An International Journal of Optimization and Control: Theories & Applications
ISSN: 2146-0957 eISSN: 2146-5703
Vol.16, No.1, pp.305-320 (2026)
https://doi.org/10.36922/IJOCTA025420177

RESEARCH ARTICLE

FracSegNet: A deep learning image segmentation model for medical images of eye diseases

Shoutong Huang*, Yu Ma*, Huitan Chang, and Bowen Xiao

Department of Electronic Science and Technology, School of Electronic and Electrical Engineering, Ningxia University, Yinchuan, Ningxia, China
12024130722@stu.nxu.edu.cnmay@nxu.edu.com12024130818@stu.nxu.edu.cn12022130647@stu.nxu.edu.cn

ARTICLE INFO

Article History:
Received: October 13, 2025
Revised: November 16, 2025
Accepted: November 21, 2025
Published online: January 13, 2026

Keywords:
Deep learning
Digital retinal images for vessel extraction
Fractional differential operator
Image segmentation
Medical image processing

ABSTRACT

Retinal vessel segmentation is essential for the diagnosis and treatment planning of retinal diseases, yet remains challenging due to weak edges, tiny branches, and complex background textures. In this paper, we propose FracSegNet, a deep segmentation network that integrates fractional-order modeling at the preprocessing, feature extraction, and loss levels to improve the continuity and robustness of retinal vessel extraction. First, a Grünwald–Letnikov fractional differential operator is used to generate multi-directional edge responses, which are concatenated with the original fundus image to form an augmented multi-channel input. Second, adaptive fractional-order convolution blocks are embedded into a U-Net–like encoder–decoder architecture, where learnable order weights dynamically fuse integer-order and fractional-order responses, enabling simultaneous modeling of local details and long-range dependencies. Third, a composite loss is designed by combining Dice loss, total variation (TV) regularization, and a fractional gradient constraint that enforces consistency between the fractional-order gradients of the prediction and the ground truth. Experiments on the DRIVE, STARE, and CHASE_DB1 datasets demonstrated that FracSegNet achieved competitive or superior performance compared with state-of-the-art methods, with F1 scores above 0.83 and clear improvements in edge continuity and fine-branch preservation. These results indicate that fractional-order modeling provides an effective and generalizable paradigm for segmenting weak edges and delicate vascular structures in medical images.

1. Introduction

Retinal vessel segmentation is fundamental for early diagnosis, treatment planning, and prognosis assessment of various fundus diseases, including diabetic retinopathy and glaucoma. Public datasets, such as the Digital Retinal Images for Vessel Extraction (DRIVE), the Child Heart and Health Study in England (CHASE_DB1), and the Structured Analysis of the Retina (STARE), provide high-resolution retinal images with pixel-wise annotations, which have greatly facilitated the development and benchmarking of vessel segmentation algorithms. However, tiny vessel branches, weak edges, and complex background textures still pose significant challenges, often leading to fragmented, discontinuous, or inaccurate segmentation results.[1]

Traditional methods based on thresholding, region growing, and level sets struggle to handle low contrast and imaging noise in a fully automatic and robust manner. With the rapid development of deep learning, convolutional neural networks (CNNs),[2] especially U-Net[3] and its

*Corresponding Author

variants, have become the de facto standard for medical image segmentation. U-Net leverages an encoder–decoder architecture with skip connections to fuse multi-scale features and has achieved remarkable performance across many tasks. Nevertheless, standard convolutions are inherently local, and their limited receptive fields restrict the ability to model long-range dependencies, which is critical for preserving the continuity of extremely thin vessels and weak boundary regions.[4,5]

In recent years, with the rapid development of deep learning, CNNs have become the mainstream method for medical image segmentation,[4] with U-Net and its variants being the most representative.[5] U-Net leverages an encoder–decoder architecture with skip connections to perform end-to-end predictions from pixel-level to semantic-level representations, enabling multi-resolution feature fusion and achieving remarkable performance across various medical segmentation tasks. Nevertheless, the limited receptive field of standard convolutions restricts their ability to model long-range dependencies, making it difficult for U-Net and its derivatives to maintain structural continuity and completeness when handling extremely thin vessels and weak boundary regions.

To overcome the limitation of local receptive fields, researchers have recently introduced the global self-attention mechanism from transformers into visual segmentation tasks.[6] Self-attention allows for modeling long-range dependencies between any pair of pixels in an image, addressing the lack of global context in purely convolutional architectures. Vision Transformer (ViT), a pioneering model that applies a transformer to visual recognition,[7] has shown superior performance across various classification and segmentation benchmarks. Subsequently, hybrid CNN–transformer architectures, such as TransUNet[8] and TransNorm,[9] have been proposed, demonstrating improved results in medical segmentation tasks, including chest computed tomography and brain magnetic resonance imaging.[10]

However, pure transformer modules often suffer from high computational cost and limited sensitivity to local details, leading to a trade-off between accuracy and efficiency when dealing with delicate structures. On the other hand, fractional calculus has demonstrated unique advantages in image enhancement, texture analysis, and edge detection due to its inherent non-locality and memory characterization. For example, Zhang et al.[11] employed fractional differential filters for edge enhancement, while Liu et al.[12] proposed fractional gradient regularization to effectively suppress noise. Nevertheless, deep integration of fractional operators with deep neural networks remains scarce and lacks systematic frameworks.

Recently, fractional calculus has attracted increasing attention in combination with deep learning and computer vision, due to its inherent non-locality and memory properties. Several studies have explored the use of fractional-order operators for image enhancement, denoising, and segmentation within neural network frameworks, showing that fractional derivatives can complement standard convolutions by emphasizing long-range dependencies and high-frequency structures. These works provide an important theoretical foundation for our design of fractional differential preprocessing, adaptive fractional convolution, and fractional gradient-based regularization in FracSegNet.[13–15]

Meanwhile, a variety of CNN-based and hybrid CNN–transformer architectures have been proposed for retinal vessel segmentation and general medical image segmentation. Transformer-enhanced U-shaped models, such as Swin-UNet, TransAttUnet, and other ViT variants, have demonstrated strong capabilities in modeling global context. At the same time, recent retinal vessel networks combine multi-scale CNNs with attention mechanisms and context-aware modules to better capture thin vessels and complex backgrounds.[16–19] Our work complements these approaches: instead of focusing solely on self-attention, we explicitly introduced fractional-order modeling at multiple levels of the network to improve weak-edge continuity and fine-branch preservation.

To address these issues, we propose FracSegNet, a retinal vessel segmentation network that integrates fractional-order modeling at three levels: (i) a fractional differential preprocessing step that enhances weak edges by concatenating multi-directional fractional gradient responses with the original fundus image; (ii) adaptive fractional-order convolution blocks embedded in a U-Net–like encoder–decoder to dynamically fuse integer-order and fractional-order features via learnable order weights; and (iii) a composite loss that incorporates a fractional gradient constraint in addition to Dice loss and TV regularization, enforcing structural consistency in the fractional-order domain. This unified framework aims to simultaneously improve edge continuity, fine-branch preservation, and robustness to complex textures.

The remainder of the paper is organized as follows: Section 2 introduces the architecture of

FracSegNet, its core modules, and the design of the loss function. Section 3 presents the experimental setup, quantitative comparisons, ablation studies, and visualization results. Section 4 discusses the results and future research directions. Section 5 concludes the paper.

2. Proposed method

It is important to clarify that the novelty of this work lies not in deriving a new fractional-derivative formula, but rather in integrating the classical Grünwald–Letnikov operator into a modern deep segmentation framework at multiple levels. Specifically, we (i) used multi-directional fractional differential filtering to generate structure-enhanced edge maps that are concatenated with the Red–Green–Blue channels, a three-channel color model widely used to represent color images, (ii) embedded adaptive fractional convolution blocks into each encoder–decoder stage so that the network can learn to balance integer-order and fractional-order responses through trainable fusion coefficients, and (iii) incorporated a fractional gradient-consistent loss term that enforces structural agreement between the prediction and the ground truth in the fractional-order domain. To the best of our knowledge, such a unified and deeply integrated use of fractional calculus for retinal vessel segmentation has not been reported previously. The overall architecture is illustrated in **Figure 1**.

As shown in **Figure 1**, FracSegNet comprises three main stages:
(i) Fractional-order preprocessing layer: Generates multi-channel inputs enhanced with edge information.
(ii) Encoder–decoder network: Embeds adaptive fractional-order convolution modules.
(iii) Fractional-order constrained loss: Jointly optimizes segmentation accuracy and boundary smoothness.

2.1. Fractional differential enhancement

To strengthen weak vessel edges and global texture features, the original retinal image $I \in \mathbb{R}^{H\times W\times 3}$ is first preprocessed using fractional-order differentiation. The Grünwald–Letnikov fractional derivative discrete approximation is utilized, and its horizontal direction operation on the pixel (x, y) is defined as:

$$ {}^{GL}D_s^{\alpha} I(x,y) \approx \sum_{k=0}^{N} (-1)^k \binom{\alpha}{k} I(x-k,y) \quad (1)$$

where $\alpha \in (0,2)$, $\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}$, and typically $N = 10$. Similarly, calculations are performed in the vertical direction. Based on the above discrete formula, a 3×3 fractional-order differentiation kernel can be constructed.

$$\begin{aligned} G_\alpha &= \begin{bmatrix} -w_2 & -w_1 & -w_2 \\ -w_1 & w_0 & -w_1 \\ -w_2 & -w_1 & -w_2 \end{bmatrix}, w_0 = 1 + \frac{\alpha}{2}, w_1 \\ &= \frac{\alpha}{4}, w_2 = \frac{\alpha(\alpha-1)}{48} \end{aligned} \quad (2)$$

The fractional-order kernel G_α is convolved channel-wise with the color image I to produce a single-channel fractional-order edge response map $E \in \mathbb{R}^{H\times W}$. Subsequently, E is concatenated with the original image I along the channel dimension to form a multi-channel input.

$$I_{\text{in}} = \text{concat}\,[I, E] \in \mathbb{R}^{H\times W\times 4} \quad (3)$$

This module captures both local abrupt changes and global texture correlations, providing richer edge information for subsequent network processing.

2.2. Overall network architecture

FracSegNet inherits the symmetrical encoder–decoder structure from U-Net but replaces each convolutional operation with an adaptive fractional convolution block. Additionally, after skip connections are concatenated with decoder features, fractional-order operators are used again to constrain boundaries, fully exploiting the non-local properties of fractional calculus. The network comprises the four main components:

(1) Encoder: Four-stage downsampling, each containing two adaptive fractional convolution blocks, followed by a 2×2 MaxPool operation.
(2) Bridge layer: Two adaptive fractional convolution blocks for deep feature extraction.
(3) Decoder: Four-stage upsampling, each stage consisting of bilinear interpolation ($2\times$ upsampling), concatenation with corresponding skip-connection features, followed by two adaptive fractional convolution blocks.
(4) Output layer: A 1×1 standard convolution and sigmoid activation to produce pixel-wise segmentation probability maps.

The subsequent sections will elaborate on the detailed implementation of the adaptive fractional convolution block, encoder, and decoder.

2.2.1. Adaptive fractional-order convolution block

The core idea of the Adaptive Fractional Convolution Block lies in performing both integer-order and fractional-order convolutions simultaneously

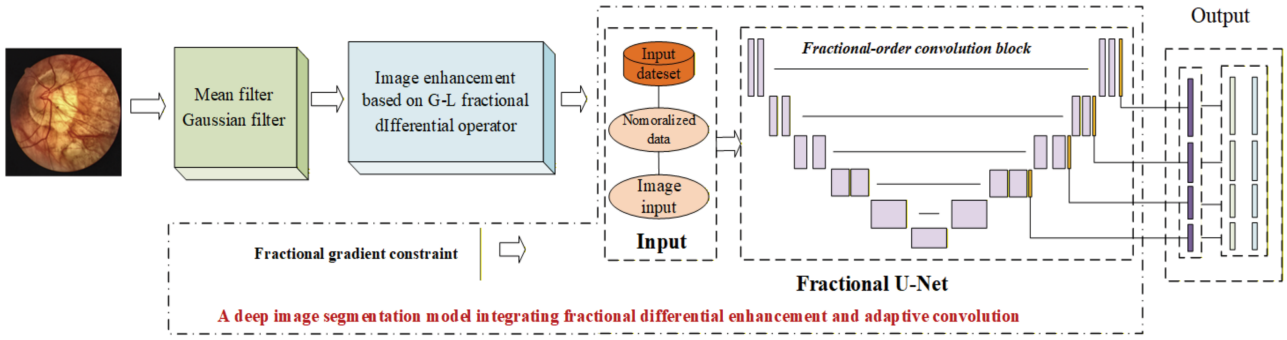

Figure 1. FracSegNet model architecture
Abbreviation: G-L: Grünwald–Letnikov.

and dynamically fusing their responses through a learnable weighting coefficient α_t. Specifically, given an input feature map F, the output of this convolution block is expressed as follows:

$$F_{\text{out}} = \text{ReLU}\left(\underbrace{\text{Conv}_{\text{int}}(F;\theta_{\text{int}})}_{\text{Integer-Order Convolution}} + \alpha_t \underbrace{\text{Conv}_{\text{frac}}(F;\theta_{\text{frac}})}_{\text{Fractional-Order Convolution}}\right) \tag{4}$$

where Conv_{int} denotes a standard 2D convolution with a kernel size of 3×3, parameterized by θ_{int}; $\text{Conv}_{\text{frac}}$ also uses a 3×3 kernel but applies Grünwald–Letnikov (G–L) fractional weighting to the input features within the convolution window; $\alpha_t \in [0, 1]$ is a learnable scalar that is updated linearly or nonlinearly during training, controlling the contribution of the fractional-order response; ReLU refers to the standard rectified linear unit activation function.

During backpropagation, the gradient of α_t is directly propagated from the fractional branch.

$$\frac{\partial \mathcal{L}}{\partial \alpha_t} = \sum_{i,j} \delta_{i,j} [\text{Conv}_{\text{frac}}(F;\ \theta_{\text{frac}})]_{i,j} \tag{5}$$

where $\delta_{i,j}$ is the error gradient backpropagated from the upstream network (through ReLU and subsequent layers), measuring the sensitivity of the loss to the output at spatial location (i, j).

$\text{Conv}_{\text{frac}}(F;\theta_{\text{frac}})$ denotes the response value of the fractional-order convolution branch at location (i, j), representing the contribution of that position's fractional feature to the final output.

Regarding adaptive adjustment, by relying directly on the contributions of fractional features, the network can automatically learn the optimal α_t for each sample and layer, dynamically balancing "memory effect" and local-detail preservation. For a simplified update, the update for α_t requires no additional complex gradient paths—it receives its signal directly from the fractional branch, ensuring stable, and efficient learning. In terms of interpretability, after training, examining the distribution of α_t provides clear insight into the model's reliance on fractional features under different conditions, guiding further architectural refinement. All these enable the network to adaptively adjust the importance of fractional-order features.

2.2.2. Encoder design

In the encoder of FracSegNet, the input feature maps are processed through a multi-scale hierarchical pipeline that combines progressive feature extraction with downsampling, enabling the network to simultaneously capture fine-grained details and global semantics.

Let the input feature map at the i-th level $(i = 1, \ldots, 4)$ be denoted as $F_{i-1} \in \mathbb{R}^{H_{i-1} \times W_{i-1} \times C_{i-1}}$ The feature is first passed through an Adaptive Fractional Convolution Block (AFCB) to obtain an intermediate representation:

$$F_i^{(1)} = \text{ReLU}\left(\text{Conv}_{\text{int}}\left(F_{i-1};\ \theta_{\text{int}}^i\right) + \alpha_t^i \text{Conv}_{\text{frac}}\left(F_{i-1};\ \theta_{\text{frac}}^i\right)\right) \tag{6}$$

followed by a second convolution of the same form:

$$F_i^{(2)} = \text{ReLU}\left(\text{Conv}_{\text{int}}\left(F_i^{(1)};\ \theta_{\text{int}}^i\right) + \alpha_t^i \text{Conv}_{\text{frac}}\left(F_i^{(1)};\ \theta_{\text{frac}}^i\right)\right) \tag{7}$$

Here, Conv_{int} and $\text{Conv}_{\text{frac}}$ denote the integer-order and fractional-order convolution operators, respectively. The coefficient $\alpha_t^i \in [0, 1]$ is a learnable fusion scalar that controls the balance between the two branches, while θ_{frac}^i and θ_{int}^i are the corresponding kernel parameters.

The resulting feature $F_i^{(2)}$ is preserved for skip connection to the decoder. To downsample the spatial resolution, a 2×2 max pooling operation is applied:

$$F_i = \text{MaxPool}\left(F_i^{(2)}\right) \in \mathbb{R}^{H_i \times W_i \times C_i}, (H_i, W_i) = \left(\frac{H_{i-1}}{2}, \frac{W_{i-1}}{2}\right) \tag{8}$$

with the number of channels typically doubled, i.e., $C_i = 2C_{i-1}$.

Constructed in this manner, the four-stage encoder introduces long-range dependencies and non-local features at each scale through adaptive fractional convolution, while preserving high-resolution texture information via skip connections. This design enables the decoder to effectively fuse multi-scale features during spatial resolution recovery, thereby significantly improving segmentation performance for weak edges and thin vessel branches.

2.2.3. Bridge module design

The bridge layer in FracSegNet serves as a semantic linkage between the encoder and decoder. It is responsible for further abstracting and reorganizing deep features, and its output directly influences the decoder's ability to capture global contextual information.

FracSegNet employs two consecutive adaptive fractional convolution blocks in the bridge layer, similar to those used in the encoder. However, the number of channels and the receptive field are deliberately increased at this stage to fully integrate the multi-scale features aggregated from all preceding encoder levels.

Let the input to the bridge layer be the feature map output from the fourth downsampling stage of the encoder, denoted as:

$$F_4 \in \mathbb{R}^{H_4 \times W_4 \times C_4} \tag{9}$$

Here, the dimensions are defined as $(H_4, W_4) = (32, 32), C_4 = 512$. The bridge layer first applies an adaptive fractional convolution to F_4, computed as:

$$B^{(1)} = \text{ReLU}\left(\text{Conv}_{\text{int}}\left(F_4; \theta_{\text{int}}^B\right) + \alpha_B \text{Conv}_{\text{frac}}\left(F_4; \theta_{\text{frac}}^B\right)\right) \tag{10}$$

On this basis, a second convolution operation of the same structure is performed:

$$B^{(2)} = \text{ReLU}\left(\text{Conv}_{\text{int}}\left(B^{(1)}; \theta_{\text{int}}^B\right) + \alpha_B \text{Conv}_{\text{frac}}\left(B^{(1)}; \theta_{\text{frac}}^B\right)\right) \tag{11}$$

Here, θ_{int}^B and θ_{frac}^B represent the integer-order and fractional-order convolution kernel parameters in the bridge layer, respectively, while $\alpha_B \in [0, 1]$ is a globally learnable fusion weight that dynamically adjusts the balance between long-range dependencies and local details during deep semantic refinement.

To further enlarge the receptive field, both convolutions in the bridge layer can adopt a dilated convolution version with a dilation rate $d > 1$,

$$\text{Conv}_{\text{frac}}^{(d)}(X)_{i,j} = \sum_{p,q} G_\alpha(p, q) X_{i-dp, j-dq} \tag{12}$$

to better capture non-local information over a larger spatial extent. The final output of the bridge layer is:

$$B = B^{(2)} \in \mathbb{R}^{32 \times 32 \times C_B} \tag{13}$$

Here, C_B can be set equal to C_4 or further doubled to enhance feature representation capacity in the subsequent decoding stage. This design enables FracSegNet to effectively fuse global long-range dependencies and local texture details at the deepest level of the network, providing an information-rich starting point for the decoder to recover high-precision segmentation.

2.2.4. Decoder design

The decoder is symmetrical to the encoder and also consists of four hierarchical stages. At each stage j ($j = 1, \ldots, 4$), spatial resolution is recovered through the following process:

(1) Upsampling: The output D_{j-1} from the previous layer (with D_0 being the bridge feature B) is upsampled by a factor of 2 via bilinear interpolation to obtain $\tilde{D}_j$

(2) Skip connection and concatenation: $\hat{D}_j$ is concatenated along the channel dimension with the stored feature map S_{5-j} from the corresponding encoder stage, producing $\hat{D}_j$;

(3) Adaptive fractional convolution: Two consecutive adaptive fractional convolution blocks are applied to $\hat{D}_j$, yielding D_j.

Finally, the output $D_4 \in \mathbb{R}^{256 \times 256 \times C}$ is passed through a 1×1 convolution to reduce it to a single channel, followed by Sigmoid activation, resulting in the pixel-wise segmentation probability map $P \in [0, 1]^{256 \times 256}$.

2.3. Fractional gradient constraint

To improve segmentation accuracy while ensuring geometric continuity of boundaries and suppressing background noise, FracSegNet incorporates a composite loss function that combines conventional region-based similarity (Dice Loss) and

TV regularization with an additional fractional-order gradient constraint term. The overall loss function is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Dice}} + \lambda_1 \mathcal{L}_{\text{TV}} + \lambda_2 \mathcal{L}_{\text{grad}} \tag{14}$$

where $\mathcal{L}_{\text{total}}$ denotes the overall loss, which jointly optimizes region overlap, boundary smoothness, and structural fidelity. $\mathcal{L}_{\text{Dice}}$ is the Dice loss, measuring the overlap between the predicted mask P and the ground-truth mask G. λ_1 is the total variation (TV) regularization weight—a nonnegative hyperparameter that balances the influence of the TV loss against the Dice loss. $\mathcal{L}_{\text{TV}}$ is the TV loss, which penalizes abrupt changes in the prediction map to promote smooth boundaries and reduce noise. λ_2 is the fractional-order gradient constraint weight—a nonnegative hyperparameter that balances the fractional gradient loss against the other terms. $\mathcal{L}_{\text{grad}}$ is the fractional-order gradient loss, which enforces consistency between the fractional-order gradients of the prediction and the ground truth, thereby enhancing the preservation of fine structural details. Some terms are described as follows:

(i) Dice loss:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i,j}\ p_{i,j} g_{i,j}}{\sum_{i,j}\ p_{i,j} + \sum_{i,j}\ g_{i,j}} \tag{15}$$

This term measures the pixel-wise overlap between the predicted probability map (p) and the ground truth (g), effectively addressing foreground–background class imbalance.

(ii) Total variation regularization:

$$\mathcal{L}_{\text{TV}} = \sum_{ij}\ \sqrt{(p_{i+1,j} - p_{i,j})^2 + (p_{i,j+1} - p_{i,j})^2} \tag{16}$$

This term penalizes local gradient magnitudes in the predicted map to suppress noise and irregular artifacts in background regions, promoting overall edge smoothness without excessively blurring meaningful structures.

(iii) Fractional-order gradient constraint:

$$\mathcal{L}_{\text{grad}} = \frac{1}{N} \sum_{i,j}\ |{}^{\text{GL}}D^{\beta}(p-g)_{i,j}| \tag{17}$$

Here, ${}^{\text{GL}}D^{\beta}$ denotes the discrete fractional-order gradient operator based on the Grünwald–Letnikov definition:

$$\begin{aligned} {}^{\text{GL}}D^{\beta}F(x,y) \approx & \sum_{k=0}^{K}\ (-1)^k \binom{\beta}{k} F(x-k,y) \\ & + \sum_{k=0}^{K}\ (-1)^k \binom{\beta}{k} F\left(x, y-k\right) \end{aligned} \tag{18}$$

where $\beta \in (0,2)$ is the fractional order, and K is the truncation term, typically set to 8–12 to balance accuracy and computational cost. The binomial coefficient is defined as $\binom{\beta}{k} = \frac{\beta(\beta-1)\cdots(\beta-k+1)}{k!}$.

$$\binom{\beta}{k} = \frac{\beta(\beta-1)\cdots(\beta-k+1)}{k!} \tag{19}$$

This operator not only captures non-local gradient variations in prediction errors but also enhances boundary detail regularization across multiple scales. Compared to integer-order gradients, fractional-order gradients are inherently sensitive to long-range dependencies, significantly improving continuity along weak and fragmented edges.

Through the careful design of this three-part composite loss, FracSegNet achieves not only high pixel-wise segmentation accuracy but also substantial improvements in boundary smoothness and global structural coherence, providing robust support for fine-grained segmentation in regions with complex textures and low contrast.

2.4. Training procedure

The overall training procedure of FracSegNet is summarized in Algorithm 1.

Training procedure of FracSegNet Input: Training images I, ground-truth vessel masks Y Output: Trained model parameters θ [1] 1: Initialize network parameters θ

2: for each training iteration do

3: Sample a mini-batch of images and masks $(I_b,\ Y_b)$

4: Fractional differential preprocessing

5: Compute fractional gradient maps $G_b$using the Grünwald–Letnikov operator

6: Concatenate I_b and G_b along the channel dimension to obtain X_b

7: Forward pass

8: $P_b \leftarrow FracSegNet\,(X_b;\ \theta)$

9:Compute composite loss

10:$L_{Dice} \leftarrow DiceLoss\,(P_b,\ Y_b)$

11: $L_{TV} \leftarrow TVLoss\,(P_b)$

12: $L_{FG} \leftarrow FractionalGradientLoss\,(P_b,\ Y_b)$

13: $L_{total} \leftarrow L_{Dice} + \lambda_{TV} L_{TV} + \lambda_{FG} L_{FG}$

14: Backward pass and parameter update

15: Compute gradients $\frac{\partial L_{total}}{\partial \theta}$

16: Update θ using the Adam optimizer

17: end for

18: return θ

2.5. Problem formulation

Given an input fundus image $I \in \mathbb{R}^{\{H\times W\times 3\}}$ and its corresponding binary vessel mask $Y \in \{0,1\}^{\{H\times W\}}$, FracSegNet aims to learn a mapping F_θ such that $P = F_{\theta(X)}$ approximates Y, where X is the augmented input obtained by concatenating I and its fractional

gradient map G along the channel dimension. The network output $P \in [0,1]^{\{H \times W\}}$ represents the pixel-wise vessel probability map.

The model parameters θ are optimized by minimizing the composite loss:

$$L_{total} = L_{Dice}(P,\ Y) + \lambda_{TV}\ L_{TV}(P) + \lambda_{FG}\ L_{FG}(P,\ Y) \quad (20)$$

where *L_Dice* measures the region-level overlap between P and Y, L_{TV} penalizes local oscillations in P to promote smooth boundaries, and L_{FG} enforces consistency between the fractional-order gradients of the prediction and the ground truth.

2.6. Experimental configuration and evaluation standards

To validate the performance of the proposed method, we conducted retinal vessel segmentation experiments on three public datasets: DRIVE, CHASE_DB1, and STARE. DRIVE consists of 40 retinal images with a resolution of 565 × 584 pixels, each image from a diabetic retinopathy screening program. CHASE_DB1 contains 28 retinal images with 999 × 960 pixels, which are larger in size and contain extensive pathological regions. STARE includes 20 retinal images with 700 × 605 pixels. Unlike other datasets, STARE is relatively balanced in terms of pathological versus normal cases, but it still contains several severely pathological images. For the retinal vessel segmentation task, the official split of the DRIVE dataset is 20/20. For fair comparison, this paper maintained consistency with the splitting standards of comparative methods for the other three datasets, such as Scale and Context Sensitive Network (SCS-Net),[22] Curvilinear Structure Segmentation Network (CS2Net), [23] and Full-Resolution Network (FR-UNet).[24] In this paper, the first 20 images from the DRIVE and CHASE_DB1 datasets and the first 15 images from the STARE dataset were used as training sets, while the remaining 20, 8, and 5 images from the three datasets, respectively, were used as test sets (**Table 1**).

The proposed network framework was implemented using PyTorch. All training and testing experiments were conducted on a workstation equipped with an Intel Core i9-10900X (20-thread) CPU and two NVIDIA GTX 5080 Ti GPUs. During training, we adopted a patch-based strategy and optimized the network using the Adam optimizer, following the established protocol of Iter-Net.[20] To mitigate GPU memory consumption during training, each patch size was set to 30% of the original image dimensions: 169 × 175 pixels for DRIVE, 299 × 288 pixels for CHASE_DB1, and 210 × 181 pixels for STARE. Data augmentation was limited to horizontal and vertical flips to preserve the structural integrity of retinal vessels, and the initial learning rate was set to 0.005. Due to memory constraints, we employed a batch size of 4 and trained the network for a maximum of 6000 iterations. To achieve an optimal balance between segmentation performance and computational efficiency, we selected a network depth (D) of 5 and individually tuned the regularization parameter λ for each dataset based on validation performance.

The evaluation metrics included the receiver operating characteristic (ROC) curve, area under the ROC curve (AUC), accuracy, sensitivity, specificity, and F1 score to objectively and comprehensively assess the performance of the proposed method.

3. Results

3.1. Quantitative comparative analysis

Table 2 presents the experimental results of the proposed method on the DRIVE, STARE, and CHASE_DB1 datasets. The proposed method achieved accuracies greater than 0.9760 and F1 scores greater than 0.8160 across all three datasets, demonstrating its universality in retinal-vessel segmentation. The specificity values were 0.9866, 0.9923, and 0.9874, respectively, indicating that the proposed method has a high recognition rate for retinal image backgrounds. The sensitivity values were 0.8191, 0.8864, and 0.8415, respectively, showing that the proposed method's capillary segmentation results closely match the ground truth annotations.

To validate the segmentation performance of the proposed method, quantitative and qualitative comparative analyses were conducted with 10 other state-of-the-art methods for the same task, including U-Net,[5] Iter-Net,[20] U-Net++,[21] SCS-Net,[22] CS2Net,[23] FR-UNet,[24] Trainable Filters,[25] Wave-Net,[26] ResDO-UNet,[27] and Dual-Path Progressive Fusion Network (DPF-Net).[28] Among these, Iter-Net is a multi-stage deep network, while the others are single U-Net architectures and their variants.[5] The same datasets were used for all experiments to ensure fair comparison.

Tables 3–5 present the segmentation performance metrics of our method compared to other state-of-the-art approaches on the three datasets. As demonstrated, our method achieved superior accuracy and specificity compared to

Table 1. Statistics of experimental datasets

Dataset	Number of images	Training/Test split	Image size (pixels)
DRIVE	40	20/20	565 × 584
CHASE_DB1	28	20/8	960 × 999
STARE	20	15/5	700 × 605

Table 2. Performance analysis of public datasets

Dataset	Accuracy	Sensitivity	Specificity	F1 score	AUC
DRIVE	0.9761	0.8191	0.9866	0.8307	0.9847
CHASE_DB1	0.9784	0.8864	0.9923	0.8400	0.9902
STARE	0.9765	0.8415	0.9874	0.8370	0.9838

Abbreviation: AUC: Area under the receiver operating characteristic curve.

all competing methods, indicating an overall advantage in segmentation performance. By employing hierarchical representation and progressively fusing local details into the global structure layer by layer, our method improved upon U-Net by 1.27%, 3.91%, 0.56%, and 0.79% in accuracy, sensitivity, specificity, and AUC, respectively. Additionally, our method outperformed SCS-Net by 0.64%, 0.28%, and 0.12% in accuracy, specificity, and AUC, respectively. These results demonstrate that our approach effectively captures vessel features across multiple scales. Furthermore, through the integration of texture enhancement and dual-contrast learning, our proposed model achieved better performance gains in accuracy, sensitivity, and specificity than the second-best methods on the CHASE_DB1 and STARE datasets. Notably, while Trainable Filters [25] achieved the best overall performance on DRIVE by specifically addressing vessel structure distortions, our method attained the second-best results on most DRIVE metrics while placing greater emphasis on improving segmentation performance in fine capillaries, the optic disc/cup region, and other pathological areas.

These findings demonstrate that our multi-scale feature representation is highly effective at modeling retinal vessels with varying sizes, shapes, and structures, while the dual-contrast learning mechanism more robustly distinguishes vessels from background regions, suppresses noise interference, and ultimately enhances segmentation accuracy. Furthermore, our proposed model achieved additional performance gains in accuracy, sensitivity, and specificity over the second-best methods on the CHASE_DB1 and STARE datasets. Notably, while Trainable Filters[25] achieved the best overall performance on DRIVE by specifically addressing vessel structure distortions, our method attained the second-best results on most DRIVE metrics while placing greater emphasis on improving segmentation performance in fine capillaries, the optic disc/cup region, and other pathological areas.

These findings demonstrate that our multi-scale feature representation, together with fractional differential preprocessing, adaptive fractional-order convolution, and the fractional gradient-constrained loss, effectively models retinal vessels of varying sizes, shapes, and structures, more robustly distinguishes vessels from background regions, suppresses noise interference, and ultimately enhances segmentation accuracy.

3.2. Qualitative comparative analysis

Figure 2 presents the visual segmentation results of our method on representative images from the DRIVE, STARE, and CHASE_DB1 datasets. Our method demonstrated accurate segmentation of vessels and capillaries with diverse complex morphologies in the DRIVE dataset, precisely delineating not only the main vessel trunks but also preserving the fundamental vessel orientation. Furthermore, for images containing extensive pathological regions in the STARE and CHASE_DB1 datasets, our method achieved accurate vessel segmentation while effectively capturing vessels and capillaries in challenging anatomical areas.

To further validate the effectiveness of our approach for retinal vessel segmentation, we conducted comprehensive visual comparisons with two high-performing U-Net variants across the three datasets, as illustrated in **Figure 3**. FR-UNet[24] integrates multi-scale feature maps from

Table 3. Performance comparison of our proposed method with others on the same task using the DRIVE dataset

Method	**Accuracy**	**Sensitivity**	**Specificity**	**F1 score**	**AUC**
U-Net[5]	0.9634	0.7800	0.9810	0.8141	0.9768
lter-Net[20]	0.9687	0.8177	0.9832	0.8250	0.9839
U-Net++ [21]	0.9686	0.8256	0.9823	0.8192	0.9854
SCS-Net[22]	0.9697	0.8289	0.9838	-	0.9835
CS2Net[23]	0.9553	0.8154	0.9757	0.8228	0.9784
Trainable Filters[25]	0.9912	0.9889	0.9886	-	-
FR-UNet[24]	0.9705	0.8356	0.9837	0.8316	0.9889
DPF-Net[28]	0.9570	0.8279	0.9776	0.8303	0.9824
ResDO-UNet[27]	0.9561	0.7985	0.9791	0.8229	-
Wave-Net[26]	0.9561	0.8164	0.9764	0.8254	-
FracSegNet	0.9761	0.8191	0.9866	0.8307	0.9847

Abbreviation: AUC: Area under the receiver operating characteristic curve.

Table 4. Performance comparison of our proposed method with others on the same task using the STARE dataset

Method	**Accuracy**	**Sensitivity″**	**Specificity**	**F1 score**	**AUC**
U-Net[5]	0.9712	0.8167	0.9838	0.8118	0.9857
lter-Net[20]	0.9729	0.8287	0.9846	0.8146	0.9874
U-Net++[21]	0.9733	0.8264	0.9851	0.8150	0.9883
SCS-Net[22]	0.9751	0.8405	0.9861	-	0.9899
CS2Net[23]	0.9752	0.8816	0.9840	0.8420	0.9932
Trainable Filters[25]	0.9708	0.8479	0.9840	-	-
FR-UNet[24]	0.9748	0.8798	0.9814	0.8330	0.9913
DPF-Net[28]	0.9655	0.8287	0.9854	0.8366	0.9898
ResDO-UNet[27]	0.9542	0.7959	0.9786	0.8231	-
Wave-Net[26]	0.9641	0.7902	0.9836	0.8140	-
FracSegNet	0.9784	0.8864	0.9923	0.8400	0.9902

Abbreviation: AUC: Area under the receiver operating characteristic curve.

Table 5. Performance comparison of our proposed method with others on the same task using the CHASE_DB1 dataset

Method	**Accuracy**	**Sensitivity**	**Specificity**	**F1 score**	**AUC**
U-Net[5]	0.9676	0.7713	0.9808	0.7898	0.9783
lter-Net[20]	0.9752	0.8303	0.9850	-	0.9861
U-Net++[21]	0.9753	0.8317	0.9850	0.8015	0.9861
SCS-Net[22]	0.9744	0.8365	0.9839	-	0.9867
CS2Net[23]	0.9721	0.8344	0.9853	0.8288	0.9864
Trainable filters[25]	0.9641	0.8301	0.9793	-	-
FR-UNet [24]	0.9752	0.8327	0.9869	0.8151	0.9920
DPF-Net [28]	0.9676	0.8303	0.9839	0.8302	0.9868
ResDO-UNet[27]	0.9664	0.8284	0.9821	0.8349	-
Wave-Net[26]	0.9672	0.8020	0.9794	0.8236	-
FracSegNet	0.9765	0.8415	0.9874	0.8370	0.9838

Abbreviation: AUC: Area under the receiver operating characteristic curve.

adjacent stages to incorporate high-level contextual information, while CS2Net[23] employs self-attention mechanisms in both encoder and decoder stages to learn rich hierarchical representations of curvilinear structures. Compared with CS2Net and FR-UNet, our method produced more complete segmentation of fine capillaries and improved vessel continuity, particularly in low-contrast regions and around the optic disc/cup. These qualitative improvements are consistent with the proposed fractional differential preprocessing and adaptive fractional-order convolution, which enhance weak vessel edges and support multi-scale vessel representation, together with the fractional gradient–constrained loss that promotes structural continuity and suppresses noise-induced artifacts. In representative examples, our method more accurately delineated thin terminal branches under low fundus contrast and exhibited robust performance in pathological regions and near the optic disc/cup. Overall, these visualizations confirm that the proposed fractional-order modeling framework effectively segments vessels across multiple morphological scales while maintaining strong continuity in challenging anatomical regions.

Figure 4 presents a magnified analysis of specific vessel details in the segmentation results. While CS2Net[23] employs self-attention mechanisms and FR-UNet[24] considers adjacent-scale features, our method's scale feature representation combines dilated convolution and average pooling to enrich multi-scale information. The first row **in Figure 4C–F** demonstrates segmentation results for normal images. Compared to the other two methods, our approach achieved more accurate segmentation of vessels and capillaries with various shapes and sizes (highlighted in red and green boxes). Compared with the other two methods, our approach yields more continuous thin-vessel segmentation in low-contrast regions, consistent with the fractional-order edge/gradient enhancement and fractional regularization. Our method improved capillary segmentation accuracy by enhancing vessel-edge definition. The second and third rows show input images with low brightness and abnormal regions, in which background areas are prone to being misclassified as vessels. Since the competing methods do not adequately address the relationship between the vessel foreground and the surrounding background in retinal images, their segmentation results were inaccurate. With fractional-order modeling and the fractional gradient–constrained loss, our method improved segmentation of crossing vessels at the optic disc (bottom panels of each input image in **Figure 4C**; red box) and achieved greater continuity for capillaries in low-contrast regions (top panels of each input image in **Figure 4C**; green box). The vessel misclassification in pathological regions (green box) and capillary omission in low-contrast areas (red box) in the second input image were also significantly improved. Experimental results demonstrate that our method produces segmentation results better aligned with the ground truth while enhancing segmentation performance for vessels and capillaries in challenging regions such as the optic cup/disc and pathological areas, thereby improving overall precision in retinal vessel segmentation.

3.3. Complexity analysis

To further evaluate the computational complexity of our method, we compared it against several state-of-the-art approaches under identical hardware conditions, measuring both the number of parameters and processing time. **Table 6** presents the complexity comparison results on the DRIVE dataset. Despite incorporating additional fractional-order components, our model maintains a moderate parameter count and competitive inference times compared with Iter-Net and other state-of-the-art approaches. Overall, our approach demonstrates a clear efficiency advantage.

Furthermore, on the DRIVE and STARE datasets, the end-to-end processing time per fundus image (including inference and input/output operations) was only 0.833 and 0.865 s, respectively—both were well under 1 s. Even for the higher-resolution CHASE_DB1 images, the processing time was 4.705 s, remaining below 5 s and well-suited for clinical deployment. Collectively, our method offers significant advantages in both segmentation accuracy and computational efficiency, establishing it as a strong candidate for practical retinal vessel analysis.

3.4. Ablation study analysis

To quantify the independent and combined effects of FracSegNet's three key innovations—fractional preprocessing, adaptive fractional-order convolution, and fractional gradient–constrained loss—we proposed the following ablation experiments on the DRIVE dataset (**Table 7**).

3.5. Loss function convergence analysis

To evaluate the optimization behavior of the proposed FracSegNet, we visualized the convergence curves of its composite loss function over 200 training epochs, as shown in **Figure 5**. The total

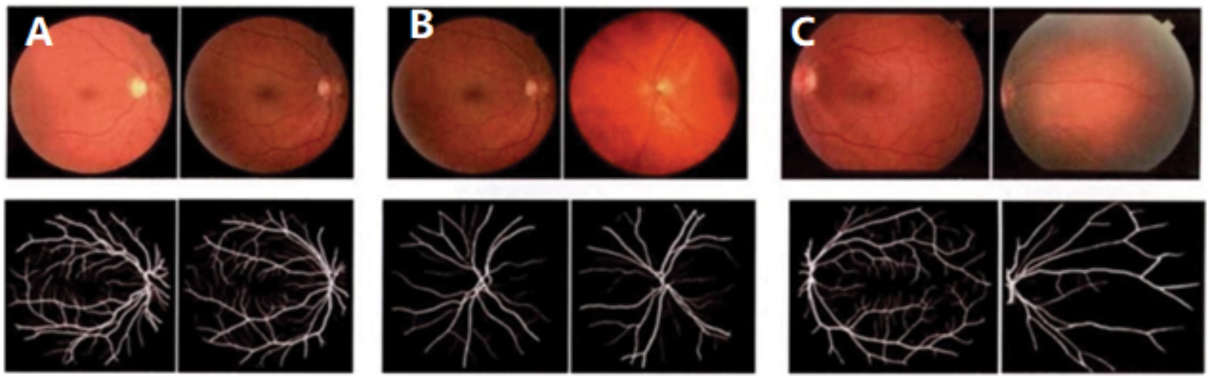

Figure 2. Segmentation results of the proposed method using the (A) DRIVE, (B) CHASE_DB1, and (C) STARE datasets

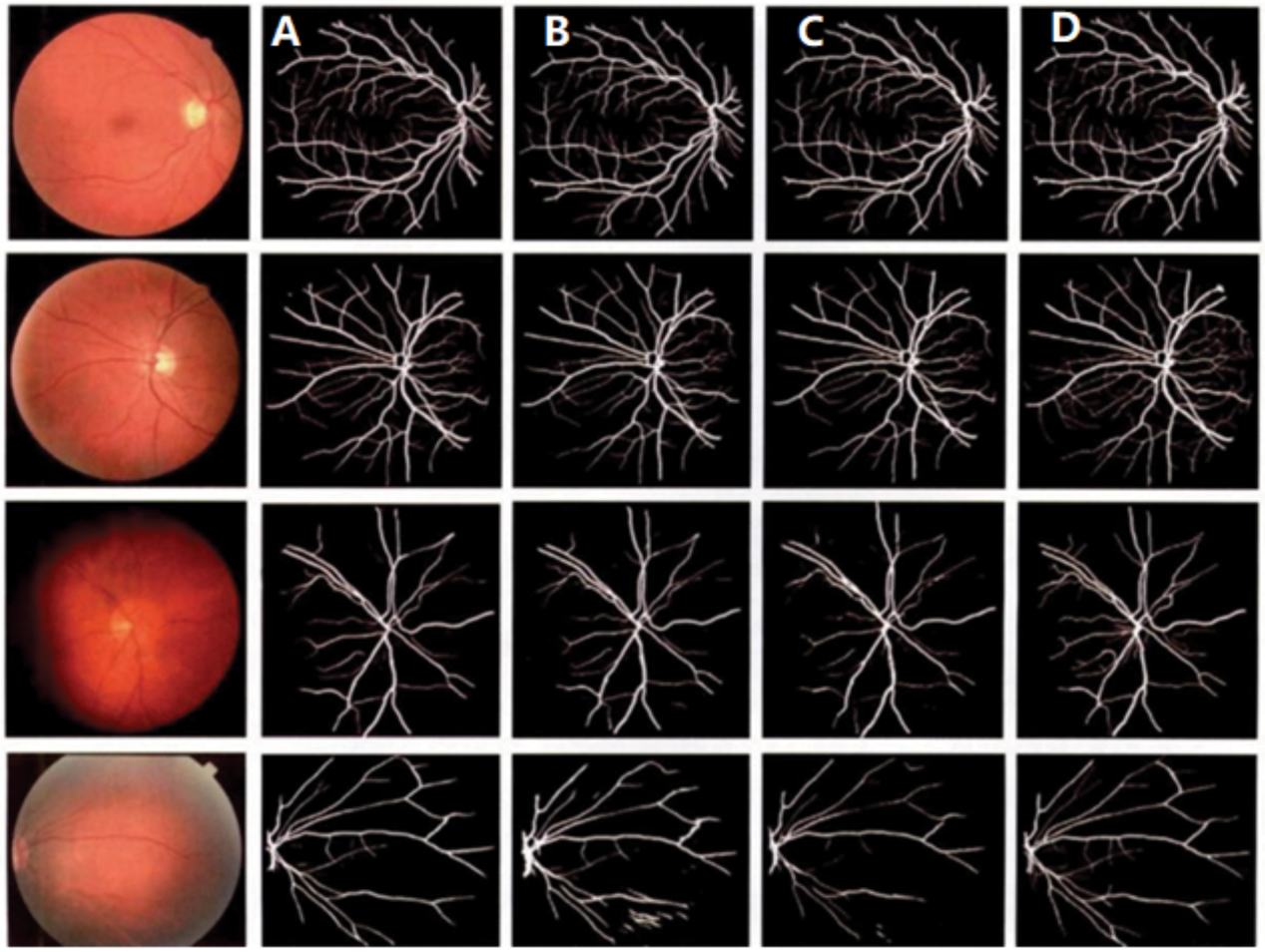

Figure 3. Qualitative comparison of the proposed method with the U-Net variants. (A) Input images. Probability maps of (B) our proposed method, (C) CS2Net, and (D) FR-UNet. (E) Ground truths

Table 6. Complexity comparison of the proposed method with the other methods using the DRIVE dataset

Method	**GPU**	**Time (s)**	**Parameter (M)**
U-Net	NVIDIA GTX 5080	0.440	4.21
Iter-Net	NVIDIA GTX 5080	0.923	10.8
FracSegNet	NVIDIA GTX 5080	0.833	9.87

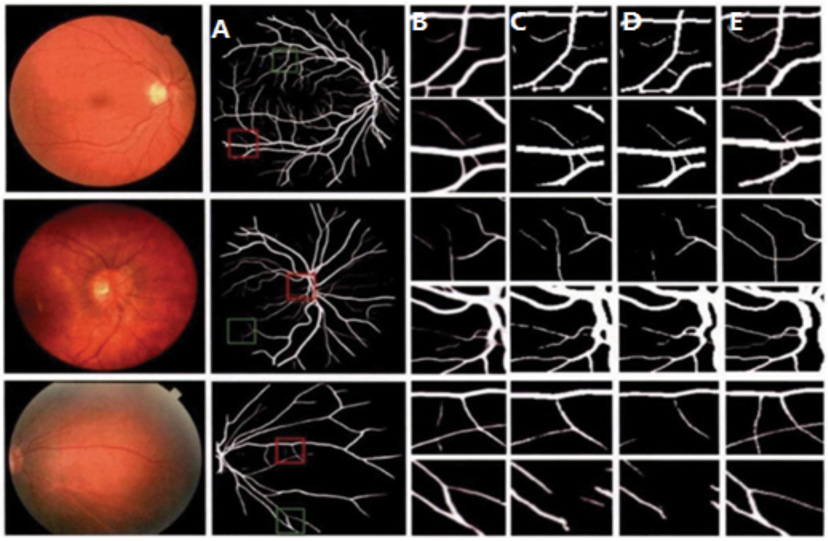

Figure 4. Comparison of patches' segmentation results using our proposed method and the U-Net variants. (A) Input images. (B) Probability maps of our proposed method. Enlarged images of (C) our method, (D) CS2Net, (E) FR-UNet, and (F) ground truths

Table 7. Ablation studies on the DRIVE dataset

Configuration	**Description**	**Dice (%)**	**HD95 (mm)**	**PSNR (dB)**	**SSIM**
B	Backbone only: standard U-Net (no FP, no AFC, Dice + TV loss)	78.5	12.3	24.1	0.842
B + FP	Backbone + FP; U-Net input augmented with FG channels	80.6	10.8	25.5	0.864
B + AFC	Backbone + AFC; all convolutions replaced by AFC blocks	81.2	10.2	25.9	0.868
B + GL	Backbone + FG–constrained loss; loss = Dice + TV + FG loss	79.8	11.5	24.8	0.855
B + FP + AFC	Backbone with FP and AFC	83.0	9.5	26.3	0.878
B + FP + GL	Backbone with FP and GL	82.1	9.8	26.1	0.873
B + AFC + GL	Backbone with AFC and GL	82.5	9.7	26.2	0.875
FracSegNet	Backbone with FP, AFC, and GL (full FracSegNet)	84.3	9.0	26.7	0.882

Abbreviations: AFC: Adaptive fractional-order convolution; FG: Fractional gradient; FP: Fractional preprocessing; GL: Fractional gradient–constrained loss; HD95: 95th Percentile hausdorff distance; PSNR: Peak signal-to-noise ratio; SSIM: Structural similarity index measure; TV: Total variation.

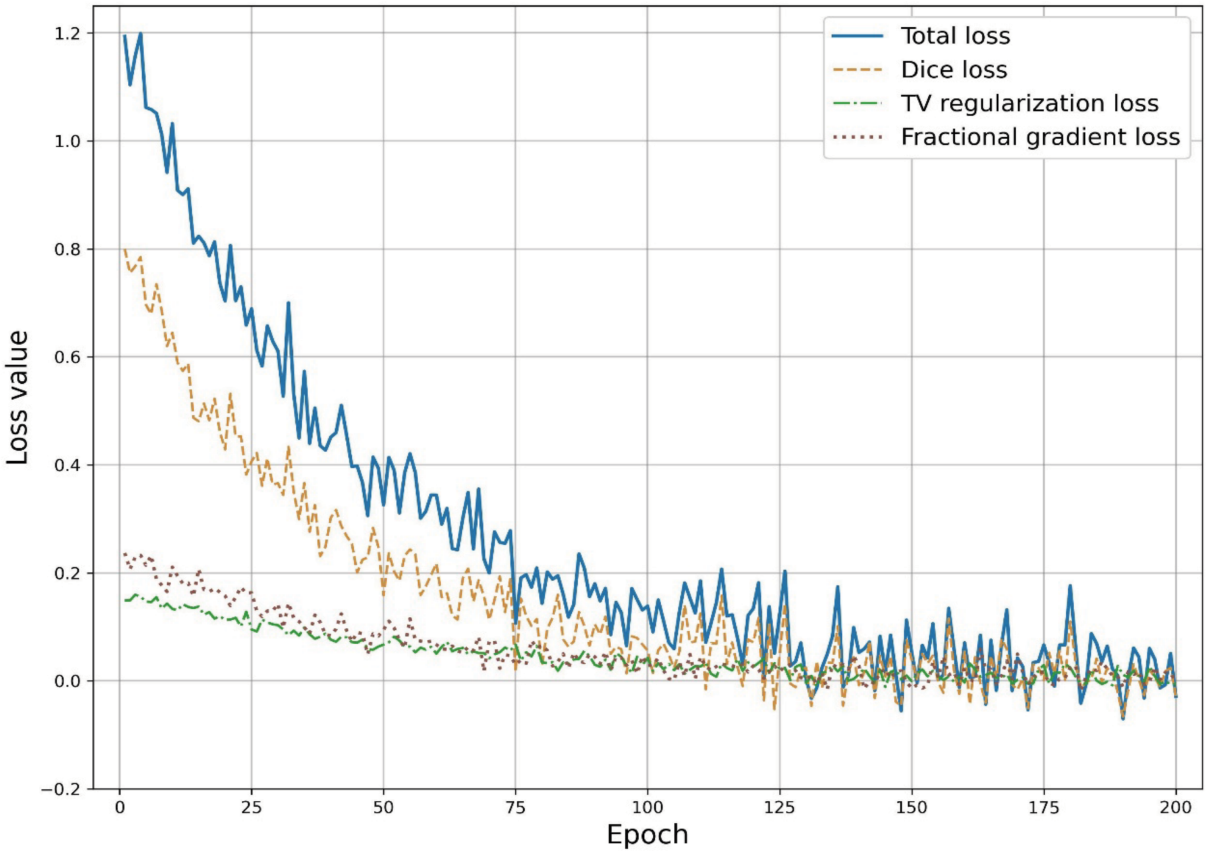

Figure 5. Loss function convergence in FracSegNet training

loss consists of three main components: Dice loss, TV regularization loss, and the fractional gradient constraint loss. The Dice loss, which directly reflects segmentation accuracy, rapidly decreased during the early stages of training and gradually stabilized as the model converged. Meanwhile, the TV loss and fractional gradient loss exhibited smoother decay trends, highlighting their roles in refining boundary smoothness and enhancing edge continuity. Overall, the total loss steadily decreased with training, indicating effective convergence and synergistic optimization among the three loss components. This result validates the effectiveness of integrating fractional-order regularization into the training process, improving pixel-wise accuracy, promoting structural consistency, and reducing edge artifacts in complex retinal images.

4. Discussion

In the current study, FracSegNet demonstrates consistent improvements over strong CNN baselines and recent U-Net variants in terms of accuracy, sensitivity, and specificity on DRIVE, STARE, and CHASE_DB1. The qualitative results further indicate that our model is particularly effective in recovering weak-edge continuity, preserving tiny terminal branches, and reducing false positives in pathological regions and around the optic disc.

However, several limitations remain. First, all experiments were conducted on public datasets with well-controlled acquisition conditions. Real-world clinical images may exhibit larger variations in illumination, artifacts, and disease patterns. Second, our implementation relies on a fixed Grünwald–Letnikov fractional operator; adaptive or learnable fractional operators could further improve flexibility. Third, we focused on 2D fundus images, while many clinical applications require 3D or multimodal data, such as optical coherence tomography and magnetic resonance angiography. These aspects will be explored in our future work.

From a clinical perspective, our current experiments were limited to publicly available datasets acquired under relatively controlled imaging conditions. In real-world practice, fundus images may exhibit substantial variations in illumination, noise, artifacts, and disease patterns across different hospitals and devices. Therefore, additional validation on large-scale, multi-center clinical datasets is required before FracSegNet can be deployed in routine screening or diagnosis workflows.

Regarding ethical and privacy considerations, this study only used fully anonymized, publicly accessible datasets and thus did not involve identifiable patient information. For future clinical deployment, strict data governance, secure storage, and privacy-preserving learning strategies (e.g., federated or distributed training) will be essential. Institutional review board approval and informed consent procedures will also be required when training or evaluating the model on hospital-owned data.

5. Conclusion

In this study, we proposed FracSegNet, a retinal vessel segmentation network that integrates fractional-order modeling at the preprocessing, feature extraction, and loss levels. The main contributions can be summarized as follows. First, we introduced a fractional differential preprocessing module based on the Grünwald–Letnikov operator, which generates multi-directional fractional gradient maps and enhances weak edges and delicate vessel structures when concatenated with the original fundus image. Second, we designed adaptive fractional convolution blocks that dynamically fuse integer-order and fractional-order responses via learnable order weights, enabling the network to capture both local details and long-range dependencies at each encoder–decoder stage. Third, we incorporated a fractional gradient-constrained loss into a composite objective that combines Dice loss and TV regularization, thereby improving boundary continuity and suppressing noise-induced artifacts.

Extensive experiments on the DRIVE, STARE, and CHASE_DB1 datasets demonstrate that FracSegNet achieved competitive or superior performance compared with several state-of-the-art methods, including U-Net, U-Net++, Iter-Net, SCS-Net, CS2Net, and FR-UNet. The quantitative results in **Tables 2–5** show consistent gains in accuracy, sensitivity, specificity, and F1 score, while the qualitative visualizations in **Figures 2–4** clearly highlight more continuous vessel networks and better preservation of tiny capillaries, especially in low-contrast and pathological regions.

Our work complements existing research on fractional-order modeling in medicine, such as fractional diabetes systems and fractional predator–prey models, by demonstrating that fractional calculus can also be effectively integrated into deep segmentation architectures for medical imaging. In future work, we plan to extend FracSegNet to 3D and multimodal data, to explore end-to-end learnable fractional operators, and to validate the model on large-scale, multi-center clinical datasets under real-world conditions.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Shoutong Huang, Yu Ma
Formal analysis: All authors
Investigation: Shoutong Huang
Methodology: All authors
Writing–original draft: Shoutong Huang
Writing–review & editing: Yu Ma, Huitan Chang, Bowen Xiao

Availability of data

The experimental data that support the findings of this study are available on GitHub https://github.com/DIAGNijmegen/drive-vessels-unet?tab=readme-ov-file, with the identifier available at https://www.isi.uu.nl/research/databases/ and https://aistudio.baidu.com/datasetdetail/28292.

AI tools statement

All authors confirm that no AI tools were used in the preparation of this manuscript.

References

1. Wang Y, Wang J, Guo P. Eye-UNet. A UNet-based network with attention mechanism for low-quality human eye image segmentation. *Signal Image Video Process.* 2023;17(4):1097-1103.https://doi.org/10.1007/s11760-022-02316-x
2. Chandra IS, Shastri RK, Kavitha D, Kumar KR, Manochitra S, Babu PB. CNN based color balancing and denoising technique for underwater images: CNN-CBDT. *Meas Sens.* 2023;28:100835. https://doi.org/10.1016/j.measen.2023.100835
3. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *MICCAI 2015.* Cham: Springer; 2015:234-241. https://doi.org/10.1007/978-3-319-24574-4_28

4. Coelho C, Costa MFP, Ferrás LL. Fractional calculus meets neural networks for computer vision: a survey. *AI.* 2024;5(3):1391-1426. https://doi.org/10.3390/ai5030067
5. Stember JN, Celik H, Krupinski E, Chang PD, Mutasa S, Lefkowicz A. Eye tracking for deep learning segmentation using convolutional neural networks. *J Digit Imaging.* 2019;32(4):597-604. https://doi.org/10.1007/s10278-019-00220-4
6. Chen BZ, Liu YS, Zhang Z, et al. TransAttUnet: multi-level attention-guided U-Net with transformer for medical image segmentation. *IEEE Trans Emerg Top Comput Intell.* 2024;8(1):55-68.https://doi.org/10.1109/TETCI.2023.3309626
7. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. In: *Proc ICLR*; 2021. Available at: https://openreview.net/forum?id=YicbFdNTTy
8. Chen JN, Cheng J, Wang X, et al. TransUNet: transformers make strong encoders for medical image segmentation. arXiv. Preprint posted online February 8, 2021. Available at: https://arxiv.org/abs/2102.04306
9. Azad R, Al-Antary MT, Heidari M, Merhof D. TransNorm: transformer provides a strong spatial normalization mechanism for a deep segmentation model. *IEEE Access.* 2022;10:108205-108215. https://doi.org/10.1109/ACCESS.2022.3211501
10. Guo DF, Terzopoulos D. A transformer-based network for anisotropic 3D medical image segmentation. In: *Proc ICPR.* Milan, Italy; 2021:8857-8861. https://doi.org/10.1109/ICPR48806.2021.9411990
11. Zhang ZA, Wu CD, Coleman S, Tsakanikas V. DENSE-INception U-net for medical image segmentation. *Comput Methods Programs Biomed.* 2020;192:105395. https://doi.org/10.1016/j.cmpb.2020.105395
12. Liu HX, Fang JX, Zhang ZJ, et al. Localized edge-region-based active contour for medical image segmentation. *IET Image Process.* 2021;15(7):1567-1582.https://doi.org/10.1049/ipr2.12126
13. Wang J, Wang W, Hu X, et al. Black-winged kite algorithm: a nature-inspired meta-heuristic for solving benchmark functions and engineering problems. *Artif Intell Rev.* 2024;57(4):1-53. https://doi.org/10.1007/s10462-024-10723-4
14. Arora R, Kumar S, Singh P. Fractional-order image enhancement using deep learning for medical imaging applications. *Neurocomputing.* 2022;501:234-246
15. Raubitzek S, Sabadas R, Latzko P. Combining fractional derivatives and convolutional neural networks for robust image denoising. *Signal Process Image Commun.* 2023;28:100835. https://doi.org/10.1016/j.imagecom.2022.116947
16. Cao H, Wang Y, Chen J, et al. Swin-Unet: UNet-like pure transformer for medical image segmentation. In: *Med Comput Vis Algorithms Big Data.* Cham: Springer; 2022:205-218. https://doi.org/10.1007/978-3-031-25066-8_9
17. Chen B, Liu Y, Zhang Z, Lu G, Kong AWK. TransAttUnet: Multi-level attention-guided U-Net with transformer for medical image segmentation. *IEEE Trans Emerg Top Comput Intell.* 2023;7(5):1150-1163.https://doi.org/10.1109/TETCI.2023.3309626
18. Jiang L, He Y, Zhang J. TransDualSegNet: Transformer dual-segmentation network for retinal vessel extraction in OCT images. *Comput Biol Med.* 2025;173:108023. https://doi.org/10.1016/j.compbiomed.2024.108023
19. Dai J, Chen W, Chen R, et al. Research on task assignment algorithm of heterogeneous aircraft cooperative cluster in dynamic scene. *Comput Electr Eng.* 2023;110:108781.https://doi.org/10.1016/j.compeleceng.2023.108781
20. Li X, Xie H, Zhang Z, Cai J. IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks. In: *Proc IEEE WACV*; 2020. https://doi.org/10.1109/WACV45572.2020.9093621
21. Zhou Z, Siddiquee MR, Tajbakhsh N, Liang J. UNet++: A nested U-Net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA 2018)*; 2018:3-11. https://doi.org/10.1007/978-3-030-00889-5_1
22. Wu Y, Sang N, Xiang T. SCS-Net: scale and context sensitive network for retinal vessel segmentation. *Med Image Anal.* 2021;70:102025.https://doi.org/10.1016/j.media.2021.102025
23. Mou L, Zhao Y, Chen L, et al. CS2-Net: Curvilinear structure segmentation network for retinal vessel extraction. *Med Image Anal.* 2021;67:101874.https://doi.org/10.1016/j.media.2020.101874
24. Liu H, Fang J, Zhang Z. FR-UNet: Full-resolution network for medical image segmentation. *IET Image Process.* 2022;16(7):1567-1582. https://doi.org/10.1109/JBHI.2022.3188710
25. Abdulsahib G, Al-Antary MT, Merhof D. Trainable filters for retinal vessel segmentation. *Signal Process Image Commun.* 2022;114:116947. https://doi.org/10.1016/j.imagecom.2022.116947
26. Liu X, Wang Y, Zhang Z. WaveNet: Multi-frequency representation learning for retinal vessel segmentation. *Comput Biol Med.* 2023;170:106337. https://doi.org/10.1016/j.compbiomed.2022.106337

27. Liu X, Fang J, Chen L. ResDO-UNet: Dual-orthogonal residual learning for retinal vessel segmentation. *Biomed Signal Process Control.* 2023;84:104810. https://doi.org/10.1016/j.bspc.2023.104810
28. Li J, Gao G, Yang L, Bian G, Liu Y. DPF-Net: A dual-path progressive fusion network for retinal vessel segmentation. *IEEE Trans Instrum Meas.* 2023;72:1-17

Shoutong Huang, *a postgraduate of the School of Electronic and Electrical Engineering, Ningxia University now; have got two bachelor degrees of engineering in Tianjin university of technology; have published several articles belong to different types; research interests are image processing, computer vision, pattern recognition, and so on.*
https://orcid.org/0009-0006-9560-3040

Yu Ma, *a professor of Ningxia University; Senior member of the China Computer Society, national graduate education evaluation and monitoring dissertation review expert, "LCD and Display" journal reviewer expert, expert of the expert database of the Autonomous Region Department of Science and Technology; comprehensive bid evaluation expert, government procurement evaluation expert, and public institution recruitment interview expert of the autonomous region; research interests include image processing, computer vision, and pattern recognition; presided over and participated in 11 scientific research projects; 23 papers with all participation (including 1 rank); has won 3 software copyrights, registered 1 scientific and technological achievement in Ningxia, and won the third prize of the Ningxia Hui Autonomous Region Science and Technology Progress Award 1 time.*
https://orcid.org/0009-0005-8764-6035

Huitan Chang, *a postgraduate of School of Electronic and Electrical Engineering, Ningxia University now, research interests include image processing, computer vision, and pattern recognition.*
https://orcid.org/0009-0006-9560-3040

Bowen Xiao, *a postgraduate of School of Electronic and Electrical Engineering, Ningxia University now; research interests include image processing, computer vision, and pattern recognition.*
https://orcid.org/0009-0006-9560-3040

An International Journal of Optimization and Control: Theories & Applications
(https://accscience.com/journal/ijocta)