Proceedings of CONF-SEML 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA31944

AF-UNet: Asymmetric Fusion U-shaped Convolutional Networks for Retinal Vessel Segmentation

Qizhe Sun

College of Computer Science, Beijing University of Technology, Beijing, China
sqz_bjut@emails.bjut.edu.cn

Abstract. Accurate segmentation of retinal vessels is very important for computer-aided diagnosis and screening of diabetic retinopathy, hypertension, glaucoma and other eye diseases. The U-Net structure based on Convolutional Neural Networks (CNNs) is excellent in the feature extraction and segmentation of retinal blood vessels. However, due to the complex structure of blood vessels in retinal images, low contrast and the interference of diseased areas, it is still difficult to accurately segment small blood vessels and marginal blood vessels by existing methods. Therefore, it is still a big challenge to realize automatic segmentation with high precision and accuracy. In this paper, we propose a new deep learning model called Asymmetric Fusion UNet (AF-UNET), which is improved on the basis of U-Net framework and is specially used to segment retinal vessels. In this model, asymmetric jump connection is added between encoder and decoder, and the shallow features of encoder are integrated into the full-scale features of decoder. The advantage of this design is that it can carry out multi-scale feature fusion, focusing on tiny vascular structures, and avoiding the problem of too dense connections, making the whole network lighter. We have done experiments on public retinal image data sets, and the results show that AF-UNet performs well on F1-score, recall and Intersection over Union(IoU), which is stronger than those classical methods based on U-Net and attention mechanism. In addition, the visualization results also prove that AF-UNet has obvious advantages in maintaining vascular connectivity and accurately segmenting small vascular structures.

Keywords: Retinal Vessel Segmentation, Deep Learning, U-Net , Multi-scale feature fusion

1. Introduction

Retinal vascular segmentation is particularly important in computer-aided diagnosis and screening of some eye diseases, such as diabetic retinopathy, hypertension and glaucoma. The shape characteristics of retinal blood vessels-such as the width, bending degree and bifurcation of blood vessels-are the key signals used by doctors to judge whether the disease has worsened. It takes a lot of time for ophthalmologists to draw these blood vessels by hand, and everyone's judgment may be different, and there is no way to use them for many people; Therefore, we especially need a method that can automatically and accurately segment blood vessels.

In the past decade, the method based on Convolutional Neural Network (CNN), especially U-Net [1] and its variants, has become the mainstream of medical image segmentation. U-Net's encoder-

DOI: 10.54254/2755-2721/2026.BA31944

decoder structure with jumping connection can combine the features of low-level and high-level so that the general outline can be seen clearly and the details can be preserved. Many works have adopted or improved U-Net to segment retinal vessels, and achieved great success [2]. In recent research, attention mechanism, multi-scale fusion and residual module have been added, which can better identify tiny blood vessels and reduce background interference [3]. However, challenges still exist, especially in areas with low contrast or lesions, and it is still difficult to accurately segment fine blood vessels and blood vessel boundaries.

These difficulties are due to several problems mixed together. First of all, the contrast between blood vessels and background in retinal images is often very low, especially those tiny blood vessels and marginal parts. Second, the structure of blood vessels is super complex, with bifurcation and crossing, and the thickness is different; It is difficult for existing models to keep their connection and coherence. Thirdly, skip connection in models such as U-Net may make the features of encoder and decoder mismatch, resulting in the result being too smooth or intermittent. Some improved methods can really segment blood vessels better, but the price is that the model becomes particularly complicated, which may make it difficult for us to use them on equipment with limited resources [4].

Therefore, we urgently need a method that can not only ensure the accuracy of segmentation (especially for those thin blood vessels), but also keep the model structure efficient. In order to solve this problem, in this paper, we propose AF-UNet, which is an improved U-Net structure, and is specially used to segment retinal blood vessels. AF-UNet introduces asymmetric hop connection between encoder and decoder, which can selectively integrate some features of the shallow layer of encoder into all scales of decoder. This method allows us to better integrate features of different scales, paying special attention to those subtle structures, and at the same time avoiding the connection mode becoming too dense or redundant. The resulting network is still relatively lightweight and efficient.

To put it simply, this paper proposes an improved U-Net segmentation framework, which adds asymmetric skip connections, which can better integrate the shallow encoder features into the full-scale decoder feature map, so that the model can capture the detailed structure more accurately. At the same time, our design maintains a lightweight architecture and does not increase unnecessary complexity, but the performance is obviously improved. A lot of experiments have been done on the public benchmark data set, and the results show that this method is really effective and performs better than the existing methods in many quantitative evaluation indexes.

2. Literature review

Retinal vascular segmentation has been studied for decades because it is very important for early detection and monitoring of eye and systemic diseases. Traditional methods mainly rely on hand-designed features and image processing technologies, such as matched filtering, morphological operations and edge detection. The early methods of Hoover et al. [5] and Staal et al. [6] used line detectors and ridge-based features to identify vascular structures. These methods are ok under ideal conditions, but the effect becomes worse when the illumination changes, the contrast is different or there is pathological noise.

With the progress of deep learning, Convolutional Neural Networks (CNNs) have basically replaced the traditional feature engineering. The appearance of U-Net [1] marks a major breakthrough in biomedical image segmentation. Its encoder-decoder structure and jumping connection can fuse multi-scale features and locate accurately, which is especially suitable for retinal blood vessel segmentation. Later, many improved versions appeared to solve the shortcomings of

DOI: 10.54254/2755-2721/2026.BA31944

the original model. For example, Residual U-Net [7] added residual learning to alleviate the problem of gradient disappearance, while Attention U-Net [8] used attention mechanism to highlight the key areas of blood vessels. These improvements improve the segmentation accuracy, especially in areas with complex vascular structures.

However, those structures based on U-Net still encounter difficulties in dealing with small blood vessel segmentation and low contrast areas. In order to make the model more sensitive to small blood vessels, some researchers have proposed a multi-scale feature fusion method. Mamba-UNet [9] uses the state space model to capture long-distance context information, and its computational complexity is linear, which makes it perform better in the task of retinal blood vessel segmentation. Similarly, multi-scale feature aggregation UNET (MFA-UNET) [3] adopts hierarchical feature fusion to maintain the continuity of blood vessels. Although the accuracy of these models has improved, they also make the network more complicated and the calculation cost greatly increased, which limits their practical application in lightweight systems.

Recently, more and more studies have explored the use of attention mechanism to adjust features, so that retinal blood vessels can be segmented more accurately. Soni et al. [10] proposed a multi-scale feature attention framework to make blood vessels more conspicuous at different resolutions; Hernandez-Gutierrez et al [11] added reverse attention to the lightweight U-Net, which effectively suppressed the background interference and made the tiny blood vessels connect more clearly. A model based on Transformer, such as TransUNet [12] also adds global self-attention, which can capture long-distance associations and perform well on large data sets. However, they occupy too much memory and are not easy to understand, so it is still a little difficult to use them in practical clinic.

Generally speaking, the research on retinal vascular segmentation now mainly follows three directions: enhancing the integration of local and global features through complex skip-connection design, adding attention or transformer-based mechanism to improve the ability of context modeling, and developing lightweight architecture to make it available in real time or with limited resources.

However, if we look closely, we will find that these methods have some shortcomings themselves. The existing multi-scale fusion strategies often use symmetric and dense hop connections, which will spread the shallow features to all decoder stages indiscriminately. Although this can better preserve the spatial details, it also brings redundant low-level information, and greatly increases the amount of calculation, resulting in less and less performance improvement.

The models based on attention and transformer can capture the global relationship better, but they usually need more parameters and memory, so they are not suitable for processing high-resolution retinal images or for clinical environment with limited resources.

At the same time, it is still difficult to design a lightweight segmentation model, because a large reduction in parameters often makes the expression ability of the model insufficient, especially when dealing with thin and low contrast blood vessels. Therefore, it is still an unsolved problem that how to accurately capture the subtle structure while ensuring the computational efficiency.

In this case, we propose AF-UNet to solve this problem, which uses asymmetric hop connection. This connection can selectively add the characteristics of shallow encoder to the multi-scale decoder expression. This design strengthens the recognition ability of small blood vessels, while avoiding repeated propagation characteristics, thus achieving a better balance between segmentation accuracy and network efficiency.

DOI: 10.54254/2755-2721/2026.BA31944

3. Methodology

3.1. Datasets and data preprocessing

We have done experiments on the published retinal vascular segmentation data sets, including DRIVE and STARE, which are commonly used standard data sets in this research field. These data sets include fundus photos taken under different imaging conditions, as well as vascular markers manually marked by ophthalmologists. The difficulties of these data sets are uneven light, low contrast and complex blood vessel shape, so they are very suitable for testing the robustness of the model. Before training, we adjusted all retinal images to the same size and normalized them to reduce the brightness difference. We also use data enhancement techniques, such as random rotation, horizontal and vertical flipping and contrast enhancement, which can improve the generalization ability and prevent over-fitting. The data set is divided into training set, verification set and test set according to the standard method, so that it can be compared with the existing methods fairly.

3.2. Network architecture

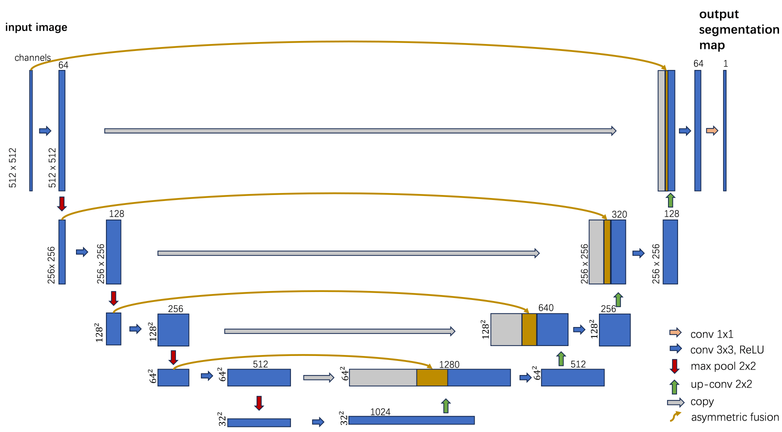

Figure 1. AF-UNet architecture (example for 512×512input resolution)

Each blue box corresponds to a multi-channel feature map. The number of channels is written at the top of each box, and the spatial resolution is displayed in the lower left corner. The gray box represents the feature map copied from the encoder by symmetrical jump connection. The yellow box represents the feature map copied from the encoder through asymmetric jump connection. Arrows represent different operations, including convolution, downsampling, upsampling and feature fusion.

The overall structure of our proposed AF-UNet can be seen in Figure 1. This AF-UNet is based on the classic U-Net framework and uses a four-layer encoder-decoder structure. The encoder part has four convolution stages, each stage contains two consecutive 3x3 convolution layers, followed by batch normalization and ReLU activation. At different stages of the encoder, the number of characteristic channels will gradually increase from 64 to 512. After each encoder block, we will use 2x2 max-pooling with a step size of 2 for spatial downsampling.

A bottleneck module with two 3×3 convolutions extends the feature dimension to 1024 channels to capture advanced semantic information.

Decoder's structure and encoder are symmetrical, and the spatial resolution is restored by transposing the convolution layer (2×2, step 2). Different from the traditional symmetric skip

DOI: 10.54254/2755-2721/2026.BA31944

connections, AF-UNet adopts an asymmetric skip connection strategy. At each layer of decoder, the up-sampled feature map will not only be fused with the encoder feature with the same resolution, but also with a shallow encoder feature at a higher resolution stage.

Specifically, we define shallow encoder features as the information that comes out of the first few layers of the encoder (such as the first to third blocks) before the picture is reduced too small. This information retains a wealth of spatial details. We will selectively integrate them into deeper decoder through a method called channel-wise concatenation. This way of full-scale but non-dense fusion allows decoder to see a variety of fine spatial clues at the same time, without making the information too repetitive.

Each decoder block uses two 3×3 convolutions to optimize the aggregated features after feature fusion. Finally, a 1×1 convolution layer will turn the decoder's output into a single-channel probability map, which can be used to segment blood vessels. This design can not only improve the continuity and boundary preservation effect of small blood vessels, but also keep the network structure compact and the calculation cost controllable.

3.3. Training strategy and loss function

We use paired retinal images and ground truth masks to supervise the training of this network. In order to solve the problem of class imbalance between blood vessels and background pixels, we use a composite loss function combined with Dice loss and binary cross-entropy loss. This loss function can not only ensure the overlap at the regional level, but also improve the accuracy at the pixel level. The model parameters are optimized by Adam optimizer, and the learning rate is fixed. In order to prevent over-fitting, we adopt the early stop strategy according to the validation loss. All experiments are implemented in PyTorch framework and run in an environment supporting GPU.

3.4. Evaluation metrics

We use several common standards to measure the segmentation effect, including precision (precision), recall (recall), F1-score(F1 score) and Intersection over Union(IoU). These criteria can help us to judge whether the segmentation is accurate, whether we can find tiny blood vessels, and how the results overlap with the standard answers.

4. Results

4.1. Quantitative evaluation

Table 1. The results of tested models on DRIVE dataset

Method	Jaccard	F1	Recall	Precision	Acc
Unet	0.6765	0.8071	0.7877	0.8276	0.9667
Unet++	0.6740	0.8053	0.7600	0.8417	0.9681
ResUNet	0.6719	0.8038	0.8040	0.8079	0.9659
ResUNet++	0.6658	0.7994	0.7525	0.8578	0.9672
AttentionUNet	0.6507	0.7884	0.7883	0.7912	0.9647
AF_UNet	0.7095	0.8301	0.7991	0.8638	0.9691

As can be seen from Table 1, AF-UNet achieved the highest Jaccard index(0.7095) and F1-score(0.8301) among all the methods compared. Compared with the standard U-Net, AF-UNet

DOI: 10.54254/2755-2721/2026.BA31944

improves the Jaccard index by about 3.3 percentage points (from 0.6765 to 0.7095), and F1-score also increases by 2.3 percentage points. This shows that the overlap between the predicted vascular region and the real marker is much better.

On the recall index, AF-UNet got 0.7991, which is similar to ResUNet(0.8040), but higher than U-Net, UNet++, ResUNet+++and Attention U-Net. This shows that AF-UNet has a strong recognition ability for blood vessel pixels, especially those thin blood vessels with low contrast and easy to be missed. At the same time, AF-UNet also achieved the highest precision(0.8638), which obviously exceeded all other baseline models. Recall and precision are improved at the same time, which shows that AF-UNet does not increase too many wrong judgments while reducing missing real blood vessels.

The overall accuracy of AF-UNet is 0.9691, which is slightly higher than other methods. Although the accuracy is usually not sensitive to the performance of blood vessel segmentation due to the imbalance of categories, this result still reflects the improvement of the overall pixel-level classification consistency. On the whole, the results on DRIVE data set show that AF-UNet provides a more balanced and robust segmentation performance, which effectively improves the ability of blood vessel detection and background discrimination.

Table 2. The results of tested models on STARE dataset

Method	Jaccard	F1	Recall	Precision	Acc
Unet	0.6557	0.7921	0.7713	0.8142	0.9642
Unet++	0.6771	0.8075	0.7823	0.8345	0.9781
ResUNet	0.6871	0.8146	0.7993	0.8343	0.9682
ResUNet++	0.6982	0.8223	0.8012	0.8446	0.9772
AttentionUNet	0.6598	0.7951	0.7912	0.7992	0.9787
AF_UNet	0.7142	0.8333	0.8134	0.8544	0.9791

As shown in Table 2, AF-UNet once again achieved the best results in all evaluation criteria on STARE data set. Specifically, AF-UNet got the highest Jaccard index (0.7142) and F1-score (0.8333), which was 1.6 percentage points higher than the Jaccard Index of ResUNet+++and the F1 score was 1.1 percentage points higher. This steady progress shows that the predicted vascular map is in good agreement with the real mark in structure.

AF-UNet also got the highest recall value (0.8134) among all the compared models, which indicated that AF-UNET was more sensitive to the vascular structure in the more challenging STARE data set, because the width and pathological features of blood vessels in this data set changed greatly. At the same time, its precision reaches 0.8544, which is higher than U-Net, U-Net, UNet++, ResUNet and Attention U-Net. This shows that AF-UNet can effectively suppress background noise while maintaining strong blood vessel detection ability.

In terms of accuracy, AF-UNet reaches a value of 0.9791, which is similar to the best methods and slightly higher than most baseline methods. There is a steady improvement in Jaccard, F1-score, recall and precision, which shows that AF-UNet does not rely on the advantages of a specific index, but can improve the overall performance under different evaluation criteria.

On the two data sets of DRIVE and STARE, AF-UNet has always performed better than the traditional U-Net and its upgraded versions (such as UNet++, Resume, ResUNet+++and Attention U-Net). The most obvious improvements are those based on overlapping indexes (Jaccard and F1-score) and recall, which shows that AF-UNet can better preserve the structure of small blood vessels and keep the blood vessels coherent. Importantly, these improvements do not reduce the precision, which proves that our proposed asymmetric skip connection strategy can integrate multi-scale

DOI: 10.54254/2755-2721/2026.BA31944

features more effectively and suppress background interference reliably. The quantitative expression of this balance shows that AF-UNet is stable and adaptable in different tasks of retinal vascular segmentation.

4.2. Qualitative analysis

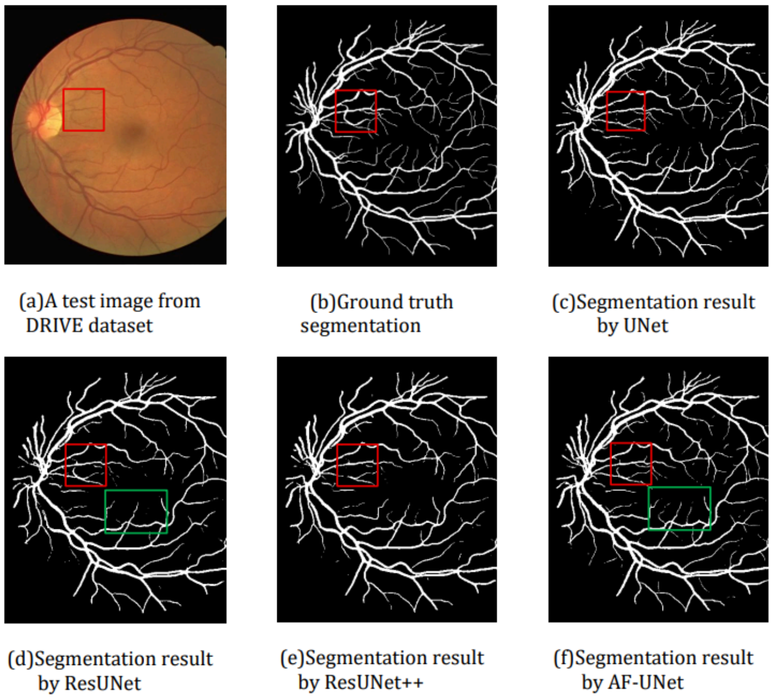

Figure 2. Qualitative segmentation results of different models on DRIVE and STARE datasets

The visualization result (Figure 2) compares the original retinal image, the real annotation, and the segmentation results generated by U-Net, Resume, ResUNet+++and our proposed AF-UNet. By observing the blood vessels marked in the red box, we can find that the segmentation result obtained by AF-UNet is the closest to the real situation under the complex blood vessel structure. In these areas, AF-UNet can maintain the continuity and branching structure of blood vessels more accurately, while U-Net and ResUNet+++will be broken or not completely segmented.

Further comparing the blood vessels marked by the green border box, it can be found that AF-UNet performs better in segmenting small blood vessels. Compared with ResUNet, AF-UNet can extract small blood vessels more completely and continuously, reducing the common problems of breakage and missed detection in baseline model.

These good results are due to the great design of AF-UNet. It uses a method called asymmetric skip connection, which can smoothly merge the detailed information in the previous encoder into the various parts of the subsequent decoder. The advantage of this is that it can combine features of different sizes well, making it easier for decoder to restore those particularly small and unclear vascular structures, especially in those places with complex shapes and light colors. Therefore, AF-UNet has a more accurate segmentation and a more complete shape, which is consistent with the numerical results we saw earlier. Generally speaking, the effects we see are matched with the measured data, which shows that AF-UNet is particularly good at preserving the shape of small blood vessels and their connections.

DOI: 10.54254/2755-2721/2026.BA31944

5. Discussion

The experimental results show that AF-UNet can better integrate the features of different scales through a method called asymmetric skip connections, and successfully solve some major problems in retinal blood vessel segmentation, and it is not as troublesome as those particularly complicated dense connection structures. Compared with the traditional U-Net variant, AF-UNet is more sensitive to tiny blood vessels and marginal areas, which is due to its selective addition of shallow encoder features, which enrich the spatial details without interfering with the main semantic information. In addition, unlike those methods based on attention or transformer, AF-UNet achieves these performance improvements with only a little additional computation, which makes it more practical in clinical environments with limited resources.

However, we have to know that this experiment uses public data sets like DRIVE and STARE, and these pictures are usually carefully arranged and marked under controlled shooting conditions. However, in a real hospital, the light, shooting equipment, noise level and the patient's condition of fundus photos will vary greatly, which may lead to a gap between the public data set and the actual clinical data. Therefore, the performance of AF-UNet on larger and more complex real clinical data needs further verification.

Moreover, although AF-UNet improves the continuity of blood vessels and the detection effect of small blood vessels, the performance of blood vessel segmentation in diseased areas is still not good enough. Lesions such as hemorrhages and exudates usually have obvious brightness changes and irregular textures, which may interfere with the identification of blood vessels, leading to misjudgment or intermittent blood vessels. There are too few samples of these lesions in the standard public data set, which limits the ability of the model to learn the changes related to complex lesions during training.

Future work will try to narrow the gap between public testing and real medical scenarios by adding more different clinical data and domain adaptive training strategies. In addition, adding a lightweight attention mechanism to suppress pathological artifacts may make the identification of lesion areas more stable, thus improving the clinical practicability of our method.

6. Conclusion

In this paper, we put forward a model called AF-UNet, which is a portable and powerful tool, specially used to segment the blood vessels in the retina. By adding asymmetric skip connections to the structure of U-Net, AF-UNet can better integrate features of different scales, which is especially helpful for finding those thin and complicated blood vessels. We have done a lot of experiments on several public data sets, and the results show that AF-UNet performs better than many top methods in F1-score, recall and IoU, and its network structure is very compact. Qualitative results further prove that it can well maintain the continuity of blood vessels and those tiny structural details. These findings show that asymmetric feature fusion is a promising direction, which can improve the effect of medical image segmentation without sacrificing efficiency. Our future research will apply this method to other medical imaging tasks and explore how to combine it with adaptive attention mechanism.

References

[1] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]//International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international publishing, 2015: 234-241.

DOI: 10.54254/2755-2721/2026.BA31944

[2] Liu Z, Sunar M S, Tan T S, et al. Deep learning for retinal vessel segmentation: a systematic review of techniques and applications [J]. Medical & Biological Engineering & Computing, 2025: 1-18.
[3] Cao J, Chen J, Gu Y, et al. MFA-UNet: A vessel segmentation method based on multi-scale feature fusion and attention module [J]. Frontiers in Neuroscience, 2023, 17: 1249331.
[4] Zhang X, Wang J, Cheng X, et al. A Review of Deep Learning-Based Medical Image Segmentation [J]. IET Image Processing, 2025, 19(1): e70163.
[5] Hoover A D, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response [J]. IEEE Transactions on Medical imaging, 2000, 19(3): 203-210.
[6] Staal J, Abràmoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina [J]. IEEE transactions on medical imaging, 2004, 23(4): 501-509.
[7] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.
[8] Das N, Das S. Attention-UNet architectures with pretrained backbones for multi-class cardiac MR image segmentation [J]. Current Problems in Cardiology, 2024, 49(1): 102129.
[9] Hu H, Xiong H, Long F, et al. Multi-scale vision Mamba-UNet: A Mamba-based method for retinal vessel segmentation [J]. Biomedical Signal Processing and Control, 2026, 112: 108435.
[10] Soni T, Gupta S, Bharany S, et al. Retinal vessel segmentation using multi scale feature attention with MobileNetV2 encoder [J]. Scientific Reports, 2025, 15(1): 43369.
[11] Hernandez-Gutierrez F D, Avina-Bravo E G, Ibarra-Manzano M A, et al. Retinal Vessel Segmentation Based on a Lightweight U-Net and Reverse Attention [J]. Mathematics, 2025, 13(13): 2203.
[12] Chen J, Mei J, Li X, et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers [J]. Medical Image Analysis, 2024, 97: 103280.