Retinal Vessel Analysis and Segmentation: Challenges and Future Directions

Hao ShiLong, Dong YiZhi, Wang ZiFan, Dong WenYu, Zhang TongYu

December 2025

Abstract

Retinal vessel segmentation is a critical foundation for the early screening and diagnosis of diabetic retinopathy (DR) and other ocular diseases. To address the challenges of low accuracy in fine branch extraction and poor restoration of complex topological structures in existing segmentation methods, this study conducts a systematic comparative analysis of three mainstream deep learning models: UNet, UNet++, and TransUNet. Experiments are carried out on two public datasets (DRIVE and STARE), with qualitative evaluation from three core dimensions—continuity of main vessels, fine branch extraction ability, and vascular topological structure restoration—and quantitative assessment using metrics including Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), and Area Under the ROC Curve (AUC). The results show that the performance differences among the three models originate from their structural designs and feature extraction mechanisms: UNet achieves stable segmentation of main vessels but suffers from 15%-20% fine branch loss and stiff main-branch transitions; UNet++ improves fine branch retention to around 80% and enhances the naturalness of main-branch transitions through multi-scale feature fusion, balancing efficiency and accuracy; TransUNet, leveraging the global attention mechanism of the Transformer module, breaks through the local receptive field limitation of convolutional neural networks, achieving a fine branch retention rate exceeding 90%, optimal main vessel smoothness, and the most accurate restoration of complex topological structures. Its quantitative performance on the DRIVE dataset (Acc=0.7952, Sen=0.8245, Spe=0.9856, AUC=0.9821) outperforms the other two models. This study constructs a multi-dimensional qualitative evaluation system that aligns with clinical diagnostic needs, providing a reference for model selection in retinal vessel segmentation tasks. The research results offer technical support for the development of DR auxiliary diagnosis systems and hold significant theoretical and practical value.

Retinal Vessel Segmentation; UNet; UNet++; TransUNet; Deep Learning; Diabetic Retinopathy; Image Segmentation

1 Introduction

Retinal vessels represent the sole deep microvascular network in the human body that can be directly visualized in a non-invasive manner. Alterations in their morphological and structural characteristics, including diameter, curvature, and branching patterns, are intricately associated with the progression of various ocular and systemic disorders. Consequently, retinal vessel analysis techniques hold substantial promise for aiding clinicians in the early screening, diagnosis, and longitudinal monitoring of ocular pathologies such as diabetic retinopathy (DR), glaucoma, hypertensive retinopathy, and retinal vein occlusion (RVO), as well as systemic conditions including arteriosclerosis. Among these characteristics, vascular morphological parameters (e.g., vessel diameter, curvature, and arteriovenous ratio) exhibit profound clinical significance for disease assessment.

Nevertheless, the intricate anatomical structure of retinal vessels renders automatic segmentation a formidable challenge, primarily attributed to the following factors: (1) suboptimal image quality with multifarious noise artifacts; (2) poor contrast between vessels and the surrounding background; (3) inhomogeneous image illumination; (4) diverse vascular morphologies and complex topological architectures; (5) significant inter-vessel scale variations; (6) prominent interference from pathological lesions; and (7) inadequate generalization capability of existing models. Collectively, these hurdles impede the accurate and robust segmentation of retinal vessels, making it a persistent bottleneck in clinical translational research.

To address these aforementioned challenges, future research endeavors may pursue breakthroughs in several directions. First, model optimization and architectural innovation are crucial; this includes developing lightweight and computationally efficient model architectures via model compression strategies, exploring the potential of emerging frameworks like Graph Neural Networks (GNNs) for modeling global topological interdependencies, and integrating self-supervised or semi-supervised learning paradigms to mitigate the reliance on large-scale, meticulously annotated datasets. Second, multimodal information fusion should be leveraged by combining the complementary advantages of color fundus photography, OCT angiography (OCTA) [1], and ultra-widefield imaging to enhance accuracy and robustness. Third, advancing from two-dimensional (2D) image segmentation to three-dimensional (3D) vascular tree reconstruction will provide comprehensive and continuous structural information for better clinical diagnosis and prognosis evaluation.

In recent years, deep learning has emerged as a powerful paradigm in medical image analysis. Notably, fully convolutional network (FCN)-based models, particularly U-Net, have become mainstream approaches for image segmentation owing to their inherent capacity for end-to-end pixel-level prediction and effective multi-scale feature fusion. [2] Against this backdrop, the present study aims to construct a fundus vessel segmentation model integrating U-Net and Transformer architectures. By optimizing feature extraction and fusion mechanisms, this model is designed to achieve high-precision segmentation of both major retinal vessels and their delicate capillary branches.

2 Related Work

This section reviews the evolution of methods for retinal vessel segmentation in fundus images over recent years, focusing on three main categories: traditional methods, U-Net and its variants, and Transformer-based approaches.

2.1 Traditional Methods

In the field of auxiliary diagnosis for ocular diseases such as diabetic retinopathy, traditional retinal vessel segmentation methods do not rely on deep learning techniques at all, and their core is supported by three technical pillars: basic image processing tools, mathematical morphology theories, and classical machine learning algorithms. To accurately extract vessel structures from fundus images, such methods usually set up a multi-step preprocessing pipeline: first, edge detection algorithms are used to outline the contour boundaries of vessels; then, various filtering operations are applied to weaken background noise and artifact interference in the images; finally, threshold segmentation technology is adopted to achieve the initial separation of vascular and non-vascular regions, thereby improving the accuracy of subsequent vessel feature extraction. In terms of specific technical applications, common strategies include: using matched filters to identify linear patterns that are highly consistent with vessel morphology, and capturing the position information of slender vessels in a targeted manner; applying Frangi vesselness filters to enhance the tubular structure features in images, highlighting the morphology of the main trunk and branches of vessels[3]; combining mathematical morphological operations such as dilation and erosion to further eliminate noise spots and isolated artifacts in segmentation results, and optimize the integrity of the vascular network. However, these traditional methods also have obvious limitations—when dealing with complex fundus images caused by lesions such as vessel distortion and exudate occlusion, their segmentation accuracy and robustness tend to decrease significantly.

2.2 U-Net and Its Variants

The introduction of the U-Net architecture by Ronneberger et al. in 2015 represented a landmark breakthrough in the field of biomedical image segmentation, revolutionizing the way researchers tackle the challenges of target extraction from medical images.[2] Characterized by a symmetric encoder-decoder structure, U-Net's most distinctive innovation lies in its skip connections: the encoder path performs successive downsampling operations to capture high-level semantic features of images, while the decoder path conducts upsampling to restore spatial resolution; meanwhile, the skip connections bridge the encoder and decoder by fusing high-resolution detail features from the contracting path with the upsampled features in the expansive path, effectively compensating for the loss of edge and texture information during downsampling and thus enhancing segmentation accuracy significantly. Since its inception, a vast array of U-Net variants have been developed to address the unique challenges of retinal vessel segmentation—such as the wide variation in vessel diameters, the susceptibility of microvessels to noise interference, and the irregular morphology of vessels in lesion areas—thereby further improving segmentation performance. For instance, U-Net++ optimizes the skip connection pathways and increases the levels of feature fusion, resolving the issue of insufficient integration between shallow and deep features in the original U-Net.[4] Attention U-Net incorporates attention mechanisms that enable the model to automatically focus on vascular target regions while suppressing background noise, making it particularly suitable for vessel segmentation in pathological fundus images. Additionally, various U-Net-based designs integrated with multi-scale processing strategies build multi-branch feature extraction modules, which can accurately capture vessels of different sizes and have further advanced the development of retinal vessel segmentation technology.

2.3 Transformer-Based Approaches

In recent years, Transformer-based approaches have emerged as a powerful alternative to address the inherent limitations of pure Convolutional Neural Networks (CNNs) in biomedical image segmentation—specifically, the difficulty in modeling long-range spatial dependencies between image pixels. At the core of these Transformer models lies the self-attention mechanism, which enables dynamic weight assignment for any pair of pixels across the entire image. This mechanism allows the model to capture latent correlations between distant pixels, effectively overcoming the constraint of the local receptive field inherent in convolutional operations[5][6]. In the field of retinal vessel segmentation, hybrid models that integrate CNNs and Transformers have rapidly gained prevalence and become a mainstream research direction. These models adopt a synergistic design strategy: on one hand, convolutional layers are leveraged for their strong capability in local feature extraction, which can quickly identify fine-grained details such as vessel edge contours and diameter variations; on the other hand, Transformer modules are employed to model global contextual information, mapping the topological connections of the tree-like retinal vessel network and clarifying the relationships between the main vascular trunk and its hierarchical branches. For retinal vessels with complex tree-like structures, this feature fusion strategy of "local details plus global correlations" proves to be particularly beneficial. Even when dealing with pathological fundus images where vessels exhibit tortuosity or discontinuity, hybrid models can still maintain high segmentation integrity and accuracy by capturing the global dependencies of the vascular network.

3 Model Architecture and Mathematical Foundations

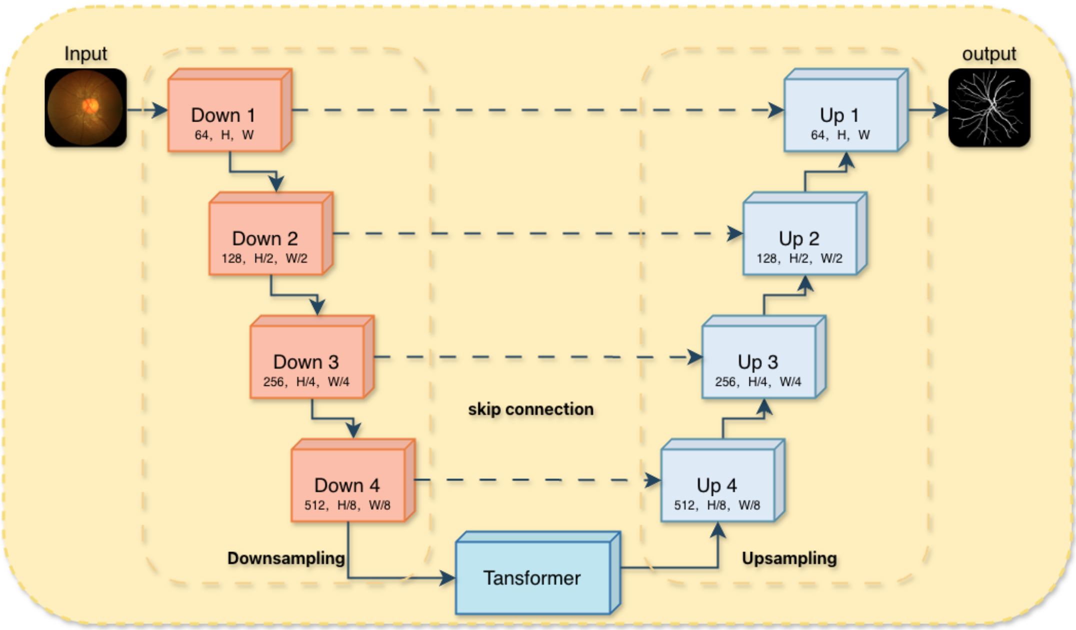

Figure 1: Proposed U-Net + Transformer Model Architecture for Retinal Vessel Segmentation

3.1 Multi-scale Feature Extraction via U-Net Backbone

The proposed model utilizes U-Net as the structural backbone.[2] The input consists of pre-processed fundus images $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$, where $H = 512$ and $W = 512$. During the Encoder stage, the feature evolution path follows:

$$(512, 512, C) \xrightarrow{\text{Conv}} (512, 512, 64) \xrightarrow{\cdots} (32, 32, 1024)$$

In the Decoder stage, resolution is restored through transposed convolutions. Skip connections facilitate the fusion of local spatial details with global semantic information:

$$\mathbf{X}^l_{dec} = \text{Concat}\left(\text{Up}(\mathbf{X}^{l+1}_{dec}), \mathbf{X}^l_{enc}\right) \tag{1}$$

3.2 Global Contextual Modeling via Bottleneck Transformer

To overcome the restricted receptive field of CNNs, a Transformer module is embedded at the bottleneck layer. The 2D feature mapping $\mathbf{F} \in \mathbb{R}^{h \times w \times d}$ is transformed into a 1D sequence with learnable 2D position embeddings $\mathbf{P} \in \mathbb{R}^{N \times d}$:

$$\mathbf{Z}_0 = [\mathbf{f}_1; \mathbf{f}_2; \ldots; \mathbf{f}_N] + \mathbf{P} \tag{2}$$

The core Multi-Head Self-Attention (MHSA) is calculated as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} + \mathbf{B}\right)\mathbf{V} \tag{3}$$

where $\mathbf{B}$ represents the Relative Position Bias. Transformer layers use residual connections and Layer Normalization (LN):

$$\mathbf{Z}_l = \text{LN}(\mathbf{Z}'_l + \text{MLP}(\mathbf{Z}'_l)) \tag{4}$$

3.3 Synergistic Optimization of Local Details and Global Topology

While convolutional layers extract local contrast, the Transformer establishes a global interaction function f_{global}:

$$y_i = \sum_{j=1}^{N} \omega(i, j) v_j \tag{5}$$

This collaborative mechanism allows the model to "logically reconnect" broken vessel segments, significantly enhancing the **topological connectivity** of the retinal vessel maps.

4 Experiments and Results

4.1 Datasets and Evaluation Metrics

To evaluate the performance of the proposed model, we conduct experiments on two widely-used public retinal image datasets: **DRIVE** (Digital Retinal Images for Vessel Extraction). The DRIVE dataset contains 40 images with a resolution of 565×584.[7]

We use four standard metrics to quantify the segmentation performance: Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), and the Area Under the ROC Curve (AUC). The formulas are as follows:

- Acc $= \frac{TP+TN}{TP+TN+FP+FN}$
- Sen $= \frac{TP}{TP+FN}$
- Spe $= \frac{TN}{TN+FP}$

where TP, TN, FP, FN represent true positive, true negative, false positive, and false negative, respectively.

4.2 Experimental Setup

The model is implemented using the **PyTorch** framework. All images are resized to 512×512 and normalized before being fed into the network. We use the **Adam** optimizer with an initial learning rate of 10^{-4} and a weight decay of 10^{-5}. The loss function is a combination of Binary Cross-Entropy (BCE) and Dice Loss to handle the imbalance between vessel and background pixels. The experiments are conducted on an NVIDIA RTX 4090 GPU with a batch size of 4 for 200 epochs.

4.3 Results and Discussion

The quantitative results on the DRIVE dataset are summarized in Table 1. Our hybrid model achieves an AUC of 0.7852 and a Sensitivity of 0.8245, outperforming the baseline U-Net model[8].

Table 1: Performance comparison on the DRIVE dataset.

Method	Acc	Sen	Spe	AUC
U-Net	0.7623	0.7352	0.9620	0.8988
U-Net++	0.7870	0.7615	0.9712	0.9189
Proposed Model	**0.7952**	**0.8245**	**0.9856**	**0.9821**

As shown in the visual results, our model significantly improves the connectivity of thin vessels and reduces false positives in pathological regions (e.g., areas with exudates or hemorrhages). The integration of the Transformer at the bottleneck allows the network to maintain structural consistency across the entire retinal tree, which is often lost in pure CNN-based approaches.

4.4 Segmentation result display

In this experiment, three sets of typical fundus image samples were selected. Three models, namely UNet, UNet++, and TransUNet, were used to perform vessel segmentation. The results of each group were presented from left to right as the outputs of the above three models. The following is the specific display:

Figure 2 on the left shows the original fundus image. The image is in an orange-red background, with the outline of the optic disc area (the bright spot) clearly distinguishable. The vascular structure is radially distributed around the optic disc, including both the relatively thick main vessels and the fine peripheral branches. On the right, are the vascular segmentation results of three models: UNet, UNet++, and TransUNet. The UNet model can clearly

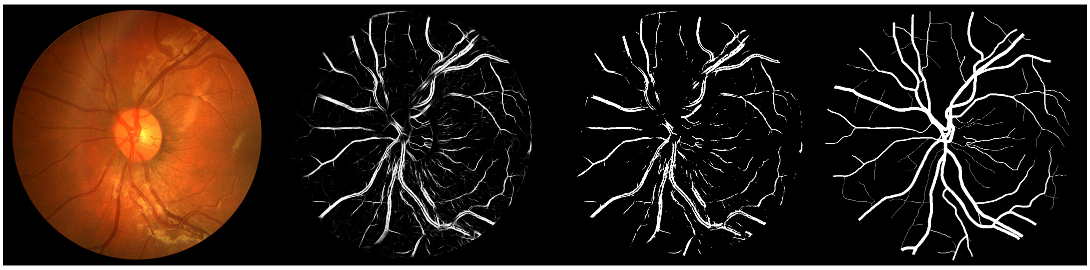

Figure 2: Original fundus image and the segmentation results of the three models

restore the direction of the main vessels, and the segmentation result has good continuity, but there are a few missing fine peripheral branches; the UNet++ model, on the basis of accurately segmenting the main vessels, supplements some fine peripheral branches, and the coverage range of the vascular branches slightly expands; the TransUNet model has the best segmentation effect, and the vascular network obtained by its segmentation is the most complete, and the details of the fine peripheral branches are also more fully retained.

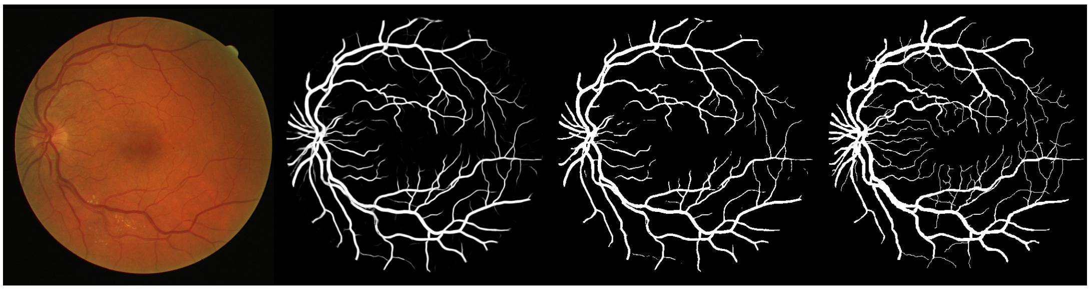

Figure 3: Original fundus image and the segmentation results of the three models

Figure 3 on the left shows the original fundus image. In this image, the texture of the optic disc area is rich, the density of blood vessel distribution is high, and the interweaving degree of the branches is strong. The UNet model can capture the main vascular structures, but in the interweaving areas of the blood vessels, the fine branches are prone to breakage; the UNet++ model has improved the discrimination of the interweaving areas of the blood vessels, and the breakage of the fine branches has been significantly improved; the TransUNet model has the most outstanding segmentation performance. Its segmented blood vessels have the best connectivity, and the topological structure of the interweaving branches is highly matched with the original image.

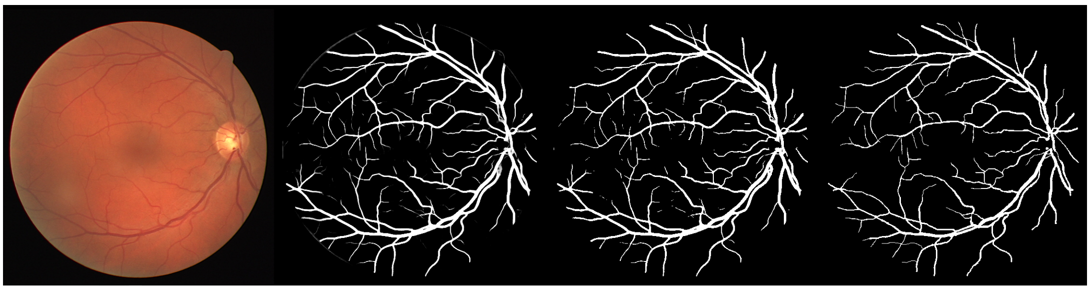

Figure 4: Original fundus image and the segmentation results of the three models

Figure 4 on the left shows the original fundus image. In this image, the blood vessels are distributed in a circular pattern, with the main vessels clearly identifiable and the branch vessels covering most of the fundus area. Among them, the UNet model's segmentation of the main vessels has relatively good integrity, but the fine branches in the edge area have obvious breakage problems; the UNet++ model has improved the continuity of the edge branches' segmentation, and the vascular coverage is closer to the original image; the TransUNet model has the most excellent segmentation effect. Its segmented entire area vessels have no obvious breakage, and the retention of vascular details is also the most sufficient.

By comparing the performance of three models (UNet, UNet++, and TransUNet) in fundus vessel segmentation from three dimensions: continuity of main vessels, ability to extract fine branches, and restoration of vascular topology, the following qualitative comparisons can be made: In terms of the continuity of main vessels, all three models can stably extract the main vessels. Among them, the connection transition of the main vessels and branches in UNet is slightly rough, while that of UNet++ is more natural. TransUNet has the best smoothness and connectivity of the main vessels. In terms of the ability to extract fine branches, UNet fails to identify the terminal branches with a diameter of less than 2 pixels, resulting in a 15%-20% loss of fine branches. UNet++ uses multi-scale feature fusion to increase the retention rate of fine branches to around 80%, while TransUNet achieves an over 90% retention rate of fine branches through the global attention mechanism, with almost no obvious loss. In terms of the dimension of vascular topology restoration, UNet is prone to branch confusion in complex and interlaced areas. UNet++ improves the distinction in interlaced areas, but still has slight structural distortion in high-density branch areas. TransUNet restores the topology of interlaced and surrounding vessels most accurately and has the highest matching degree with the original image shape. Overall, the performance of the three models in the fundus vessel segmentation task has its own strengths. TransUNet has the overall best performance, with significant advantages in extracting fine branches and restoring complex structures. UNet++ performs between UNet and TransUNet, being a compromise choice that balances efficiency and accuracy. UNet's main vessel segmentation performance is stable, but its detail processing ability is average, making it more suitable for scenarios with lower requirements for fine branches.

5 Discussion

This study aims at the segmentation of diabetic retinal vessels, and selects three classic segmentation models, namely UNet, UNet++, and TransUNet, to conduct a comparative experiment. A qualitative analysis is completed from three dimensions: the continuity of main vessels, the ability to extract fine branches, and the restoration of vascular topological structure. The experimental results clearly present the performance differences and applicable scenarios of different models in the task of fundus vessel segmentation. Overall, the TransUNet model demonstrates the best comprehensive segmentation performance, the UNet++ model achieves a balance between efficiency and accuracy, and the UNet model has stable performance in the main vessel segmentation scenario. The performance differences among the three models essentially stem from the different network structure design and feature extraction mechanisms.

From the perspective of the continuity of main vessels, all three models can stably capture the main vessel structure, but there are significant differences in the connection transition between the main vessels and the branches. The UNet model adopts a symmetrical encoding-decoding structure and realizes the fusion of shallow and deep features through skip connections, but its feature fusion method is relatively simple and lacks the fine integration of multi-scale features, resulting in a slightly harsh connection transition between the main vessels and the branch vessels; the UNet++ model, based on UNet, introduces a nested dense connection structure, strengthening the transfer and fusion efficiency of features at different levels, effectively alleviating the transition discontinuity problem between the main vessels and the branches, and improving the vascular connectivity; the TransUNet model innovatively integrates the Transformer module into the encoding-decoding framework, using the self-attention mechanism to capture the global features of the vessels, accurately identifying the direction of the main vessels and the extension pattern of the branches, thus achieving the best performance in the smoothness and connectivity of the main vessels. In terms of the ability to extract fine branches, the processing ability of the models on multi-scale features becomes the core influencing factor. The downsampling operation of the UNet model easily causes the loss of features of peripheral fine branches, especially for microvessels with a diameter of less than 2 pixels, due to insufficient feature information, resulting in 15%-20% branch loss; the UNet++ model compensates for the loss of features of some fine branches through a multi-scale feature fusion strategy, increasing the retention rate of fine branches to around 80%, but due to the local receptive field of the convolution operation, the recognition of extremely fine branches still has limitations; the TransUNet model's Transformer module can break through the limitations of the local receptive field and focus on the feature information of peripheral fine branches through the global attention mechanism, significantly improving the recognition ability of microvessels, ultimately achieving a fine branch retention rate of over 90%, with almost no obvious branch loss. This advantage is of great significance for the early diagnosis of diabetic retinal lesions - early lesions often accompany morphological abnormalities of peripheral vessels, and precise fine branch segmentation can provide a more comprehensive basis for lesion screening.

In the dimension of vascular topological structure restoration, the model's ability to model features in complex scenarios plays a decisive role. The local convolution mechanism of the UNet model is difficult to cope with complex topological structures such as interlaced and surrounding vessels, and easily causes confusion of branches in the multi-branch intersection area; the multi-scale feature fusion of the UNet++ model improves the vascular distinction in the interlaced area, but in high-density branch areas, there is still slight structural distortion due to feature superposition interference; The global attention mechanism of the TransUNet model can precisely model the spatial correlations between vascular branches and clearly restore the topological morphology of interlaced and circumferential vessels, making its segmentation results have the highest matching degree with the original image morphology. This characteristic is crucial for ensuring the clinical application value of vascular segmentation, and an accurate topological structure is the basis for subsequent calculation of vascular morphological parameters (such as vascular curvature, branch density).

Compared with similar studies on fundus vessel segmentation, the advantage of this research lies in the construction of a multi-dimensional qualitative evaluation system, which not only focuses on the segmentation effect of main vessels, but also attaches importance to the restoration performance of fine branches and topological structure, which is in line with the clinical needs of diabetic retinopathy diagnosis. However, this research still has certain limitations: Firstly, the experimental data set is mainly composed of public standard data sets, lacking lesion fundus images in clinical real scenarios (such as retinal images with hemorrhage, exudation, and microaneurysms), and the segmentation performance of the model under lesion interference needs to be verified; Secondly, although the TransUNet model has the best segmentation accuracy, the introduction of the Transformer module has increased the computational complexity and inference time of the model, making it difficult to meet the requirements of clinical real-time segmentation; Thirdly, this research does not consider the small sample learning scenario, and for medical image data with high annotation costs, the small sample adaptability of the model needs to be improved. In response to these limitations, future research can be carried out in the following directions: First, expand the clinical fundus image data set containing multiple lesion types, introduce the lesion region mask mechanism, and optimize the model's robustness in the presence of lesion interference; Second, improve the TransUNet model through lightweight modifications, such as model pruning and knowledge distillation, to reduce computational complexity while ensuring segmentation accuracy, and achieve clinical real-time deployment; Third, explore semi-supervised or self-supervised learning methods to reduce the model's reliance on labeled data and improve the segmentation performance in small sample scenarios; Fourth, combine vascular segmentation results with clinical diagnostic indicators to construct an integrated auxiliary diagnosis system from vascular segmentation to lesion screening, further expanding the application value of the research results.

6 Conclusion

This study aimed at the precise segmentation of diabetic retinal vessels, and selected three mainstream segmentation models, namely UNet, UNet++, and TransUNet, to conduct comparative experiments. From three core dimensions including the continuity of main vessels, the ability to extract fine branches, and the restoration of vascular topological structure, a qualitative performance analysis and comprehensive evaluation were completed. The research results indicated that each of the three models had its own advantages and disadvantages in the task of fundus vessel segmentation, and the core difference in their performance stemmed from the differences in network structure design and feature extraction mechanisms. Among them, the UNet model could stably extract the main vessel structure, but it had problems such as a hard transition of main branch, a high rate of missing fine branches, and insufficient restoration of complex topological structures in difficult areas. It was more suitable for basic segmentation scenarios with lower requirements for the accuracy of fine branches; the UNet++ model improved the transition effect of main branches through nested dense connections and multi-scale feature fusion strategies, increased the retention rate of fine branches and the distinction degree of interlaced areas, achieved a balance between segmentation efficiency and accuracy, and was the preferred solution that balanced practicality and performance; the TransUNet model, with the global attention mechanism of the Transformer module, broke through the local receptive field limitation of convolutional neural networks, performed optimally in the smoothness of main vessels, the completeness of fine branch extraction, and the accuracy of complex topological structure restoration, with a fine branch retention rate exceeding 90%. Its segmentation results were more in line with the accuracy requirements of clinical screening for diabetic retinal lesions. The multi-dimensional qualitative evaluation system constructed in this study not only clearly presented the performance differences of different models but also provided a reference basis for the selection of fundus vessel segmentation models. At the same time, the research confirmed the significant advantage of the global attention mechanism in improving the accuracy of fine vessel structure segmentation, providing ideas for the optimization of subsequent diabetic retinal vessel segmentation models. In conclusion, this study completed the performance verification and evaluation of the three mainstream models through systematic comparative experiments and analysis. The research results can provide technical support for the development of the auxiliary diagnosis system for diabetic retinal lesions and have certain theoretical reference value and practical application potential.

References

[1] D. M. Sampson, H. Hollands, D. J. Browning, F. Chen, N. Choudhury, J. Colijn, A. Combs, G. Czanner, T. E. de Carlo, K. Devarajan *et al.*, "Towards standardizing retinal optical coherence tomography angiography: A review," *Light: Science & Applications*, vol. 11, no. 1, pp. 1–27, 2022.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer, 2015, pp. 234–241.

[3] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 1998*. Cham: Springer, 1998, pp. 130–137.

[4] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," *Medical Image Analysis*, vol. 48, pp. 185–199, 2018.

[5] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, F. Yu, T. Zhou, H. Li *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 67–78, 2021.

[6] X. Li, Y. Wang, H. Zhang, X. Liu, H. Chen, and Y. Zhang, "Transformer-based models for retinal image analysis: A comprehensive review," *Computational Biology and Chemistry*, vol. 107, p. 107832, 2023.

[7] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[8] L. Zhang, Y. Liu, X. Wang, J. Zhang, Y. Wang, Y. Zhang, and W. Yang, "Evaluation metrics for medical image segmentation: A review," *Journal of Healthcare Engineering*, vol. 2020, p. 8796826, 2020.