MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Liu *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 22 (2026) 466–484

RESEARCH ARTICLE

Boundary Preserving with Attention-Guided Filtering Network for Retinal Vessel Segmentation

Zhihui Liu[a,b], Mohd Shahrizal Sunar[a,b,*], Tian Swee Tan[c,d], Wan Hazabbah Wan Hitam[e]

[a]Faculty of Computing, Universiti Teknologi Malaysia, 81310, UTM Johor Bahru, Johor, Malaysia; [b]Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia; [c]Department of Biomedical Engineering & Health Sciences, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, UTM Johor Bahru, Johor, Malaysia; [d]IJN-UTM Cardiovascular Engineering Centre, Institute of Human Centered Engineering, Universiti Teknologi Malaysia, 81310, UTM Johor Bahru, Johor, Malaysia; [e]Department of Ophthalmology & Visual Science, School of Medical Sciences, Health Campus Universiti Sains Malaysia, 16150, Kubang Kerian, Kelantan, Malaysia

---

Abstract Accurate segmentation of retinal vessels is critical for the early detection of vision-threatening diseases. Although U-Net-based methods have shown strong performance, they often fail to capture thin vessels and preserve boundary details due to repeated downsampling. To overcome these limitations, we propose an enhanced U-shaped network that incorporates a multi-scale attention guided filtering module, allowing the model to retain edge details and suppress noise more effectively. Experiments conducted on the DRIVE, STARE, CHASE_DB1, and HRF datasets demonstrate that the proposed method consistently achieves the best results across multiple metrics. The improvements in F1 score and sensitivity confirm its capability to recover fine vascular structures and its potential for clinical application.

Keywords: Retinal vessel segmentation, deep learning, U-Net, attention guided filtering, medical image analysis.

Introduction

Retinal vessel segmentation plays a critical role in the early detection and diagnosis of various ophthalmic and systemic diseases, such as diabetic retinopathy, hypertension, and glaucoma[1,2]. Accurate vessel delineation provides quantitative biomarkers for computer-aided diagnosis systems and facilitates clinical decision-making[3]. However, the complex morphology of retinal vessels, including thin and tortuous capillaries, bifurcations, and crossings, poses significant challenges for automatic segmentation[4]. Furthermore, variations in image quality caused by uneven illumination, noise, and low contrast frequently lead to discontinuities, missed detections, and inaccurate boundary localization.

Deep convolutional neural networks (DCNNs) have achieved remarkable progress in medical image segmentation tasks, including retinal vessel analysis[5]. Architectures such as U-Net[6] and its variants have become the de baseline for pixel-level prediction. Despite these advances, DCNNs still suffer from a fundamental limitation: the loss of fine structural details due to repeated down-sampling operations. This problem becomes particularly critical in retinal vessel segmentation, where precise boundary preservation and the detection of small vessels are essential. Existing methods often produce blurred boundaries, merge adjacent vessels, or fail to recover the thin structures, thereby limiting their clinical applicability[7,8].

***For correspondence:**
shahrizal@utm.my

Received: 5 Nov. 2025
Accepted: 28 March 2026

Guided filtering has emerged as a promising strategy for enhancing spatial details, as it can effectively transfer structural information from a guidance image to refine the prediction map[9]. Recent works have attempted to integrate guided filtering into deep learning frameworks to improve boundary sharpness[10,11]. However, conventional guided filters rely on fixed local operations and lack adaptive mechanisms to focus on important vessel regions, which constrains their ability to capture complex contextual dependencies.

To address these limitations, we propose a Boundary-Preserving Network with Attention-Guided Filtering (BAGF-Net) for retinal vessel segmentation. Specifically, our method introduces a Boundary Aware Guided Filtering module that leverages structural priors to enhance vessel boundaries and restore spatial details lost in down-sampling. In addition, we design a Spatial Pooling Attention (SPA) mechanism to adaptively highlight informative vessel regions and suppress irrelevant background noise. These components are integrated into a unified network that preserves vessel continuity, improves small-vessel detection, and achieves robust segmentation under varying image qualities.

We evaluate our approach on benchmark datasets, including DRIVE, STARE, CHASE_DB1 and HRF. Extensive experiments demonstrate that our method outperforms existing methods in terms of both quantitative performance and visual quality. Ablation studies further verify the effectiveness of the proposed BAGF and SPA modules in boundary preservation and small-vessel segmentation. The main contributions of this work are summarized as follows:

(1) A Boundary Aware Guided Filtering module is proposed to preserve fine vessel structures and improve spatial detail restoration.

(2) A Spatial Pooling Attention mechanism is introduced to enhance vessel representation by adaptively focusing on relevant regions.

(3) A boundary-preserving segmentation network (BAGF-Net) is developed, which integrates Boundary Aware Guided Filtering and Spatial Pooling Attention into a unified framework and achieves superior performance on multiple retinal vessel benchmarks

Related Work

Retinal Vessel Segmentation

Over the past decades, owing to the outstanding performance of U-Net in image segmentation, numerous variants of U-Net have been proposed for retinal vessel segmentation. Since the use of fixed-shaped convolution kernels restricts the receptive field of CNNs, Jin *et al*. [12] developed DUNet with deformable convolutions, allowing the receptive field to adapt to vessel scale and morphology for more accurate feature extraction. IterNet [13] refine vascular segmentation by leveraging skip connections to propagate prediction maps from shallow layers into deeper layers, leading to progressive enhancement of segmentation quality. In addition, to design lightweight yet effective architectures, Wei *et al*. [14] employed an enhanced genetic algorithm to automatically search for models with improved performance. To mitigate the imbalance between vessel and background regions, Zhang *et al*. [15] developed Bridge-Net, which integrates convolutional neural networks with recurrent neural networks to effectively deliver contextual information. The study proposed a network that incorporates a multiple-dimension attention enhancement block to capture local edge information, a deep guidance fusion block and a cross-pooling semantic enhancement block to strengthen global context, and an adaptive weight learner unit to fuse multi-stage predictions effectively. To overcome the difficulty of simultaneously modeling global context and local structures in retinal images, CoVi-Net integrates convolutional networks with transformers and employs advanced feature fusion modules for enhanced segmentation[16].

Despite these successes, existing methods still struggle with preserving fine vessel boundaries and detecting small, tortuous vessels, which are clinically important but easily lost during down-sampling operations.

Guided Filtering

Guided filtering has attracted increasing attention in retinal image segmentation due to its edge-preserving property. The technique, originally proposed by He *et al*.[17], is a linear operator that transfers structural and boundary information from a guidance image (such as the original fundus image or low-level feature maps) to higher-level features. This mechanism allows the filter to retain sharp edges while effectively suppressing noise. In practice, the guidance image plays a critical role: in edge regions, it constrains the filter to avoid excessive smoothing, while in flat regions it permits more aggressive averaging. With its linear-time complexity, guided filtering strikes a favorable balance between computational efficiency and accuracy.

Building on this concept, several studies have incorporated guided filtering into deep segmentation frameworks. Zhang *et al*. applied it to retinal vessel segmentation and introduced an attention map into the optimization process to improve parameter estimation[9]. Yin *et al*. observed that traditional guided filtering often oversmooths subtle structures in nearly homogeneous regions. To address this limitation, a guided filtering module was designed and embedded into feature maps[11]. By concatenating the filter output with the original feature representation, richer and more robust features were obtained. Moreover, the module was applied after each downsampling and upsampling stage, helping to recover fine vascular details lost during scale transformations.

More recent approaches further extend the capability of guided filtering. For instance, a Guided Filter Network for semantic image segmentation incorporated a purification block to remove irrelevant background noise before passing structural information from low-level features to high-level one s[10]. BGF-Net introduced both channel and spatial boundary-guided modules, followed by a boundary-guided filter within DeepLabv3+ to better preserve structural cues[7]. Similarly, FSG-Net appended a vessel feature-guided fusion block after each decoder path to restore fine vessel structures suppressed by downsampling[18]. Li *et al*. enhanced U-Net skip connections by embedding a Deep Guidance Fusion block, which integrates high-resolution spatial and semantic information, along with a Multi-Dimensional Attention Enhancement block to strengthen local vessel representation and alleviate the degradation of thin vessel segmentation[19].

Despite these advances, most existing methods either lack attention mechanisms within guided filtering or restrict the use of guidance to the decoder stage, which limits fine-grained structural recovery across layers. This motivates the design of boundary-aware guided filtering mechanisms that can learn structural priors while being seamlessly incorporated into neural networks.

Attention Mechanisms in Segmentation

Attention mechanisms have become an important component in deep learning models for semantic segmentation, as they enable networks to selectively emphasize relevant features while suppressing irrelevant background information. By doing so, attention improves feature representation, enhances long-range dependency modeling, and strengthens the discriminative ability of segmentation frameworks.

Channel attention focuses on inter-channel relationships. A representative work is the squeeze-and-excitation (SE) block [20], which adaptively recalibrates feature responses using channel-wise statistics. Building upon this idea, the convolutional block attention module (CBAM) extended attention learning by sequentially applying channel and spatial attention, thereby enabling joint feature refinement across dimensions [21]. In parallel, Wang *et al*. [22] proposed an efficient channel attention mechanism that leverages fast one-dimensional convolution to compute channel weights in a lightweight manner. Spatial attention, on the other hand, highlights significant spatial locations in feature maps. Yu *et al*. [23] designed a spatial attention module that exploits texture and edge information to generate a self-attention matrix. Fu *et al*. [24] proposed a transformer-based spatial attention module that aggregates spatial information across image patches, further enhancing the representation of discriminative features.

Research Motivation

Existing retinal vessel segmentation methods have explored various strategies to improve segmentation performance, including guided filtering techniques and attention mechanisms. Guided filtering is often used to preserve edge information and refine feature representations, while attention mechanisms help networks focus on informative spatial or channel features.

However, most existing approaches treat guided filtering and attention mechanisms independently or integrate them in a limited manner. Conventional guided filtering methods primarily focus on edge-preserving smoothing and are typically applied only at the decoder stage, without explicitly incorporating boundary-aware guidance during feature refinement. In addition, many attention mechanisms rely on single pooling strategies, which limits their ability to capture rich spatial context and highlight vessel boundaries, particularly in low-contrast regions.

To address these limitations, the proposed BAGF-Net introduces a boundary aware guided filtering module enhanced by spatial pooling attention, enabling the network to adaptively emphasize vessel boundaries while suppressing background noise. The BAGF module introduces boundary-sensitive guidance into feature refinement, enabling better preservation of thin vessel structures. Meanwhile, the SPA module captures spatial context through multiple pooling operations to enhance the detection of blurred and low-contrast vessel boundaries. Moreover, the filtering module is integrated at multiple stages of the decoder architecture, allowing structural details to be progressively recovered across different feature levels.

To further clarify the differences between the proposed approach and recent retinal vessel segmentation networks, Table 1 summarizes the architectural differences between the proposed method and several representative guided filtering and attention based retinal vessel segmentation networks.

Table 1. Architectural comparison of guided filtering and attention based retinal vessel segmentation networks.

Method	Backbone	Guided Filtering	Attention	Boundary Awareness	Multi-level Integration	Key Idea
Deep Guidance Network (2020)	CNN	✓	✗	✗	Partial	Use guided filtering to refine segmentation output
Guided Filter Network (2022)	CNN	✓	✗	✗	Decoder	Feature refinement using guidance features
BGF-Net (2024)	DeepLabv3+	✓	✗	✓	Decoder	Boundary guided filtering
EDAE-Net (2024)	U-Net	✗	✓	Partial	Multi-stage	Multi-dimensional attention enhancement
MSTP-Net (2025)	Multi-path CNN	✗	✓	✗	Multi-scale	Multi-scale three-path feature learning
WFL-VNet (2025)	CNN	✗	✓	✗	Full-process	Whole-process vessel localization using ConvNeXt-UNet with residual coordinate attention
Ours	U-Net	✓	✓	✓	Decoder	Boundary-aware attention-guided filtering for structural detail preservation

Methodology

The proposed framework is illustrated in Figure 1. An attention guided filtering module is designed to preserve edge information while suppressing noise in the feature maps. The following subsections provide detailed descriptions of the overall framework, including boundary aware guided filtering and spatial pooling attention.

Network Architecture

Unlike conventional guided filtering approaches that apply filtering as a post-processing step or only at the decoder stage, the proposed BAGF module integrates boundary-aware guidance directly into intermediate feature refinement. In addition, the SPA module is designed to explicitly guide the filtering process by generating spatial attention weights that emphasize vessel boundaries and suppress homogeneous background regions. Therefore, the proposed framework forms a coupled attention–filtering mechanism, where SPA provides adaptive spatial guidance and BAGF performs boundary-preserving feature refinement. This design enables the network to progressively recover structural vessel details across multiple feature levels, rather than relying on a single-stage filtering operation.

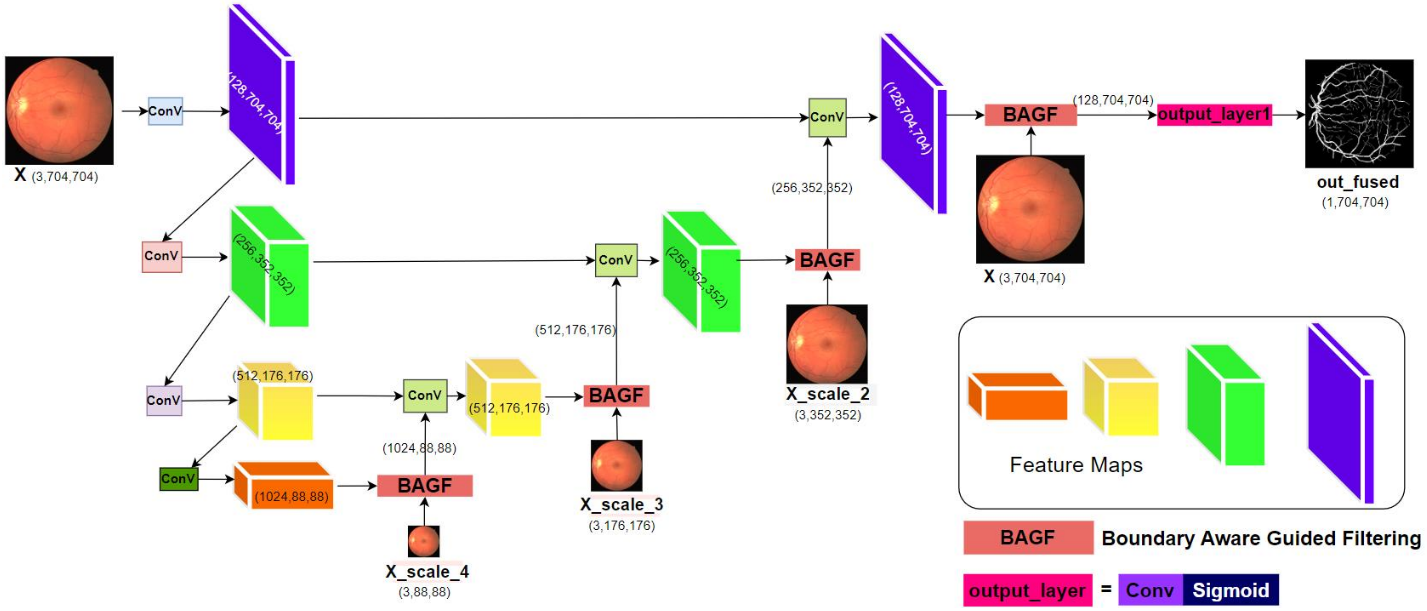


Figure 1. Overall architecture of the proposed network

Our proposed network is composed of a U-shaped backbone, a multi-scale boundary aware guided filtering module, and a spatial pooling attention module, with the RGB image as input. As shown in Figure 1, the proposed Boundary-Aware Guided Filtering (BAGF) module is applied to the intermediate feature maps before they are fused with the decoder features, enabling boundary-aware refinement during multi-scale feature integration. Within each BAGF module, a Spatial Pooling Attention (SPA) mechanism is introduced to generate an attention map from the decoder feature maps. This attention map serves as the guidance for the subsequent guided filtering operation, allowing the network to emphasize vessel boundaries while suppressing background noise. The refined features are then passed to the decoder for further reconstruction. U-Net is a classical fully convolutional architecture that has been widely applied in medical image segmentation. Our framework inherits its symmetric encoder–decoder design. The encoder progressively extracts semantic representations from the input image using convolutional layers and ReLU activations, while the decoder gradually restores spatial resolution through convolution operations. Skip connections are employed to fuse shallow detail features with deep semantic representations, which improves the segmentation of object boundaries and fine structures.

However, repeated downsampling operations may still cause the loss of important spatial details. The multi-scale attention guided filtering module is introduced to address this issue. In retinal vessel segmentation, small and thin vessels are often lost during downsampling, and in optic disc and cup segmentation, the boundary between the two regions tends to be blurry. The module leverages grayscale guidance images to preserve edge-aware structural details by taking both the feature map and the guidance image as inputs and producing a structurally enhanced feature map. It is incorporated after each downsampling and upsampling operation, mitigating the spatial information loss and enhancing robustness against noise.

Boundary Aware Guided Filtering

The Guided Filter was first introduced for image processing tasks[17]. The idea is that the output of the guided filter can be expressed as a locally linear transformation of the guidance image. For instance, [9] incorporated guided filtering into retinal vessel segmentation and further introduced attention-related parameters into the minimization problem. Different from conventional guided filtering used in image processing, the proposed BAGF module introduces two important modifications for retinal vessel segmentation. First, boundary-aware guidance is incorporated through an attention-weighted filtering formulation, allowing the filter to focus more strongly on vessel boundaries and thin structures. Second, the filtering process is embedded within the decoder network and applied to intermediate feature maps rather than the final prediction map. This design allows structural information from the guidance image to be progressively injected during feature learning, improving the recovery of fine vascular details that are often lost during downsampling.

In guided filtering, let G denote the guidance image, I the input image, and O the output image. For a window W_p centered at pixel p (with radius r, i.e., size $(2r+1) \times (2r+1)$), the guided filter assumes that the output O within this window can be represented as a linear transform of G:

$$O_i = a_p G_i + b_p,\ \forall i \epsilon W_p \tag{1}$$

Here, i denotes the pixel index in the guidance image G. The coefficients a_p and b_p are determined by minimizing a local cost function. For retinal vessel segmentation, small details are expected to be preserved while suppressing noise, hence attention block M is introduced. The objective is:

$$min\ E(a_p, b_p) = \sum_{i \in W_p} (M_i^2 (a_p G_i + b_p - I_i)^2 + \varepsilon a_p^2) \tag{2}$$

The first term measures fidelity, enforcing O to approximate the input I. The second term regularizes a_p to avoid instability. ε is a regularization constant, typically very small (e.g., 1e–6). Solving (2) yields:

$$a_p = \frac{E[G \cdot I] - E[G] \cdot E[I]}{E[G^2] - (E[G])^2 + \varepsilon} \tag{3}$$

$$b_p = E[I] - a_p \cdot E[G] \tag{4}$$

Here, $E[I]$ is the mean of I in window W_p, $|W_p|$ is the number of pixels in the window, $E[G^2] - (E[G])^2$ and $E[G]$ denote the variance and mean of G in the window. f_μ denotes the mean filtering operation within a local window of radius r. These values are efficiently computed via box filtering f_μ.

Since each pixel i belongs to multiple overlapping windows, the estimates of a_p and b_p may differ across windows. To resolve this, the standard approach averages the results across all possible windows:

$$O_i = \bar{a}_{p'} G_i - \bar{b}_{p'} \tag{5}$$

Here, $\bar{a}_{p'}$ and $\bar{b}_{p'}$ denote the averages of a_p and b_p across all windows covering pixel i:

$$\bar{a}_{p'} = \frac{1}{|W_p|} \sum_{p \in W_p} a_p \tag{6}$$

$$\bar{b}_{p'} = \frac{1}{|W_p|} \sum_{p \in W_p} b_p \tag{7}$$

Through this averaging strategy, the output image O can retain sharp transitions in G while reducing noise. However, conventional guided filtering is not always optimal for retinal image analysis, as it tends to over smooth fine structures in nearly flat regions. Motivated by this limitation, we design a Boundary Aware Guided Filtering module that integrates attention mechanisms.

This module is appended after each decoder stage of the U-Net architecture, enabling multi-level guided filtering and fine-grained feature fusion. The filtering is applied directly on feature maps so that structural details from the guidance image are injected back into the representation, recovering spatial information lost during downsampling. Moreover, the averaging process further suppresses noise.

As illustrated in Figure 2 (Boundary Aware Guided Filtering), the example demonstrates the operation on the feature maps from the third skip connection. The implementation details are provided in Algorithm 1. Specifically, the original RGB image is used as the guidance image G, since it encodes abundant edge and fine-structure priors. The decoder feature map F serves as the input to be refined. A spatial pooling attention mechanism produces an attention map M (see Section 3.3 for details), which highlights edges and salient regions. This attention map is then introduced into the guided filtering process, weighting the coefficients a and b, thereby improving boundary preservation while suppressing noise in homogeneous regions.

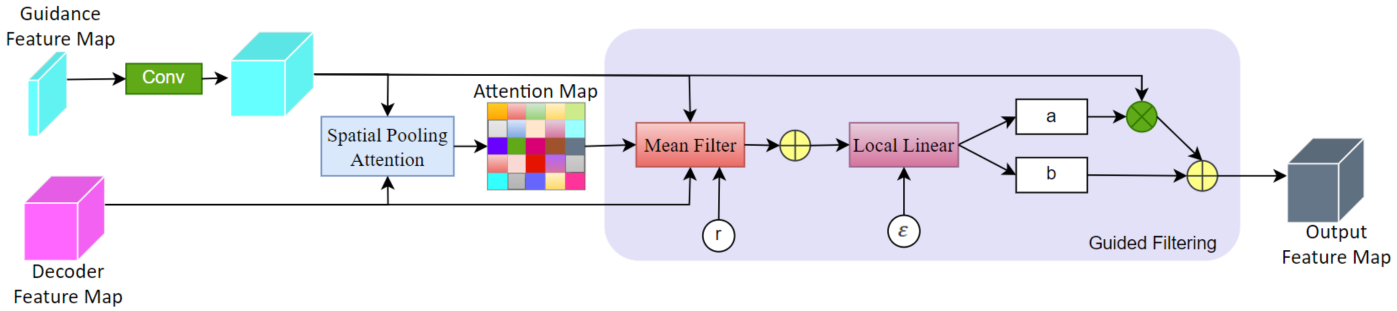


Figure 2. Boundary Aware Guided Filtering module, with feature map dimensions indicated for the third skip connection as an example

```
Algorithm 1: Boundary Aware Guided Filtering (BAGF)
Input   Decoder feature map I
        Guidance map $G'$ (RGB image)
        Attention map M (from Spatial Pooling Attention)
        Radius $r$, regularization $\varepsilon$
Output  Refined feature map O
Generate the guidance feature map by applying a 1 × 1convolution:
    $G = Conv_{1\times1}(G')$
Compute the spatial attention map using the Spatial Pooling Attention module:
    $M = \text{Spatial Pooling Attention}(\mathrm{G}, \mathrm{I})$
Calculate the local mean of the decoder feature map:
    $E[I] = M \cdot f_{\mu}(I, r)$
Calculate the local mean of the guidance feature map:
    $E[G] = M \cdot f_{\mu}(G, r)$
Compute the second-order statistics of the guidance feature map:
    $E[G^2] = M \cdot M \cdot f_{\mu}(G * G, r)$
Compute the correlation between guidance and decoder features:
    $E[G \cdot I] = M \cdot M \cdot f_{\mu}(G * I, r)$
Estimate the linear coefficient $a_p$:
    $a_p = \frac{E[G \cdot I] - E[G] \cdot E[I]}{E[G^2] - (E[G])^2 + \varepsilon}$
Compute the bias term $b_p$:
    $b_p = E[I] - a_p \cdot E[G]$
Smooth the coefficients using mean filtering:
    $\bar{a}_{p'} = f_{\mu}(a_p),\ \bar{b}_{p'} = f_{\mu}(b_p)$
Generate the refined feature map:
    $O = \bar{a}_{p'} G + \bar{b}_{p'}$
Return O
```

Spatial Pooling Attention

To improve the edge-preserving capability of guided filtering, we propose a spatial pooling-based attention mechanism that adaptively highlights vessel boundaries and fine structural details, as shown in Figure 3. Given a guidance feature map $G_{conv} \in R^{H\times W\times C_g}$ derived from the RGB image and a decoder feature $X \in R^{H\times W\times C_f}$, we first construct a joint representation by

$$K = \varphi([G, X]) \tag{8}$$

$$Z = \sigma(BN(f_{1\times1}(ReLU(\mathrm{K})))) \tag{9}$$

Where φ denotes channel concatenation, followed by a transformation comprising nonlinear activation, pointwise convolution, and normalization.

From Z, three complementary spatial statistics are extracted through pooling operators along the channel dimension:

$$u_{max} = MaxPool_c(Z) \tag{10}$$
$$u_{avg} = AvgPool_c(Z) \tag{11}$$
$$u_{min} = MinPool_c(Z) \tag{12}$$

These descriptors encode high-activation responses, global contextual trends, and low-activation background patterns, respectively. To generate the attention distribution, we employ a contrastive fusion strategy:

$$M = u_{max} \oplus u_{avg} \ominus u_{min} \tag{13}$$

where $\oplus$ and $\ominus$ denote additive and subtractive interactions. The resulting map $M \in R^{H\times W}$ emphasizes salient and edge-associated regions while suppressing homogeneous background areas.

Unlike conventional attention mechanisms that rely solely on maximum or average responses, this formulation explicitly incorporates a suppressive term from the minimum pooling, allowing the network to better distinguish subtle and low-contrast vascular structures. In practice, the attention map effectively guides subsequent filtering to preserve vessel continuity and refine boundary details.

Furthermore, this module is computationally lightweight and seamlessly integrable into segmentation frameworks. Its design makes it particularly suited for applications requiring fine-scale boundary recovery under noisy or low-contrast imaging conditions.

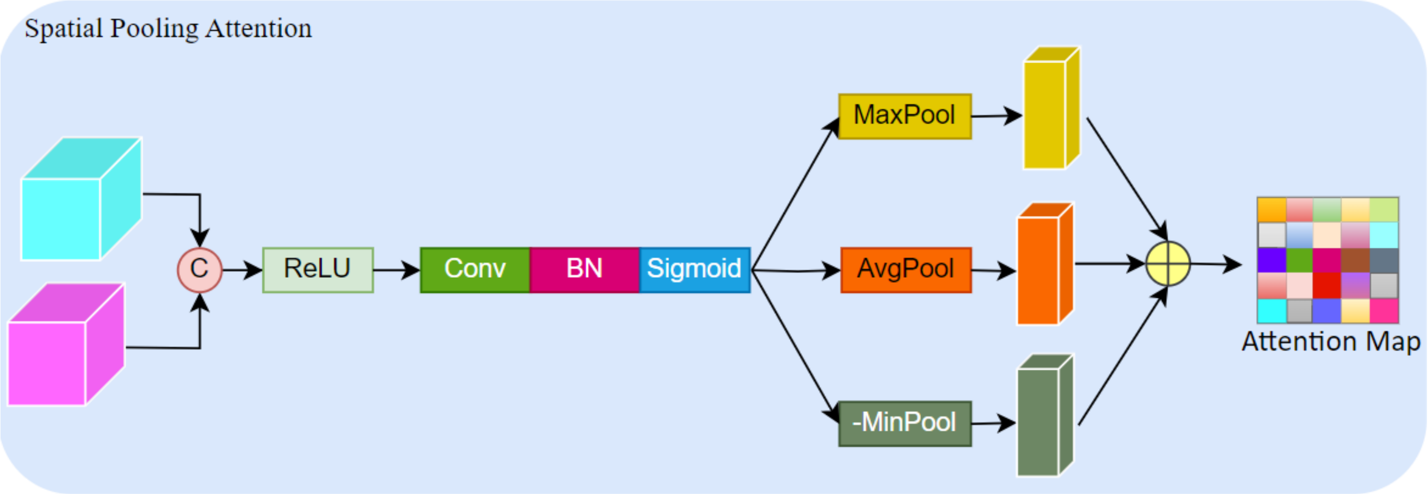


Figure 3. spatial pooling attention mechanism

Compared with conventional spatial attention mechanisms that rely only on average or maximum pooling, the proposed SPA module introduces an additional minimum pooling descriptor to capture low-response background statistics. This contrastive pooling strategy enhances the network's ability to distinguish subtle vessel structures from surrounding tissues. When combined with the BAGF module, SPA provides spatially adaptive guidance for feature refinement, enabling the filtering process to selectively preserve vessel boundaries and suppress noise.

Experiments

Datasets

Our study was conducted on four widely used retinal vessel segmentation datasets: DRIVE [25], STARE [26], CHASE_DB1 [27], and HRF [28]. Their essential characteristics are summarized in Table 2 and Figure 4 illustrates several representative examples.

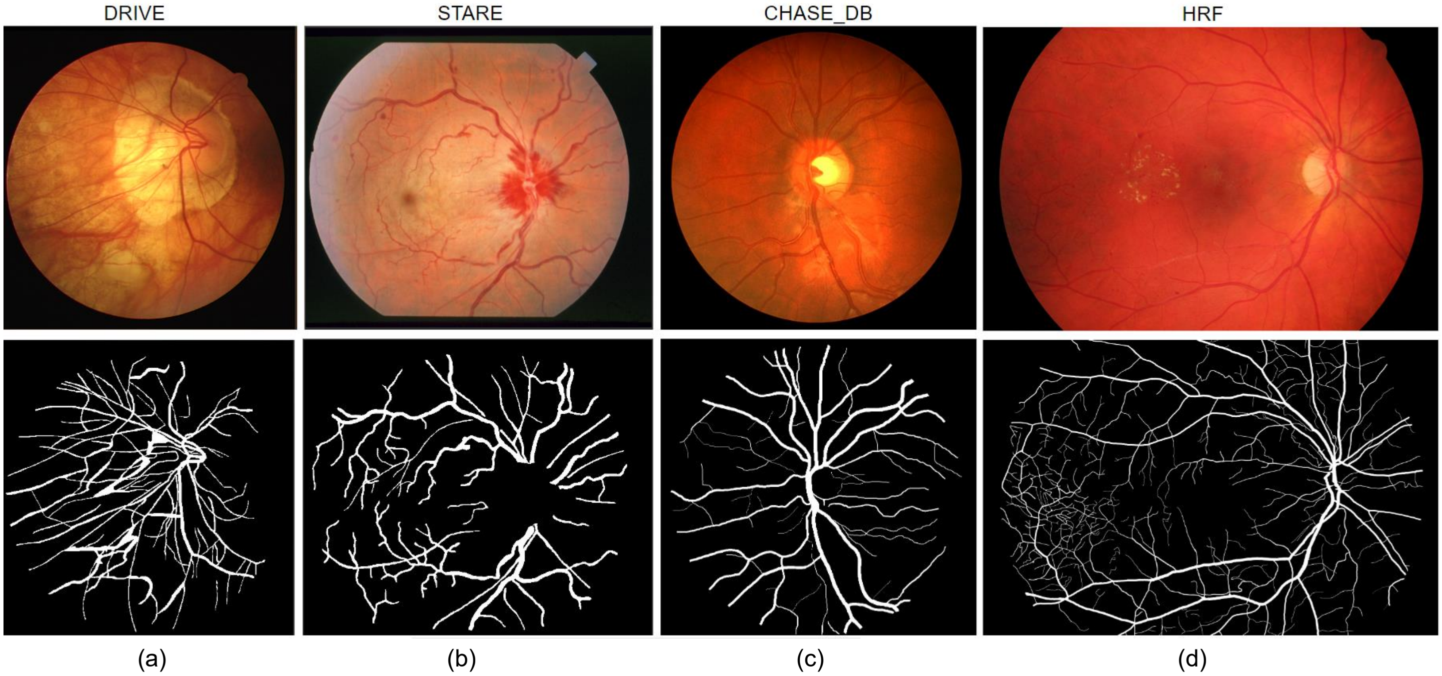


Figure 4. Examples of challenging retinal vessel segmentation. (a) DRIVE dataset, (b) CHASE-DB1 dataset, (c) STARE dataset, (d) HRF dataset. The original images are shown on the up, with their corresponding ground truth annotations displayed on the bottom

Table 2. Details of the retinal vessel segmentation datasets.

Dataset	Year	Modality	Resolution	Image total number	Abnormal Number	Training set/test set
DRIVE	2004	Fundus	565*584	40	7	20/20
STARE	2000	Fundus	605*700	20	10	10/10
CHASE_DB1	2009	Fundus	1280*960	28	0	14/14
HRF	2013	Fundus	3504* 2336	45	30	24/21

The DRIVE dataset contains 40 fundus photographs, including 33 normal cases and 7 with early diabetic retinopathy, captured with a Canon CR5 camera at 565 × 584 resolution; the official split allocates 20 images for training and 20 for testing, with dual annotations available for the test set. STARE consists of 20 images of size 605 × 700, evenly divided between healthy and pathological retinas, obtained using a TopCon TRV-50 device, with vessel masks provided by two experts. CHASE_DB1 includes 28 images from 14 children, acquired with a Nidek NM-200-D camera at 1280 × 960 resolution; two graders performed manual labeling, and the dataset is typically partitioned into 14 training and 14 testing images. The HRF dataset offers 45 high-resolution images (3504 × 2336) from three categories (normal, glaucoma, and diabetic retinopathy) recorded with a Canon CR-1 system; vessel annotations are available from one observer, and we adopted a 22/23 train-test split.

Since the public retinal vessel datasets are relatively small and widely evaluated under fixed train–test splits, we followed the commonly adopted protocol for fair comparison with previous methods. To further reduce the risk of result bias caused by a single split, additional repeated-run statistical analysis was conducted and is reported in Section 5.4.

Preprocessing

All images were standardized to square dimensions to ensure consistent input sizes across datasets. As shown in Table 3, each dataset underwent specific preprocessing strategies—such as center padding, resizing, or a combination of both—while preserving aspect ratios. For example, DRIVE images (565×584) were center-padded to 608×608, and HRF images (3504×2336) were resized to 1344×896 before applying center padding to reach 1344×1344. These tailored operations ensured uniform input dimensions and compatibility with the network architecture.

Table 3. Resolution adjustments and preprocessing strategies for validation and testing datasets

Dataset	Original Resolution	Resized Resolution	Preprocessing Type (Scaling Factor)	Ratio Preserved	Purpose
DRIVE	565×584	608×608	Center padding	Yes	Standardize input size
STARE	605×700	704×704	Center padding	Yes	Standardize input size
CHASE_DB1	999×960	1024×1024	Center padding	Yes	Standardize input size
HRF	3504×2336	1344×1344	Resize(×0.38) + center padding	Yes	Reduce memory cost, standardize input

Augmentation

Due to the limited number of images in retinal fundus vessel datasets, the model is prone to overfitting and may exhibit suboptimal segmentation accuracy. To address this, various data augmentation strategies were employed to expand the dataset and enhance model performance under data-scarce conditions while mitigating overfitting.

Each image is first randomly resized using a scaling factor ranging from 0.5 to 2.0, followed by a random crop to a fixed size to meet the input requirements of the network. Subsequently, a series of stochastic augmentation techniques are applied, including CutMix, horizontal flipping, color jittering, Gaussian blur, and perspective transformation, as illustrated in Figure 5.

Data augmentation was applied only to the training sets, while the validation and test sets remained unchanged to ensure fair evaluation. After augmentation, the size of each training set was expanded to ten times the original size. In particular, the DRIVE training set increased from 20 images to 200 images, the STARE training set increased from 10 images to 100 images, the CHASE_DB1 training set increased from 14 images to 140 images, and the HRF training set increased from 24 images to 240 images.

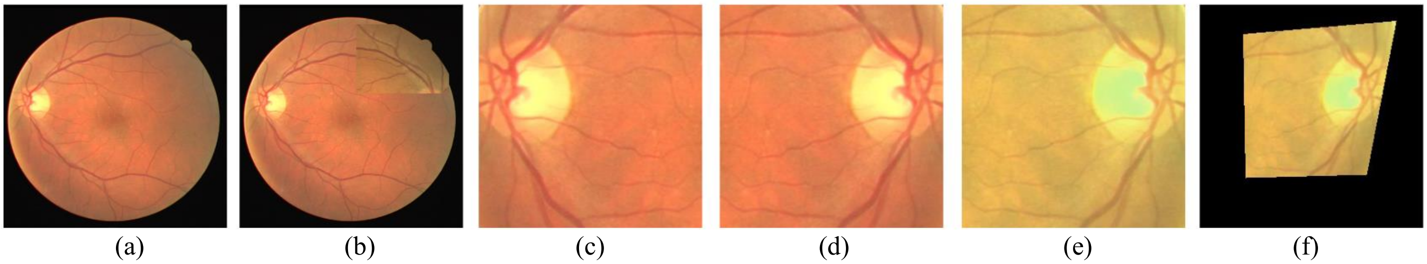


Figure 5. Visualization of training-time data augmentation methods, including (a)raw image, (b) CutMix, (c) cropping, (d) flipping, (e) color jittering and (f) perspective transformation

Implementation Details

All experiments were carried out on a cloud server equipped with an NVIDIA GeForce RTX 4090 GPU (48 GB memory, driver 535.230.02) and an AMD EPYC 7542 CPU, where 16 cores were allocated. The system was configured with 62.9 GB RAM, a 20 GB system partition, and a 50 GB NVMe data disk. The software setup included Ubuntu 20.04, Python 3.8, PyTorch 2.4.1, CUDA 12.1, and cuDNN 8.9.7, while PyCharm 2024.2 served as the development environment. Data loading was parallelized with six worker threads. Model training was performed with a batch size of 4 for up to 1000 epochs. To avoid overfitting and unnecessary training, an early stopping strategy was adopted, where training was terminated when the validation performance did not improve for a predefined number of epochs. The network was optimized using the Adam optimizer with an initial learning rate of 0.001 and a weight decay of 0.05. The neural network is trained from scratch without relying on any pretrained weights, using a batch size of 8.

For the classification layer, the network is trained using the binary cross-entropy (BCE) loss, which measures the discrepancy between the predicted probability map and the ground-truth vessel annotation.

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

where N denotes the total number of pixels in the image, $y_i \in \{0,1\}$represents the ground-truth label of pixel i, and p_i denotes the predicted probability produced by the network after the sigmoid activation. The loss is computed over all pixels of the predicted vessel probability map and optimized using back-propagation during training.

Evaluation Metrics

To quantitatively assess the segmentation performance, four indicators were employed: F1-score (F1), Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE). The F1-score is the harmonic mean of precision and recall, which is particularly effective for class-imbalanced problems. Accuracy measures the overall proportion of correctly classified pixels, while Sensitivity evaluates the fraction of vessel pixels correctly detected. Specificity, in contrast, measures the proportion of background pixels correctly identified as non-vessels. These metrics jointly provide a balanced evaluation of model performance across both vessel and non-vessel classes.The corresponding mathematical definitions are as follows:

$$\mathrm{F1} = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2TP + FN + FP}$$

$$ACC = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$SEN = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$SPE = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

Here, TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Results and Model Performance

Ablation Studies

To assess the effectiveness of the proposed modules, we performed ablation studies on the DRIVE dataset. Table 4 reports the results in terms of F1-score (F1), Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE). All experiments were repeated ten times, and the results are reported as mean ± standard deviation.

To statistically assess the contribution of each module, we conducted the non-parametric Wilcoxon signed-rank test using the F1 scores obtained from the ten independent runs of each configuration. The resulting p-value was used to evaluate whether the performance improvement is statistically significant. A smaller p-value indicates that the improvement in F1 resulting from the inclusion of each configuration is statistically significant. The reason for using the p-value is that it quantifies the likelihood of the observed performance differences occurring by chance. This helps determine whether the performance improvement observed after adding a specific module is primarily due to the module itself, rather than random fluctuations in the data. By setting a significance threshold (typically 0.05), we can infer with a certain level of confidence whether the module has a real impact on the model's segmentation performance. This method allows for objective statistical validation of each module's effect, rather than relying solely on the observed performance metrics.

The baseline U-Net already achieves reasonable segmentation performance with an F1 of 79.35%. Incorporating the Boundary Aware Guided Filtering (BAGF) module without spatial pooling attention (SPA) yields noticeable improvements, especially in sensitivity (80.64% vs. 76.72%), which indicates better detection of thin and low-contrast vessels. The overall accuracy and specificity also increase slightly, confirming that BAGF effectively reduces boundary smoothing while retaining structural details. When SPA is introduced into BAGF, the performance is further enhanced across all metrics. In particular, the F1-score rises to 81.71% and sensitivity improves to 83.43%, demonstrating that SPA provides complementary benefits by adaptively emphasizing vessel boundaries and suppressing background noise. The combination of BAGF and SPA achieves the best balance between sensitivity and specificity, leading to more accurate and stable segmentation results.

Figure 6 illustrates the visual comparison of the ablation study on representative DRIVE images. The baseline U-Net tends to miss thin vessels and produces fragmented vessel maps. The introduction of BAGF improves structural consistency but still leaves some weak capillaries underrepresented. With SPA incorporated into BAGF, the model produces the most complete and continuous vessel structures, with clearer boundaries and fewer false positives in the background.

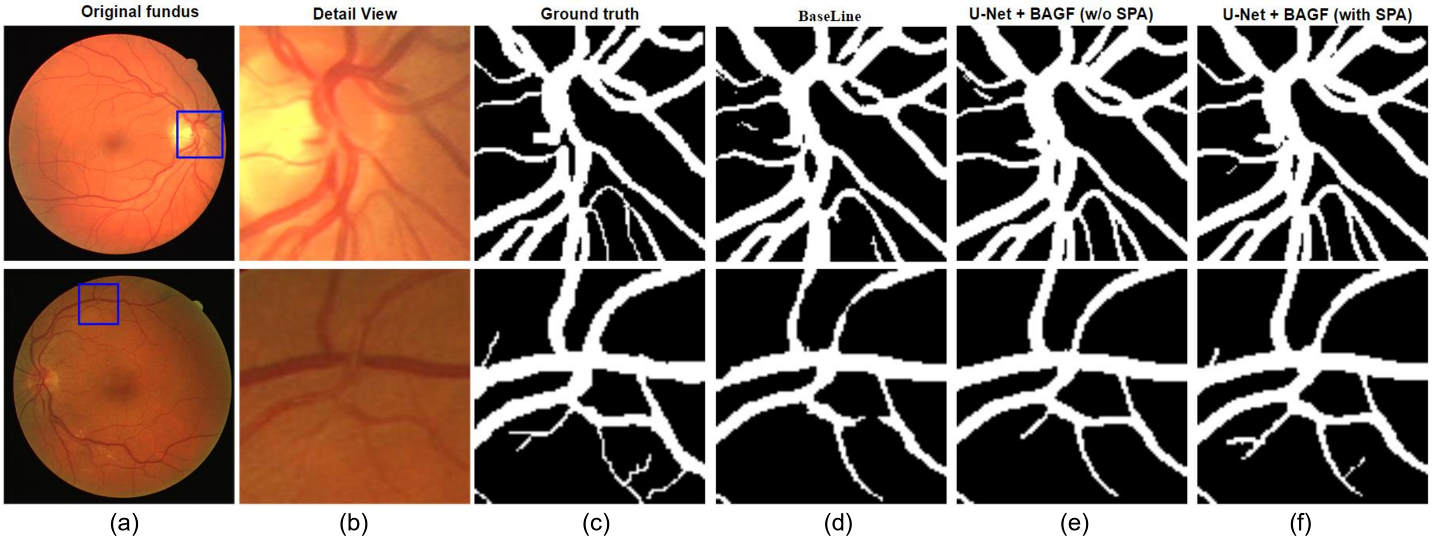


Figure 6. Visual comparison of ablation study on the DRIVE dataset. From left to right: (a) original image, (b) detail view, (c) ground truth, (d) baseline U-Net, (e) U-Net + BAGF (w/o SPA), and (f) U-Net + BAGF (with SPA)

Table 4. Ablation study results on the DRIVE dataset. Best results are highlighted in bold.

Model	F1 (%)	ACC (%)	SEN (%)	SPE (%)	p-value (F1)
Baseline U-Net	79.35 ± 0.15	95.75 ± 0.08	76.72 ± 0.22	98.28 ± 0.05	–
U-Net + BAGF (w/o SPA)	79.80 ± 0.13	96.20 ± 0.07	80.64 ± 0.18	98.40 ± 0.04	<0.05
U-Net + BAGF (with SPA)	**81.71 ± 0.12**	**97.63 ± 0.05**	**83.43 ± 0.16**	**98.53 ± 0.03**	<0.01

Quantitative Comparison with Different Methods

The quantitative comparisons across four public datasets are summarized in Tables 5–8. The results reveal that the proposed method delivers consistent advantages over both the classical U-Net and several recent existing methods.

For the DRIVE dataset (Table 5), our approach achieves 81.72% F1, 97.64% ACC, 83.49% SEN, and 98.52% SPE. The improvement in sensitivity is especially pronounced, showing a gain of +6.63% compared with the baseline U-Net (76.86%). Even against ResMU-Net, which attains 81.49% F1 and 78.43% SEN, our method provides superior performance, underscoring its strength in detecting fine-grained vessels.

Turning to the STARE dataset (Table 6), the proposed model records an F1 of 85.03%, an accuracy of 97.65%, and the best sensitivity of 86.37%. Compared to MSTP-Net (84.68% F1, 97.61% ACC, 86.03% SEN), our method yields modest yet consistent improvements in both F1 (+0.35%) and SEN (+0.34%), demonstrating strong generalization under varying imaging conditions.

On the CHASE_DB1 dataset (Table 7), our network again secures leading results, with 81.42% F1, 85.21% SEN, and the highest specificity (99.10%). These results indicate that the method not only enhances vessel detection but also effectively reduces false positives. Relative to MSTP-Net (80.74% F1, 84.85% SEN, 98.30% SPE), our model achieves balanced gains across all major metrics.

Finally, in the HRF dataset (Table 8), the proposed approach obtains 83.05% sensitivity and 98.68% specificity, together with competitive performance in F1 (81.48%) and accuracy (97.10%). While MSTP-Net reports a slightly higher F1 of 81.08%, our method achieves a substantial sensitivity improvement (+1.97%), which is critical for the wide-field and high-resolution characteristics of HRF images.

In summary, across all four datasets, the proposed network either attains the best performance outright or closely rivals the strongest existing methods. The particularly strong gains in sensitivity highlight its capability to recover thin and low-contrast vessels that are often missed by competing models.

Table 5. Metric comparison of results on public dataset DRIVE (all values in %). Best values are in bold. "–" denotes the corresponding value was not reported in the original paper. Results of the compared methods are taken from their original publications

Method	F1	ACC	SEN	SPE
U-Net(2015) [6]	79.37	95.78	76.86	98.30
SCS-Net(2021) [4]	-	96.97	82.89	98.38
DF-Net(2022) [29]	-	96.14	76.89	98.48
AFNet(2023) [30]	-	95.80	81.39	98.18
BCU-Net(2023) [31]	80.89	96.62	82.38	98.00
ResMU-Net(2024) [32]	81.49	96.85	78.43	98.64
EDAE-Net(2024) [19]	83.04	95.76	82.28	98.32
WFL-Vnet(2026) [33]	**84.74**	97.40	82.53	**98.83**
Ours	81.72	**97.64**	**83.49**	98.52

Table 6. Metric comparison of results on public dataset STARE (all values in %). Best values are in bold. "–" denotes the corresponding value was not reported in the original paper. Results of the compared methods are taken from their original publications.

Method	F1	ACC	SEN	SPE
U-Net(2015) [6]	79.16	95.34	78.40	97.67
DF-Net(2022) [29]	-	97.04	78.58	98.79
AFNet(2023) [30]	-	97.12	79.72	97.79
EDAE-Net(2024) [19]	83.27	96.44	80.36	98.43
MSTP-Net(2025) [34]	84.68	97.61	86.03	98.58
WFL-Vnet(2026) [33]	84.54	**97.86**	84.67	**99.34**
Ours	**85.03**	97.65	**86.37**	98.26

Table 7. Metric comparison of results on public dataset CHASE_DB1 (all values in %). Best values are in bold. "–" denotes the corresponding value was not reported in the original paper. Results of the compared methods are taken from their original publications.

Method	F1	ACC	SEN	SPE
U-Net(2015) [6]	78.32	95.66	77.50	97.84
DF-Net(2022) [29]	-	**98.02**	81.19	98.82
AFNet(2023) [30]	-	96.69	81.94	98.17
MSTP-Net(2025) [34]	80.74	97.45	84.85	98.30
WFL-Vnet(2026) [33]	**83.64**	97.55	82.93	98.95
Ours	81.42	97.14	**85.21**	**99.10**

Table 8. Metric comparison of results on public dataset HRF (all values in %). Best values are in bold. "–" denotes the corresponding value was not reported in the original paper. Results of the compared methods are taken from their original publications.

Method	F1	ACC	SEN	SPE
U-Net(2015) [6]	79.91	96.97	72.19	98.33
SCS-Net (2021) [4]	-	96.87	81.14	98.23
MFA-UNet(2023) [35]	78.60	96.82	77.33	98.45
ResMU-Net(2024) [32]	77.94	97.04	81.56	98.10
MSTP-Net(2025) [34]	81.08	97.05	81.08	98.38
Ours	**81.48**	**97.10**	**83.05**	**98.68**

Visual Comparison

To further demonstrate the effectiveness of the proposed method, qualitative results are presented in Figure 7, which provides visual comparisons across four public datasets: DRIVE, STARE, CHASE_DB1, and HRF. From left to right, each example includes the original fundus image, ground truth, a detailed view, the result of the proposed method, and the U-Net baseline.

As shown in Figure 7, the baseline U-Net often fails to delineate thin vessels and produces blurred boundaries, especially in low contrast areas. Our method generates more continuous and complete vessel structures, with sharper and more precise boundaries that are consistent with the ground truth.

The integration of the Boundary Aware Guided Filtering with Spatial Pooling Attention allows our model to preserve delicate vascular details while effectively suppressing background noise. These improvements are particularly evident in the visualization of capillary networks and bifurcation points, where the proposed method achieves closer alignment with manual annotations.

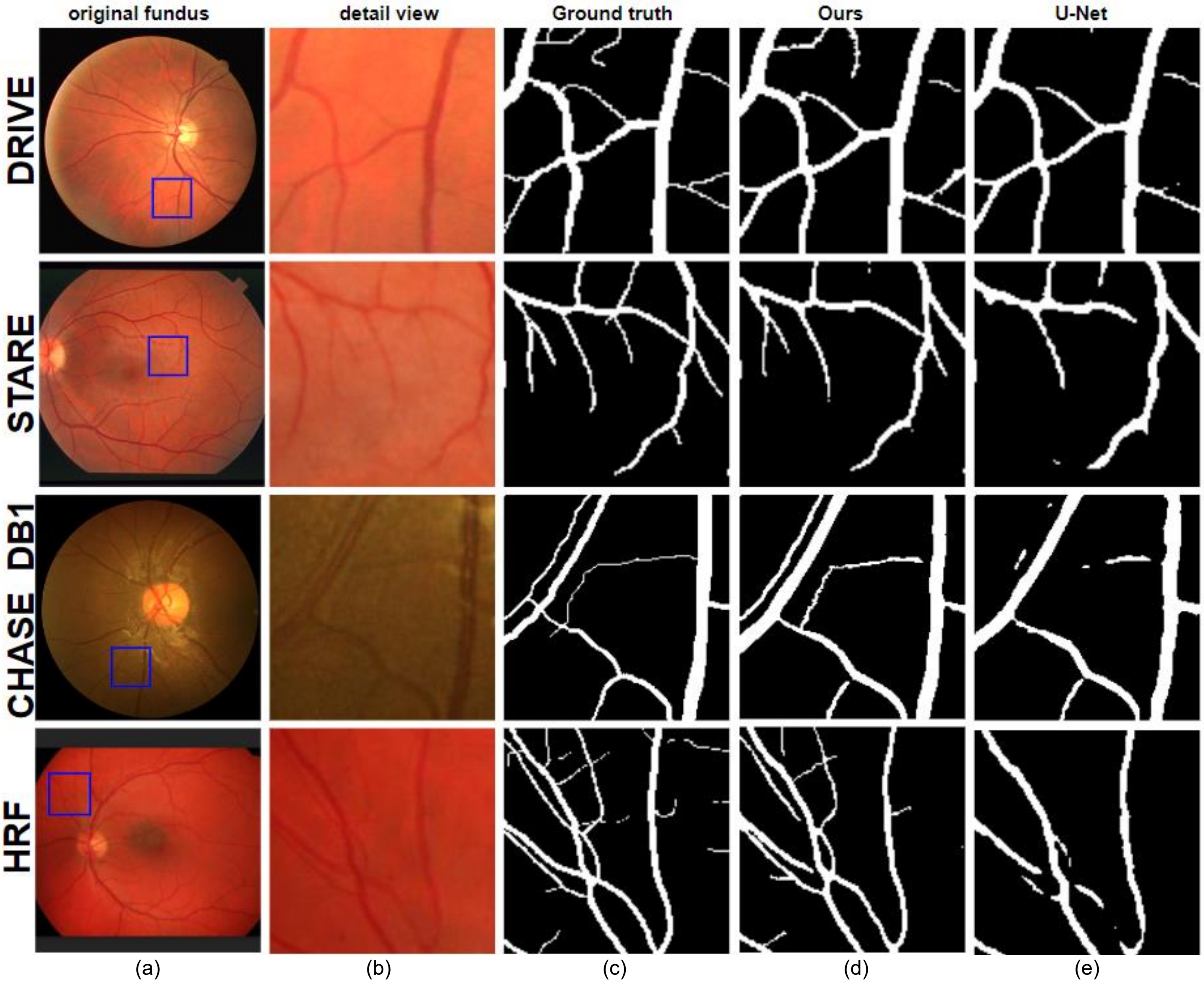


Figure 7. Visual comparison of retinal vessel segmentation results on four public datasets (DRIVE, STARE, CHASE_DB1, and HRF). From left to right: (a) original fundus image, (b) ground truth, (c) detail view, (d) our result, (e) U-Net Baseline

Robustness Evaluation under Different Data Splitting Strategies

To further examine the robustness and generalization capability of the proposed method, we conducted additional experiments using different data partition settings. For the DRIVE dataset, besides the standard official split, two supplementary settings were considered. In the first setting, the original training and test subsets were exchanged to form a reverse split, which provides a more challenging evaluation because the model is trained on the images originally used for testing. In the second setting, we constructed a random 20/20 split based on a fixed random seed. The 20 images were randomly selected for training and the remaining 20 images were used for testing. The results are presented in Table 9.

As shown in Table 9, both methods exhibit a certain degree of performance fluctuation when the data partition changes, which is expected given the small scale of the DRIVE dataset. Nevertheless, the proposed method consistently outperforms the baseline U-Net under both alternative settings. Our model preserves a clear advantage in sensitivity, indicating that the BAGF and SPA modules remain effective in recovering thin vessels even when the train–test composition changes. Although the reverse split leads to a slight decrease compared with the standard split, the overall performance remains stable, suggesting that the proposed network is not overly dependent on a single predefined partition.

Table 9. Segmentation results on the DRIVE dataset under different data splitting strategies. (all values in %). Best values are in bold

Method	Split	F1	ACC	SEN	SPE
U-Net(2015) [6]	Reverse split	78.41	95.92	77.28	98.17
	Random 20/20 split	79.02	96.08	78.11	98.22
Ours	Reverse split	**80.86**	**97.31**	**82.14**	**98.39**
	Random 20/20 split	**81.27**	**97.48**	**82.96**	**98.44**

For the CHASE_DB1 dataset, we further adopted the four-fold cross-validation protocol reported by Oliveira *et al.* [36]. The 28 images are divided into four subsets and each subset is used as the test set while the remaining three subsets are used for training. The fold results are summarized in Table 10. The average results across the four folds show that the proposed method consistently achieves higher F1-score and sensitivity than the baseline U-Net, while also maintaining high specificity. The relatively small variation across folds indicates that the proposed method has good robustness to data variation and maintains stable segmentation quality under different partitioning schemes.

Table 10. Experimental results on the CHASE_DB1 dataset under four-fold cross-validation. (all values in %). Best values are in bold.

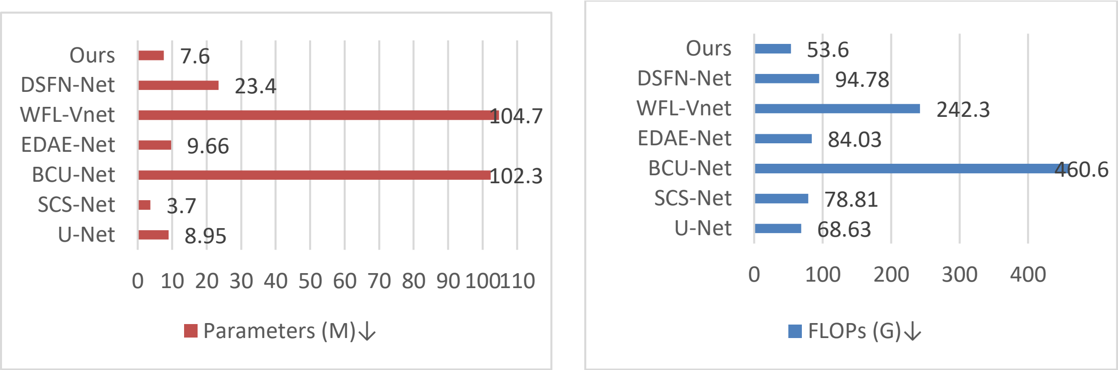
Method	Fold	F1	ACC	SEN	SPE
U-Net(2015) [6]	Fold 1	78.06	95.44	76.91	97.76
	Fold 2	78.58	95.71	77.42	97.88
	Fold 3	78.21	95.60	77.11	97.82
	Fold 4	78.74	95.83	77.86	97.93
	Average	78.40 ± 0.31	95.65 ± 0.16	77.33 ± 0.43	97.85 ± 0.07
Ours	Fold 1	81.19	96.98	84.66	98.91
	Fold 2	81.64	97.16	85.08	99.02
	Fold 3	81.37	97.08	84.79	98.97
	Fold 4	81.88	97.24	85.41	99.05
	Average	81.52 ± 0.29	97.12 ± 0.11	84.99 ± 0.33	98.99 ± 0.06

Computional Complexity

In addition to segmentation accuracy, model complexity is an important factor in evaluating the practicality of retinal vessel segmentation methods, as it directly affects computational efficiency. The additional computational cost introduced by the BAGF and SPA modules is relatively small. BAGF mainly consists of lightweight convolution and guided filtering operations, while SPA employs a simple spatial pooling mechanism with negligible parameter increase. Table 11 compares our method with several representative retinal vessel segmentation models in terms of sensitivity (Sen), model parameters (M), and FLOPs (G), including U-Net, SCS-Net, BCU-Net, EDAE-Net, WFL-Vnet, and DSFN-Net. Figure 8(a) and Figure 8(b) further illustrate the comparisons of model parameters and FLOPs, respectively. As shown in Table 11 our method achieves the highest sensitivity (83.49%) while maintaining a moderate model complexity, with 7.6M parameters and 53.6G FLOPs. Although our model has more parameters than SCS-Net, it requires fewer parameters than U-Net, BCU-Net, EDAE-Net, WFL-Vnet, and DSFN-Net. In terms of FLOPs, our method is lower than U-Net, SCS-Net, BCU-Net, EDAE-Net, WFL-Vnet, and DSFN-Net. These results indicate that the proposed method achieves a favorable balance between segmentation sensitivity and computational efficiency, making it suitable for practical retinal vessel segmentation applications.

Table 11. Comparison of Model Complexity on DRIVE Dataset. Sen indicates sensitivity, Parameters (M) represents the number of model parameters in millions, and FLOPs (G) denotes floating-point operations in gigaflops

Method	**Sen ↑**	**Parameters (M)↓**	**FLOPs (G)↓**
U-Net(2015) [6]	76.86	8.95	68.63
SCS-Net(2021) [4]	82.89	**3.7**	78.81
BCU-Net(2023) [31]	82.38	102.3	460.6
EDAE-Net(2024) [19]	82.28	9.66	84.03
WFL-Vnet(2025) [33]	82.53	104.7	242.3
DSFN-Net(2026) [37]	82.21	23.40	94.78
Ours	**83.49**	7.6	**53.6**



(a) FLOPs comparison (b)Parameters comparison

Figure 8. Visualization of Model Complexity

Discussion

Although existing methods are capable of segmenting retinal vessels to some extent, the segmentation task remains challenging due to factors such as low image contrast, variations in vessel thickness, complex vascular topology, and the presence of various lesions. In contrast, we propose a Boundary-Preserving Network with Attention-Guided Filtering for retinal vessel segmentation. Our method significantly addresses challenges such as small vessel detection, vessel boundary preservation, and robustness across images of varying quality. One of the main advantages of our approach is that, through the proposed Boundary Aware Guided Filtering (BAGF) module, we effectively enhance vessel boundary detection and restore spatial details lost during down-sampling. This module leverages structural priors and focuses on preserving boundaries, ensuring that even small vessels remain well-preserved during the down-sampling process.

we introduce a Spatial Pooling Attention (SPA) mechanism that adaptively emphasizes the importance of vessel regions and suppresses background noise, thus enhancing the vessel representation. This mechanism enables the network to effectively focus on the vessel regions in areas with low contrast or heavy noise, avoiding interference from the background. We evaluate our approach on several benchmark datasets, including DRIVE, STARE, CHASE_DB1, and HRF. The experimental results demonstrate that our method achieves competitive performance compared with existing approaches in both quantitative performance and visual quality. Extensive experiments show that the BAGF module significantly improves vessel boundary preservation and restores spatial details, while the SPA mechanism ensures that the network accurately focuses on the most informative vessel regions. Ablation studies further confirm the effectiveness of the BAGF and SPA modules in boundary preservation and small-vessel segmentation. These results indicate that by focusing on boundary preservation and adaptively emphasizing vessel regions, our method provides a comprehensive understanding of retinal vessel structures, particularly in challenging conditions such as low contrast or low-resolution images, showcasing its stronger robustness.

From Tables 5 to 8, it can be seen that all methods exhibit relatively low Sen (Sensitivity) values, which can be attributed to several reasons: on one hand, the complex morphological structure of retinal vessels makes accurate segmentation difficult; on the other hand, balancing Sen and Spe (Specificity) during supervised training is challenging. A higher Spe value means that the model is better at accurately detecting non-vessel pixels in the background, reducing the number of false positives (FP). A higher Sen value indicates a stronger ability of the model to correctly identify vessel regions, thus reducing false negatives. However, this also increases the probability of vessel pixels being misidentified as non-vessel pixels. Therefore, in this study, we chose to sacrifice a small portion of Spe in order to improve Sen. As can be seen from these tables, while some methods achieve higher results than ours on specific datasets, they do not outperform our approach across all datasets. In contrast, our proposed method consistently demonstrates higher sensitivity and relatively stable performance across all datasets compared to the state-of-the-art methods.

However, several competing methods achieve slightly higher performance on certain metrics in specific datasets. For example, WFL-VNet reports a higher F1 score on the DRIVE and CHASE_DB1 dataset, which attributed to its deeper architecture and extensive feature localization strategy throughout the network. Nevertheless, this approach typically introduce higher computational complexity and rely on more complicated network structures. In contrast, the proposed method aims to improve boundary preservation and thin-vessel detection through the combination of Boundary Aware Guided Filtering (BAGF) and Spatial Pooling Attention (SPA). This design enables the model to achieve higher sensitivity while maintaining moderate computational complexity, resulting in a balanced performance across different datasets.

Although our vessel segmentation model performs excellently, there are still some limitations, and there is considerable room for improvement. In extreme scenarios, such as extremely low contrast or when vessels are completely occluded by lesions, the model fails to accurately identify the entire vessel structure. The model's sensitivity to small vessels in low-contrast regions is reduced, which may lead to both false positives and false negatives. When vessel pixels are completely obscured by large lesions, accurate vessel identification becomes extremely difficult.

Conclusions

In this work, we introduced a retinal vessel segmentation framework that extends U-Net with a multi-scale attention guided filtering module. The proposed design effectively preserves edge information and enhances the segmentation of thin vessels. Experimental results on DRIVE, STARE, CHASE_DB1, and HRF show that the method achieves competitive performance compared with previous approaches, delivering the best results on every dataset. These findings confirm the robustness and practical value of the framework for accurate retinal image analysis.

Future work will focus on enhancing the model's sensitivity to fine structures in low-contrast scenarios. Although this paper focuses on retinal vessel segmentation, it is necessary to apply the proposed model to other domains with large databases to validate its variability and scalability. In future work, we may consider expanding the application domains, including crack detection, facial wrinkle detection, and general medical image processing.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Statement

This study utilizes publicly available retinal fundus image datasets for retinal vessel segmentation. As the data used in this research are publicly accessible and de-identified, individual informed consent was not required. The study was conducted in accordance with the principles outlined in the Declaration of Helsinki (2013), and no direct involvement with human participants occurred. All data handling and processing were carried out in compliance with privacy and ethical standards to ensure the protection of individuals' privacy.

Acknowledgment

The authors would like to thank the Faculty of Computing at Universiti Teknologi Malaysia for providing them with access to the research resources that were essential to carrying out this work. This acknowledgement signifies their gratitude and recognition of the support received from this institution during the course of their research.

References

[1] Liu, Z., Sunar, M. S., Tan, T. S., & Hitam, W. H. W. (2025). Deep learning for retinal vessel segmentation: A systematic review of techniques and applications. *Medical & Biological Engineering & Computing, 63*(8), 2191–2208.

[2] Verma, P. K., & Kaur, J. (2024). Systematic review of retinal blood vessels segmentation based on AI-driven technique. *Journal of Imaging Informatics in Medicine, 37*(4), 1783–1799.

[3] Sule, O. O. (2022). A survey of deep learning for retinal blood vessel segmentation methods: Taxonomy, trends, challenges and future directions. *IEEE Access, 10*, 38202–38236.

[4] Wu, H., Wang, W., Zhong, J., Lei, B., Wen, Z., & Qin, J. (2021). SCS-Net: A scale and context sensitive network for retinal vessel segmentation. *Medical Image Analysis, 70*, 102025.

[5] Li, T., Bo, W., Hu, C., Kang, H., Liu, H., Wang, K., & Fu, H. (2021). Applications of deep learning in fundus images: A review. *Medical Image Analysis, 69*, 101971.

[6] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention (MICCAI)* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28.

[7] He, Y., Yi, Y., Zheng, C., & Kong, J. (2024). BGF-Net: Boundary guided filter network for medical image segmentation. *Computers in Biology and Medicine, 171*, 108184.

[8] Tan, Y., Yang, K. F., Zhao, S. X., Wang, J., Liu, L., & Li, Y. J. (2024). Deep matched filtering for retinal vessel segmentation. *Knowledge-Based Systems, 283*, 111185.

[9] Zhang, S., Fu, H., Yan, Y., Zhang, Y., Wu, Q., Yang, M., & Xu, Y. (2019). Attention guided network for retinal image segmentation. In *Medical image computing and computer-assisted intervention (MICCAI)* (pp. 797–805). Springer. https://doi.org/10.1007/978-3-030-32239-7_88.

[10] Zhang, X., Zhao, W., Zhang, W., Peng, J., & Fan, J. (2022). Guided filter network for semantic image segmentation. *IEEE Transactions on Image Processing, 31*, 2695–2709.

[11] Yin, P., Yuan, R., Cheng, Y., & Wu, Q. (2020). Deep guidance network for biomedical image segmentation. *IEEE Access, 8*, 116106–116116.

[12] Jin, Q., Meng, Z., Pham, T. D., Chen, Q., Wei, L., & Su, R. (2019). DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems, 178*, 149–162.

[13] Wu, Y., Xia, Y., Song, Y., Zhang, Y., & Cai, W. (2020). NFN+: A novel network followed network for retinal vessel segmentation. *Neural Networks, 126*, 153–162.

[14] Wei, J., Zhu, G., Fan, Z., Liu, J., Rong, Y., Mo, J., & Chen, X. (2021). Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm. *IEEE Transactions on Medical Imaging, 41*(2), 292–307.

[15] Zhang, Y., He, M., Chen, Z., Hu, K., Li, X., & Gao, X. (2022). Bridge-Net: Context-involved U-Net with patch-based loss weight mapping for retinal blood vessel segmentation. *Expert Systems with Applications, 195*, 116526.

[16] Jiang, M., Zhu, Y., & Zhang, X. (2024). CoVi-Net: A hybrid convolutional and vision transformer neural network for retinal vessel segmentation. *Computers in Biology and Medicine, 170*, 108047.

[17] He, K., Sun, J., & Tang, X. (2012). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(6), 1397–1409.

[18] Seo, S., Yoo, S., & Yoon, H. (2025). Full-scale representation guided network for retinal vessel segmentation. *BMC Medical Imaging, 25*(1), 484.

[19] Li, J., Gao, G., Yang, L., & Liu, Y. (2024). A retinal vessel segmentation network with multiple-dimension attention and adaptive feature fusion. *Computers in Biology and Medicine, 172*, 108315.

[20] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7132–7141). https://doi.org/10.1109/CVPR.2018.00745

[21] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *European Conference on Computer Vision (ECCV)* (pp. 3–19). https://doi.org/10.1007/978-3-030-01234-2_1.

[22] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11534–11542).

[23] Yu, M., Han, M., Li, X., Wei, X., Jiang, H., Chen, H., & Yu, R. (2022). Adaptive soft erasure with edge self-attention for weakly supervised semantic segmentation: Thyroid ultrasound image case study. *Computers in Biology and Medicine, 144*, 105347.

[24] Fu, Y., Liu, J., & Shi, J. (2024). TSCA-Net: Transformer-based spatial-channel attention segmentation network for medical images. *Computers in Biology and Medicine, 170*, 107938.

[25] Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging, 23*(4), 501–509.

[26] Hoover, A., Kouznetsova, V., & Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging, 19*(3), 203–210.

[27] Owen, C. G., Rudnicka, A. R., Mullen, R., Barman, S. A., Monekosso, D., Whincup, P. H., & Paterson, C. (2009). Measuring retinal vessel tortuosity in 10-year-old children: Validation of the CAIAR program. *Investigative Ophthalmology & Visual Science, 50*(5), 2004–2010.

[28] Budai, A., Bock, R., Maier, A., Hornegger, J., & Michelson, G. (2013). Robust vessel segmentation in fundus images. *International Journal of Biomedical Imaging, 2013*, 154860.

[29] Yin, P., Cai, H., & Wu, Q. (2022). DF-Net: Deep fusion network for multi-source vessel segmentation. *Information Fusion, 78*, 199–208.

[30] Li, D., Peng, L., Peng, S., Xiao, H., & Zhang, Y. (2023). Retinal vessel segmentation by using AFNet. *The Visual Computer, 39*(5), 1929–1941.

[31] Zhang, H., Zhong, X., Li, G., Liu, W., Liu, J., Ji, D., & Wu, J. (2023). BCU-Net: Bridging ConvNeXt and U-Net for medical image segmentation. *Computers in Biology and Medicine, 159*, 106960.

[32] Panchal, S., & Kokare, M. (2024). ResMU-Net: Residual multi-kernel U-Net for blood vessel segmentation in retinal fundus images. *Biomedical Signal Processing and Control, 90*, 105859.

[33] Song, P., & Wu, Y. (2025). WFL-VNet: Retinal vessel segmentation method using whole-process feature localization. *Signal, Image and Video Processing, 19*(5), 356.

[34] Wang, J., Li, X., & Ma, Z. (2025). Multi-scale three-path network (MSTP-Net): A new architecture for retinal vessel segmentation. *Measurement, 250*, 117100.

[35] Cao, J., Chen, J., Gu, Y., & Liu, J. (2023). MFA-UNet: A vessel segmentation method based on multi-scale feature fusion and attention module. *Frontiers in Neuroscience, 17*, 1249331.

[36] Oliveira, A., Pereira, S., & Silva, C. A. (2018). Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications, 112*, 229–242.

[37] Li, Z., Wang, H., Tang, N., Lan, R., Li, B., & Luo, X. (2026). A dual-layer semantic fusion network for retinal vessel segmentation. *Biomedical Signal Processing and Control, 112*, 108463.