Overview of the EPConv

that adaptively balances the contributions of local and global attention using learnable parameters. This mechanism enables the network to dynamically adjust the influence of each attention branch according to the semantic and structural requirements of the input features. A sigmoid function is subsequently applied to generate the corresponding local and global attention weights. This is achieved through a learnable parameter l, which is processed by the sigmoid function to ensure that the weight values lie within the range [0, 1]:

$$w = \sigma(l) \tag{2}$$

where σ represents the sigmoid function. The final attention map A is obtained by dynamically fusing the local and global attention maps with learned weights:

$$A = G \times (1 - w) + L \times w \tag{3}$$

where G is the global feature, L is the local feature, and w is the dynamic weighting value. This dynamic weighting enables the model to adaptively modulate the contributions of local and global information based on the input features, thereby improving its ability to handle diverse vessel structures and image characteristics.

Finally, through an Unpooling of Average Pooling (UNAP) operation, the DMCA resizes the fused attention maps A' to match the dimensions of the original input feature maps and performs element-wise multiplication with the original input features:

$$Y = X \times A' \tag{4}$$

In this way, the DMCA effectively fuses and dynamically balances local and global information. The UNAP operation plays a crucial role in this process, being used not only to align the global attention dimensions with the local attention but also to resize the final fused attention map to match the dimensions of the original input. This design enables DMCA to efficiently capture multi-scale features while maintaining computational efficiency, thereby enhancing performance in complex medical image analysis tasks.

C. Enhanced Partial Convolution (EPConv)

This subsection introduces a novel convolution operation, Enhanced Partial Convolution (EPConv). Fig. 6 illustrates its structure and operational mechanism. The method is designed to improve both the performance and efficiency of deep neural networks in processing complex medical images. The core idea of EPConv is to selectively process a subset of input feature map channels while integrating Batch Normalization (BN) and ReLU activation, thereby enhancing the model's representational capacity and training stability.

EPConv works as follows: given an input feature map X belongs to $\mathrm{R}^{\mathrm{C\times H\times W}}$, where C is the number of channels, and H and W are the height and width of the feature map, respectively, it first splits the channels of the input feature map into two parts, one of which is processed while the other is kept unchanged. For the channels to be processed, EPConv applies a standard convolution operation, immediately followed by batch normalization and ReLU activation. The unprocessed channels are retained as is. Finally, EPConv concatenates the processed and unprocessed channels along the channel dimension to produce the final output.

The introduction of batch normalization and nonlinear activation functions is a key innovation of EPConv, and this design offers several important advantages. First, by processing only a subset of channels, EPConv significantly reduces computational cost and parameter count, thereby improving model efficiency. This is particularly important for processing high-resolution retinal vessel images, as it enables the construction of deeper networks under limited computational resources. At the same time, retaining some unprocessed channels preserves the original input information within the network, which is crucial for capturing the subtle features of retinal vessels. For example, retaining original color and texture information may be critical for distinguishing fine arteries from veins. Second, batch normalization mitigates the problem of internal covariate shift. In deep neural networks, the distribution of inputs at each layer varies with the parameters of previous layers, a phenomenon known as internal covariate shift. Batch normalization stabilizes this distribution by normalizing the inputs of each layer to have zero mean and unit variance. This not only accelerates training but also enables the use of higher learning rates, further improving training efficiency. This property is particularly important when dealing with retinal vessel images, which often exhibit significant variations in brightness and contrast. Third, the ReLU activation function introduces nonlinearity, thereby enhancing the representational capacity of the model. Nonlinear activation allows the network to learn more complex feature representations, which is essential for capturing the intricate morphology and topology of the retinal vasculature. An additional advantage of ReLU is the simplicity of its derivatives, which helps mitigate the vanishing gradient problem and supports more stable and efficient network training.

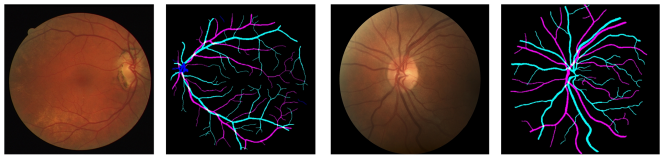
Overall, EPConv provides an efficient and practical solution for the task of retinal arteriovenous vessel segmentation

by integrating partial processing, batch normalization, and nonlinear activation in its structure. This design not only improves the accuracy of the model but also maintains computational efficiency, making it a powerful tool for analyzing complex medical images. With further research and optimization, EPConv is expected to play an important role in advancing medical image analysis and supporting the early diagnosis of related diseases.

IV. Experiments

A. Datasets

In this study, we utilized the RITE dataset[20] and the LES-AV dataset[21] to evaluate the performance of our proposed method. The RITE dataset comprises 40 image sets, evenly divided into training and testing subsets. Each set contains a fundus photograph, a vascular reference standard, and an arterial/venous reference standard. All images have a resolution of 565×584 pixels and are derived from the DRIVE dataset[22], ensuring consistent quality and structure. The LES-AV dataset consists of 22 fundus images with vessels manually annotated as arteries and veins by experts, providing high-quality labels with strong clinical relevance. Representative examples from both datasets, along with their arterial and venous annotations, are shown in Fig. 7.



(a) RITE dataset **(b) LES-AV dataset**

Fig. 7. Examples of fundus images from two datasets

B. Implementation details

The experiments were conducted on a single NVIDIA GeForce RTX 3060Ti GPU using the PyTorch deep learning framework and evaluated with 4-fold cross-validation on the RITE and LES-AV datasets. To enhance model generalization, various data augmentation techniques were applied, including stochastic affine transformations, HSV adjustments, flipping, and cropping. DIMAN was trained using the Adam optimizer with a learning rate of 1e-4 and a batch size of 1 for 20,000 epochs. Mixed-precision training was employed to accelerate convergence and reduce memory consumption. Multi-process training was adopted, with each fold trained in a separate process, and early stopping was used to prevent overfitting. For reproducibility, a fixed random seed was set. During training, GPU resources were fully utilized with CUDA acceleration and cudnn.benchmark enabled, and memory usage was optimized. The training results for each fold, including optimal model weights, configurations, and logs, were saved separately for subsequent analysis.

C. Evaluation metrics

To objectively evaluate the performance of the proposed method for retinal arteriovenous vessel segmentation, three widely used evaluation metrics were adopted: sensitivity, specificity, and accuracy. These metrics collectively reflect the model's discriminative ability and robustness, providing a reliable basis for performance assessment. In the binarized segmentation map, pixels correctly identified as vessels are denoted as True Positives (TP), while those incorrectly identified as vessels are designated as False Positives (FP). Pixels correctly identified as non-vessels are labeled True Negatives (TN), while those incorrectly classified as non-vessels are labeled False Negatives (FN). The definitions and significance of these evaluation metrics are detailed below.

Sensitivity (Sen) evaluates the model's ability to correctly identify arterial and venous vessels, reflecting the proportion of true positives among all actual vessels. It is calculated as follows:

$$Sen = \frac{TP}{TP + FN} \tag{5}$$

Specificity (Spe) measures the model's ability to correctly identify non-target regions, such as background areas or vessels of the opposite class, reflecting the proportion of true negatives among all actual non-target pixels. It is calculated as follows:

$$Spe = \frac{TN}{(TN + FP)} \tag{6}$$

Accuracy (Acc) is the most intuitive metric for assessment, but it may not fully reflect the model's performance in cases of class imbalance. It is calculated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

D. Experimental Results and Analysis

(1) Comparative Experimental Analysis

To validate the effectiveness of the proposed DIMAN network for retinal arteriovenous vessel segmentation, we conducted systematic comparative experiments with several existing mainstream methods to evaluate its performance on two publicly available datasets, RITE and LES-AV. The results of the comparison on the RITE dataset were presented in Table I, while those on the LES-AV dataset were shown in Table II.

As shown in Table I, the proposed DIMAN network outperformed other state-of-the-art methods on the RITE dataset in terms of three widely adopted evaluation metrics: sensitivity, specificity, and accuracy. Specifically, DIMAN achieved the best performance across all three metrics, with a sensitivity of 96.04%, a specificity of 97.40%, and an accuracy of 96.50%, demonstrating a clear improvement over existing approaches. In sensitivity, DIMAN surpassed the next best method, TW-GAN, which reached 95.38%, by 0.66%, highlighting its superior capability in detecting vessel pixels, particularly thin or low-contrast ones. In specificity, DIMAN slightly outperformed TW-GAN at 97.20% and NM-GAN-IUNet at 97.25%, demonstrating an enhanced capacity to discriminate vessels from background regions. With respect to accuracy, DIMAN exceeded NM-GAN-IUNet at 96.46% by 0.16% and RRWNet at 96.16% by 0.34%. CMP-OSNet and SegNeXt showed lower sensitivities and accuracies, while Coupling performed close to DIMAN but remained slightly below, emphasizing DIMAN's superior and balanced segmentation performance. Although the gains were modest, they were meaningful for medical image analysis, where

TABLE I
COMPARATIVE EXPERIMENTAL RESULTS FOR THE RITE DATASET

Methods	Sen(%)	Spe(%)	Acc(%)
ISAVC	90.00	90.00	92.30
GenS	71.00	74.00	72.00
Multi-task CNN	93.40	95.50	94.50
FCN	95.13	92.78	93.81
UA-AV	89.00	90.00	89.00
TR-GAN	94.53	96.31	95.46
AVNet	88.63	92.72	90.81
AV-NET	93.70	94.30	94.30
MTLB	87.47	90.89	89.24
SegRAVIR	93.10	94.31	95.13
TW-GAN	95.38	97.20	96.34
WNet	88.86	96.04	92.76
MMF-Net	94.10	93.79	93.95
CMP-OSNet	70.00	95.00	95.00
NM-GAN-IUNet	91.78	97.25	96.46
SPC-Net	93.37	95.37	94.42
RRWNet	94.87	96.37	95.7
Coupling	96.01	97.20	96.47
SegNeXt	93.70	95.60	94.70
Ours	**96.04**	**97.40**	**96.50**

TABLE II
COMPARATIVE EXPERIMENTAL RESULTS FOR THE LES-AV DATASET

Methods	Sen(%)	Spe(%)	Acc(%)
U-Net	86.40	90.44	87.71
DeepLabv3+	89.27	87.94	88.46
UA-Net	88.00	85.88	86.04
UA-AV	88.00	85.00	86.00
AVNet	94.26	90.90	92.19
AV-NET	94.40	94.60	93.60
VC-Net	94.25	94.67	93.46
WNet	86.86	93.56	90.47
RMHAS	94.00	90.00	92.00
MSC-Net	90.30	91.55	90.72
CMP-OSNet	69.00	90.00	92.00
RRWNet	93.06	92.21	93.42
Coupling	94.03	94.14	93.18
DBMAE-Net	89.53	94.20	93.10
Ours	**95.33**	**94.81**	**93.75**

small improvements can enhance clinical decision-making. Methods like TW-GAN and RRWNet achieved competitive results on some metrics but did not consistently outperform DIMAN, with TW-GAN achieving high specificity but lower sensitivity, and RRWNet showing balanced performance yet inferior accuracy and sensitivity.

As presented in Table II, the proposed DIMAN network also achieved outstanding results on the LES-AV dataset, consistently surpassing existing methods in sensitivity, specificity, and accuracy. DIMAN obtained a sensitivity of 95.33%, a specificity of 94.81%, and an accuracy of 93.75%, ranking first across all three evaluation metrics. In terms of sensitivity, DIMAN significantly outperformed most competitive approaches, such as AV-NET and AVNet, which reported 94.40% and 94.26%, as well as Coupling and DBMAE-Net, which achieved 94.03% and 89.53%, demonstrating its superior capability in accurately identifying vessel regions, particularly in challenging conditions such as bifurcations or low-contrast areas. With respect to specificity, DIMAN exceeded the second-best method, VC-Net,

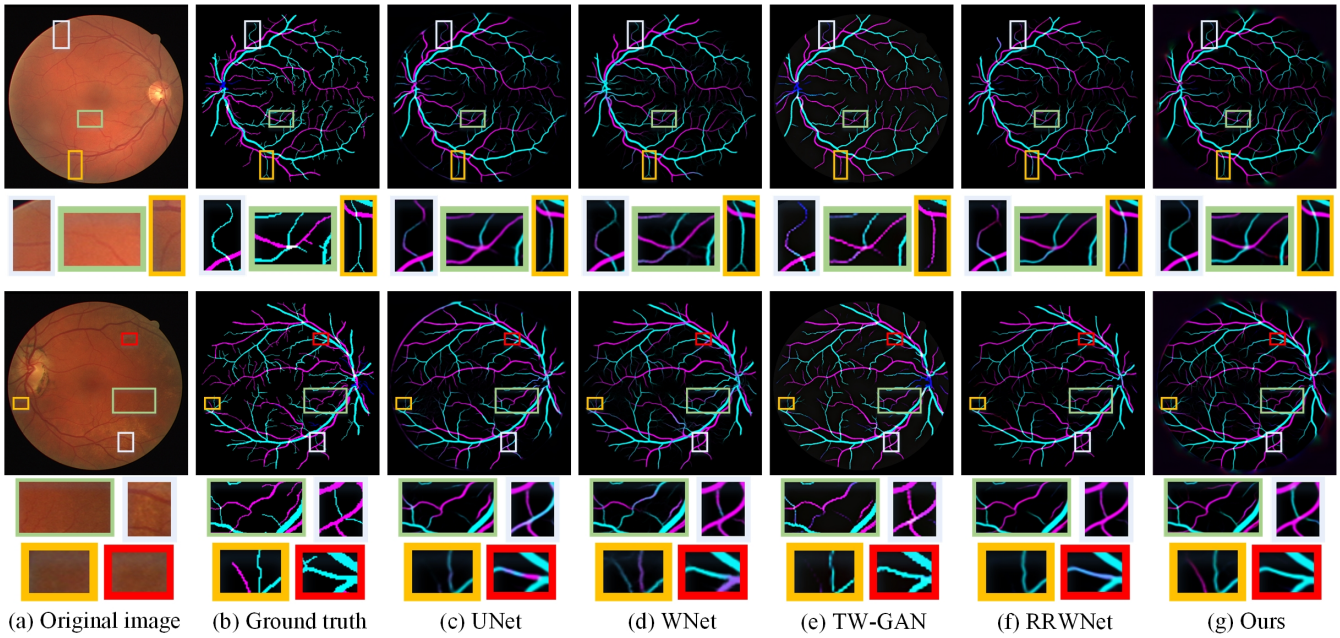


Fig. 8. Comparison of visualization results on the RITE dataset. (a) original fundus image; (b) labeled true values; (c) U-Net; (d) WNet; (e) TW-GAN; (f) RRWNet; (g) DIMAN (our method). Cyan color indicates veins and purple color indicates arteries.

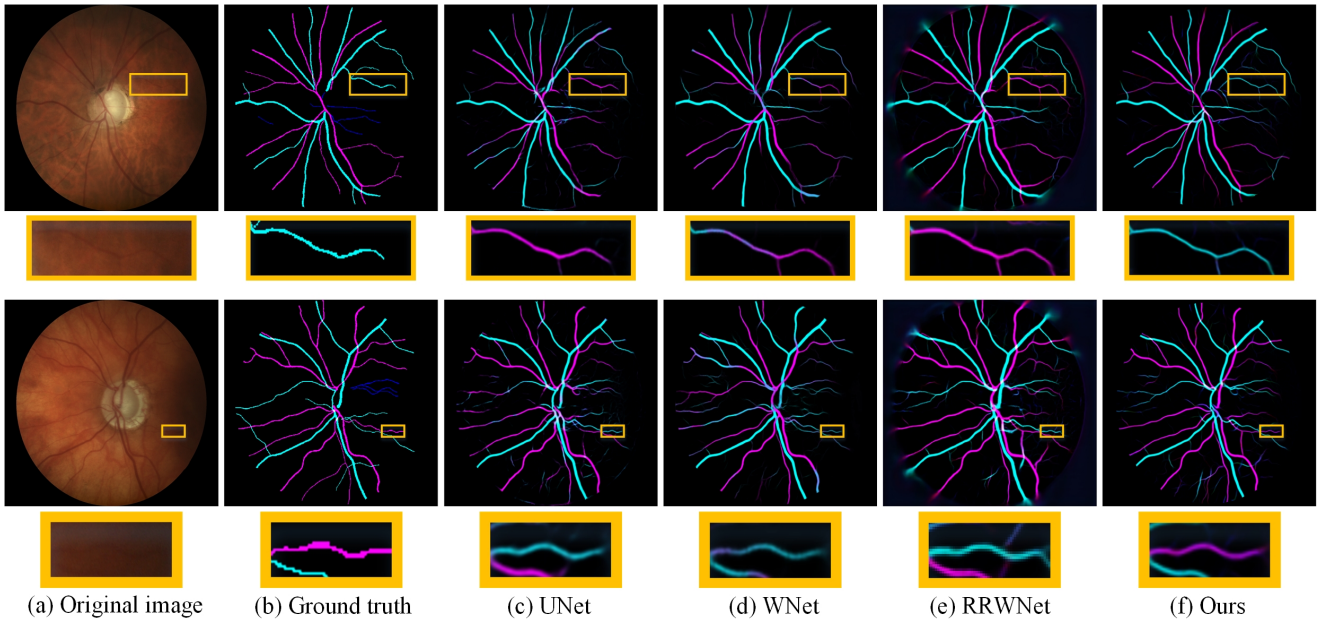


Fig. 9. Comparison of visualization results on the LES-AV dataset. (a) original fundus image; (b) labeled true values; (c) U-Net; (d) WNet; (e) RRWNet; (f) DIMAN(our method). Cyan color indicates veins and purple color indicates arteries.

which achieved 94.67%, and also outperformed Coupling and DBMAE-Net, which obtained 94.14% and 94.20%, indicating its enhanced ability to suppress false positives and avoid misclassification of background areas. Furthermore, DIMAN achieved the highest overall segmentation accuracy of 93.75%, outperforming strong baselines such as VC-Net, AV-NET, Coupling, and DBMAE-Net, highlighting its robust generalization capacity across diverse vessel structures. Several methods demonstrated strong performance in individual metrics; for example, VC-Net achieved the highest specificity, while AV-NET achieved strong accuracy; however, they did not consistently achieve high scores across all evaluation aspects. In contrast, DIMAN maintained balanced and top-level performance, reinforcing its effectiveness and reliability for practical applications.

To visualize the performance advantages of the DIMAN model, we presented its segmentation results on the RITE and LES-AV datasets and compared them in detail with those of other methods for the arterial-venous segmentation task in Figs. 8 and 9. These figures displayed the original fundus images, ground truth annotations, and A/V segmentation results generated by U-Net[23], WNet[24], TW-GAN[25], RRWNet[26], and DIMAN. Visually, DIMAN achieved the best performance on both datasets, particularly in handling fine blood vessels and complex intersection regions. In Figs. 8 and 9, colored boxes highlighted the differences in performance for fine vessel segmentation among the methods. It was evident that DIMAN more accurately recognized and segmented these fine vessels, whereas methods such as U-Net and WNet tended to miss or incorrectly segment these areas. Furthermore, DIMAN effectively preserved vessel continuity and segmentation consistency, especially at crossings and bifurcations, yielding stable segmentation results.

Combining quantitative evaluation and visualization results, the DIMAN network demonstrated excellent performance and good generalization ability in the retinal artery-vein segmentation task. It not only achieved improvements in individual metrics but also realized a comprehensive improvement in three key metrics: sensitivity, specificity, and accuracy. The visualization results further confirmed DIMAN's superiority in handling complex and fine vascular structures. This demonstrated that DIMAN was able to recognize vascular structures more accurately while effectively reducing segmentation errors, providing a more reliable tool for clinical applications. DIMAN's success validated the effectiveness of its innovative architecture, especially the dynamic mixed channel attention mechanism and the enhanced partial convolution module, in processing complex medical images.

(2) Ablation Experimental Analysis

Ablation experiments are a commonly used analytical tool for assessing the impact of individual key modules on a model's overall performance. In this study, we examined performance variations by selectively removing or replacing specific components of the model, thereby revealing the actual contribution of each network component. This approach not only facilitated a deeper understanding of the model architecture but also provided a theoretical foundation for subsequent optimization.

To validate the practical effectiveness of the key modules in the proposed DIMAN, we conducted a series of ablation experiments focusing on two core innovative modules in the model: dynamic mixed channel attention and enhanced partial convolution. These experiments aimed to systematically analyze the specific contribution of each module to the model's performance and further substantiate its critical role in improving the accuracy and robustness of retinal arteriovenous vessel segmentation.

In this experiment, ablation experiments were conducted on the RITE and LES-AV datasets to evaluate the effects of the base model, the DMCA mechanism, the EPConv module, and their combinations. The results of the ablation experiments on the RITE dataset and LES-AV dataset were shown in Table III.

TABLE III
RESULTS OF ABLATION EXPERIMENTS ON RITE AND LES-AV DATASETS

Dataset	DMCA	EPConv	Sen(%)	Spe(%)	Acc(%)
RITE	×	×	94.87	96.37	95.7
	✓	×	95.32	97.16	96.35
	×	✓	95.32	96.78	96.23
	✓	✓	**96.04**	**97.4**	**96.5**
LES-AV	×	×	93.06	92.21	93.42
	✓	×	93.22	92.31	93.62
	×	✓	93.24	92.99	93.58
	✓	✓	**95.33**	**94.81**	**93.75**

As shown in Table III, the ablation experiments on both the RITE and LES-AV datasets demonstrated the contributions of the two modules, DMCA and EPConv, to overall model performance. On the RITE dataset, the baseline model achieved an accuracy of 95.7% without either module. Incorporating the DMCA or EPConv module individually improved the accuracy to 96.35% and 96.23%, respectively, indicating that each module independently enhanced the model's effectiveness. When both modules were integrated, the accuracy further increased to 96.5%, achieving the best performance and confirming the synergistic effect of DMCA and EPConv. In addition, the sensitivity and specificity also reached their highest values of 96.04% and 97.4%, respectively, demonstrating an improved capability to identify fine vessels while reducing false detections.

A similar trend was observed on the LES-AV dataset. The base model achieved an accuracy of 93.42%, which increased to 93.62% and 93.58% with the inclusion of the DMCA and EPConv modules, respectively. The highest accuracy of 93.75% was obtained when both modules were used in combination. Compared with the base model, the addition of DMCA and EPConv not only significantly improved segmentation accuracy but also enhanced the model's generalization and robustness, particularly in specificity, which increased markedly from 92.21% to 94.81%. This notable improvement indicated that the model was more reliable in distinguishing arteries from veins.

In summary, the ablation experiments fully validated the effectiveness of the DMCA module for dynamic feature enhancement and the advantages of the EPConv module in improving feature representation and computational efficiency. Their combination not only consistently improved performance across different datasets but also demonstrated the strong potential of the DIMAN architecture for handling complex vascular structure tasks.

V. CONCLUSION

The proposed Dynamic Iterative Multi-scale Attention Network (DIMAN) provides a novel and effective approach for retinal arteriovenous vessel segmentation. By incorporating the Dynamic Mixed Channel Attention (DMCA) mechanism and the Enhanced Partial Convolution (EPConv) module, DIMAN addresses several critical challenges in retinal vessel segmentation. Experimental results demonstrate that DIMAN achieves outstanding performance on two public datasets, RITE and LES-AV, with significant improvements in segmentation accuracy, sensitivity, and specificity. The ablation studies further confirm the effectiveness of both the DMCA and EPConv modules, which improve performance individually and show significant synergistic effects when combined. Specifically, the DMCA mechanism enhances the model's ability to perceive complex vascular structures by adaptively balancing local and global features, while the EPConv module improves feature representation efficiency without increasing computational cost. These innovations advance the state of the art in retinal vessel segmentation and offer new perspectives for broader medical image analysis. The successful application of DIMAN highlights its potential in handling high-resolution medical images and complex vascular networks, offering a reliable tool for automated retinal analysis in clinical settings. Future work may focus on optimizing the DMCA and EPConv modules, exploring more efficient iterative refinement strategies, and extending DIMAN to other medical imaging domains to facilitate early disease diagnosis and prevention.

REFERENCES

[1] F. Argüello, D. L. Vilariño, D. B. Heras, and A. Nieto, "GPU-based segmentation of retinal blood vessels," *Journal of Real-Time Image Processing*, vol. 14, pp773-782, 2018.

[2] A. A. Abdulsahib, M. A. Mahmoud, M. A. Mohammed, H. H. Rasheed, S. A. Mostafa, and M. S. Maashi, "Comprehensive review of retinal blood vessel segmentation and classification techniques: intelligent solutions for green computing in medical images, current challenges, open issues, and knowledge gaps in fundus medical images," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 10, pp1-32, 2021.

[3] V. Sathananthavathi, and G. Indumathi, "BAT optimization based retinal artery vein classification," *Soft Computing*, vol. 25, no. 4, pp2821-2835, 2021.

[4] X. Xu, W. Ding, M. D. Abràmoff and R. Cao, "An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image," *Computer Methods and Programs in Biomedicine*, vol. 141, pp3-9, 2017.

[5] B. Dashtbozorg, A. M. Mendonça, and A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp1073-1083, 2013.

[6] S. Vázquez et al., "Improving retinal artery and vein classification by means of a minimal path approach," *Machine Vision and Applications*, vol. 24, pp919-930, 2013.

[7] Y. Zhao et al., "Retinal vascular network topology reconstruction and artery/vein classification via dominant set clustering," *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp341-356, 2019.

[8] Y. Yan, D. Wen, M. A. A. Dewan, and W.-B. Huang, "Classification of artery and vein in retinal fundus images based on the context-dependent features," in *Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Ergonomics and Design: 8th International Conference, DHM 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I 8*. Springer, 2017, pp198-213.

[9] E. Pellegrini, G. Robertson, T. MacGillivray, J. van Hemert, G. Houston, and E. Trucco, "A graph cut approach to artery/vein classification in ultra-widefield scanning laser ophthalmoscopy," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp516-526, 2017.

[10] J. Yi, C. Chen, and G. Yang, "Retinal artery/vein classification by multi-channel multi-scale fusion network," *Applied Intelligence*, vol. 53, no. 22, pp26400-26417, 2023.

[11] J. Zhang et al., "End-to-End Automatic Classification of Retinal Vessel Based on Generative Adversarial Networks with Improved U-Net," *Diagnostics*, vol. 13, no. 6, p1148, 2023.

[12] R. K. Diwakar, P. Kumari, P. Saxena, and R. Poddar, "An efficient multitasking cascade network for arteriovenous segmentation using dual-modal fundus images," *Multimedia Tools and Applications*, vol. 83, no. 16, pp48399-48414, 2024.

[13] Y. Cui, J. Zhu, L. Chen, G. Zhang, and S. Gao, "MGA-Net: multiscale global feature aggregation network for arteriovenous classification," *Signal, Image and Video Processing*, vol. 18, no. 8, pp5563-5577, 2024.

[14] J. Liu et al., "VSG-GAN: A high-fidelity image synthesis method with semantic manipulation in retinal fundus image," *Biophysical Journal*, vol. 123, no. 17, pp2815-2829, 2024.

[15] Y.-H. Zhang, J.-S. Wang, and Z.-H. Zhang, "Retinal Vessel Segmentation Algorithm Based on U-NET Convolutional Neural Network," *Engineering Letters*, vol. 31, no. 4, pp1837-1846, 2023.

[16] X. Chen, L. Niu, and S. Guo, "Efficient retinal artery/vein classification with dense color-invariant feature learning," *Neural Computing and Applications*, vol. 37, no. 3, pp1255-1270, 2025.

[17] S. Patil, Y. Talekar, S. Tathe, and S. Takale, "Classification of retinal blood vessels into arteries and veins using CNN and likelihood propagation," *Journal of Integrated Science and Technology*, vol. 13, no. 1, pp1006-1006, 2025.

[18] U. Bhimavarapu, "Retina Blood Vessels Segmentation and Classification with the Multi-featured Approach," *Journal of Imaging Informatics in Medicine*, vol. 38, no. 1, pp520-533, 2025.

[19] 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, pp11534-11542.

[20] Q. Hu, M. D. Abràmoff, and M. K. Garvin, "Automated separation of binary overlapping trees in low-contrast color retinal images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16*. Springer, 2013, pp436-443.

[21] J. I. Orlando, J. Barbosa Breda, K. Van Keer, M. B. Blaschko, P. J. Blanco, and C. A. Bulant, "Towards a glaucoma risk index based on simulated hemodynamics from fundus images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*. Springer, 2018, pp65-73.

[22] 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, pp501-509, 2004.

[23] 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, pp234-241.

[24] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, and I. B. Ayed, "State-of-the-art retinal vessel segmentation with minimalistic models," *Scientific Reports*, vol. 12, no. 1, p6174, 2022.

[25] W. Chen et al., "TW-GAN: Topology and width aware GAN for retinal artery/vein classification," *Medical Image Analysis*, vol. 77, p102340, 2022.

[26] J. Morano, G. Aresta, and H. Bogunović, "RRWNet: Recursive Refinement Network for effective retinal artery/vein segmentation and classification," *Expert Systems with Applications*, vol. 256, p124970, 2024.