Ni *et al*. *Intell. Robot*. **2025**, *5*(3), 679-94
DOI: 10.20517/ir.2025.35

Intelligence & Robotics

Research Article

Open Access

Check for updates

An improved U-Net model with multiscale fusion for retinal vessel segmentation

Jianjun Ni[1], Jiamei Shi[1], Qiao Zhan[2], Ziru Zhang[1], Yang Gu[1]

[1]College of Artificial Intelligence and Automation, Hohai University, Changzhou 213200, Jiangsu, China.
[2]Department of Infectious Diseases, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, Jiangsu, China.

Correspondence to: Prof. Jianjun Ni, College of Artificial Intelligence and Automation, Hohai University, Changzhou 213200, Jiangsu, China. E-mail: njjhhuc@gmail.com; ORCID: 0000-0002-7130-8331

How to cite this article: Ni, J.; Shi, J.; Zhan, Q.; Zhang, Z.; Gu, Y. An improved U-Net model with multiscale fusion for retinal vessel segmentation. *Intell. Robot.* **2025**, *5*(3), 679-94. http://dx.doi.org/10.20517/ir.2025.35

Received: 12 May 2025 **First Decision:** 27 Jun 2025 **Revised:** 9 Jul 2025 **Accepted:** 17 Jul 2025 **Published:** 21 Aug 2025

Academic Editor: Muhammad Adnan Khan **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

The condition of the retinal vessels is involved in various ocular diseases, such as diabetes, cardiovascular and cerebrovascular diseases. Accurate and early diagnosis of eye diseases is important to human health. Recently, deep learning has been widely used in retinal vessel segmentation. However, problems such as complex vessel structures, low contrast, and blurred boundaries affect the accuracy of segmentation. To address these problems, this paper proposes an improved model based on U-Net. In the proposed model, pyramid pooling structure is introduced to help the network capture the contextual information of the images at different levels, thus enhancing the receptive field. In the decoder, a dual attention block module is designed to improve the perception and selection of fine vessel features while reducing the interference of redundant information. In addition, an optimization method for morphological processing in image pre-processing is proposed, which can enhance segmentation details while removing some background noise. Experiments are conducted on three recognized datasets, namely DRIVE, CHASE-DB1 and STARE. The results show that our model has excellent performance in retinal vessel segmentation tasks.

Keywords: Retinal vessel segmentation, deep learning, U-Net network, contextual information, attention mechanism

1. INTRODUCTION

The retina is a unique region for direct visualization and imaging. It can reflect the signs of many systemic diseases. By examining the shape, curvature and branching patterns of retinal vessels, clinicians can provide

crucial diagnostic services for the early identification and ongoing care of eye diseases[1,2]. However, it takes a lot of work to manually identify blood vessels in fundus images and cumbersome task relying on the experience of ophthalmologists, and there is a lack of specialists who can perform this task independently[3]. Therefore, the development of the precise and effective segmentation method will be beneficial in the diagnosis of eye diseases[4,5].

Traditional retinal vessel segmentation methods utilize the morphological features of the vessels, and mainly rely on morphological operations[6]. Early vessel segmentation methods include feature pixel-based methods, matched filter-based methods, and vessel tracking-based methods[7]. However, these traditional methods rely on domain experience, which are sensitive to image quality and noise, leading to poor segmentation results. So their performances are limited.

In recent years, retinal vessel segmentation technology based on deep learning has been more widely studied and applied[7]. In retinal vessel segmentation, deep-learning networks have outperformed conventional methods[8]. For example, convolutional neural networks (CNNs) have achieved excellent performance in numerous visual tasks with their powerful nonlinear feature extraction capabilities[9–11], bringing new possibilities for retinal vessel segmentation. Now, the network-derived model based on U-Net achieves good vessel segmentation results[12,13]. However, these methods exhibit significant limitations when processing retinal vasculature: (1) Traditional U-Net frequently neglects local structural features, particularly for thin vessels or irregular vascular morphologies; (2) Background artifacts and noise inherent in fundus photography substantially degrade segmentation precision; (3) Despite strong local feature extraction, the architecture struggles to integrate global context with fine-grained details required for vessels exhibiting large scale variations. The structure in retinal vessel images is intricate and complex, and the images will be affected by light, resolution, noise and other factors during the information acquisition process of fundus camera. Therefore, the vessels in the image may appear blurred, broken or overlapped in different degrees. These factors make the segmentation task very difficult.

To address above problems, we improve U-Net and propose a new retinal vessel segmentation model. Firstly, we highlight the image features in the image pre-processing stage after grayscale conversion, normalization, CLAHE and Gamma correction, to emphasize the detailed characteristics, thereby accentuating the differences between vascular structures and surrounding backgrounds to provide clearer input for subsequent segmentation. Then, morphological processing is applied to enhance accuracy by eliminating tiny, noisy components and fine-tuning the vessel structure to increase the contrast between the vessel edges and the background, significantly improving the discernibility of vascular structures and enabling more precise segmentation of fine vessels. Secondly, the pyramid pooling structure (PSP) is integrated in the decoder to capture information of the image in different scales, thus realizing the fusion of multiscale information, and enhancing the characterization of local details. It enables the model to better segment the entire vessel structure. In addition, to enhance the segmentation capability of fine vessels, we add a dual attention block (DAB) in the decoder. The network may dynamically learn the relative relevance of each channel, which enables it to focus more on fine vessels by adaptively adjusting the importance of different channels to improve fine vessel segmentation capability. The network may modify its attention based on the spatial location of the pixel according to the spatial attention mechanism, which is crucial for modifying the attention of various regions to accommodate fine vessels, which effectively addresses the challenges of morphological variability and scale heterogeneity in retinal vessel images, further enhancing the model's focus on local details.

The main contributions of this paper are summarized as follows. (1) A new model for retinal vessel segmentation is designed. The proposed model integrates the PSP structure to capture scale features and improve the characterization of local detailed vessels; (2) A DAB module is presented in the model. It improves the segmentation of retinal fine vessels; (3) We conduct experimental comparison of our model using the DRIVE,

CHASE-DB1, and STARE datasets with models proposed recently. The results show that our model outperforms previous models.

2. RELATED WORK

2.1. Medical image segmentation

Deep learning has gradually become a commonly used technology for medical images segmentation [14]. Among them, U-Net has a simple but effective structural design and good applicability for domain-specific tasks such as medical images [15,16]. In addition, it can achieve better segmentation results with relatively less labeled data and training time. Currently, most segmentation methods based on deep learning use the encoder-decoder architecture. On this basis, the segmentation effect is improved by introducing attention mechanisms and optimizing the loss [17].

Some representative models based on U-Net are introduced as follows. Ahmad *et al.* [18] developed MH-UNet, which is a new architecture for medical image segmentation. It combines multiscale information and expands the receptive field size of feature maps. For example, Deng *et al.* used ELU-Net, which is an efficient lightweight U-Net where the deep skip connections have the same massive skip connections as the encoder for comprehensive feature extraction [19]. Nguyen *et al.* proposed a cascaded context module (CCM) and a balanced-attention module (BAM) [20]. Xiang *et al.* proposed a network that adds a multiscale feature extraction (MFE) module while reducing the number of convolutional kernel filters per layer [21]. It enhances the sensitivity of the model.

Those models above show that U-Net has become one of the most popular deep learning frameworks. However, it should be further studied in retinal vessel segmentation.

2.2. Retinal vessels segmentation

Deep neural networks demonstrate significant potential for retinal vasculature segmentation [22]. But the vessels are characterized by a complex structure, weak texture, and low edge contrast. Therefore, precise retinal vessel segmentation continues to be a difficult undertaking [23].

Scholars have conducted many studies to improve segmentation. Some researchers adopted the strategy of integrating multiscale features. For instance, Ye *et al.* proposed a U-shaped network, which enhances scale features and preserves more vessel details during decoding [24]. In order to learn adaptive weights to improve vessel features, Shi *et al.* computed the multiscale force field to capture its semantic relationship with the mask image pixels [25]. Zhou *et al.* divided the fundus image into plaques of different sizes and designed a cascaded dilated spatial pyramid pooling (CDSPP) module in the model to enhance the multiscale features and vessel continuity [26]. Despite the fact that multiscale fusion has many advantages and achieves respectable segmentation performance, they are nevertheless constrained by the following issues. On the one hand, multiscale fusion makes the network structure more complex, which reduces the interpretability of the model and makes it difficult to intuitively understand the exact process of segmentation of retinal images by the network. On the other hand, the introduction of multiscale fusion increases the number of parameters of the network, which requires more training data for adequate training, leading to the problem of over-fitting.

In retinal vessel segmentation, some scholars used attention mechanisms to improve models. For example, Guo *et al.* introduced a spatial attention module and replaced the original convolutional blocks of U-Net with dropout blocks to achieve network lightweighting [27]. Su *et al.* proposed a cross-scale attention (CSA) block to model multiscale interactions within vascular regions, thereby extracting rich contextual semantics [28]. Dong *et al.* proposed CRAUNet [29]. In CRAUNet, dropblock regularization is introduced to reduce over-fitting. Mahmoud *et al.* extended the Residual U-Net architecture through a cross-attention U-Net framework, en-

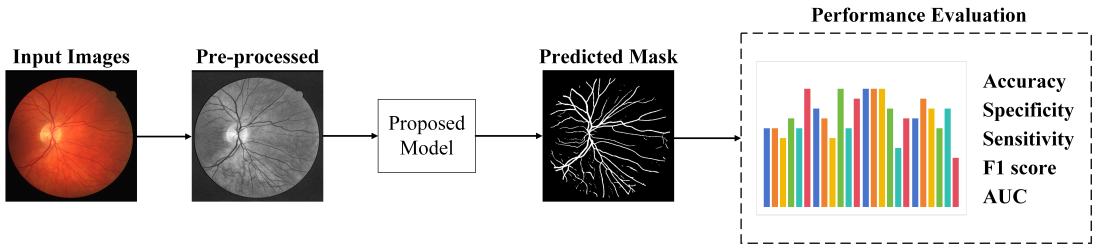

Figure 1. The overall workflow of our proposed method for retinal vessel segmentation.

hancing network adaptability and contextual comprehension capabilities to address critical limitations inherent in residual U-Net designs[30]. Wang *et al.* proposed a multi-attention method based on directed graph search[31]. By building a directed graph representation of the vessels, the method improves the topology and spatial connectivity. Although the introduction of the attention mechanism has become a mainstream method, the existing approaches are still unable to properly balance the global information capture of vessel structure and the segmentation accuracy of fine vessels.

To address these limitations, a morphological optimization strategy is proposed at the image processing stage. When processing retinal vessel images, the noise points in the images are reduced and the edge burrs of vessels are minimized, resulting in smoother and more continuous boundaries. In addition, the PSP structure and dual attention module are integrated in our proposed model. This allows the model to enhance the detailed features of vessels while capturing and fusing information at different scales, improving segmentation performance and reducing redundant information.

3. METHODOLOGY

To deal with the problems that exist in retinal vessel segmentation, an improved deep neural network based on U-Net is proposed. Figure 1 is the overall workflow of our method. In this paper, input images are processed by converting to grayscale, normalization, CLAHE, Gamma correction and Morphology. Then, data enhancement is performed using inversion and rotation. Finally, the proposed model is used for retinal vessel segmentation.

3.1. Image processing

The fundus images may be affected by lighting conditions, lens quality and other factors, resulting in poor image quality or insufficient contrast. Image processing is a crucial routine step in medical image segmentation, which can enhance image quality and contrast. In this paper, fundus images are processed using grayscale conversion, CLAHE, normalization, Gamma correction and morphological processing. Figure 2 presents the steps used for image pre-processing of input images.

At first, the input color images are converted to grayscale. Converting color images to grayscale for vessel segmentation does not entail complete information loss. Instead, it simplifies input data by reducing redundant color channels, allowing the model to focus on critical luminance and contrast properties essential for vascular structure identification. Grayscale imaging further enhances vessel boundaries and structural details. Additionally, this approach reduces computational burden by streamlining data processing, thereby improving training efficiency and model performance. Normalization adjusts the value of each pixel to standardize. CLAHE enhances the contrast of fundus images[32]. Gamma correction strengthens the morphology of vessels by adjusting the brightness of images. Morphological processing removes fine noise from an image by convolving its structural elements. It makes the vessel boundaries clearer. Retinal vessels face the problem of limited data availability. After completing dataset pre-processing, we implemented data augmentation on the

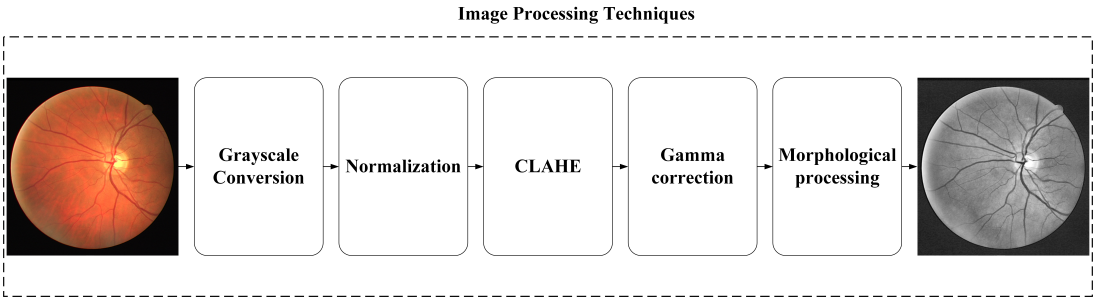

Figure 2. Image processing steps of the proposed method.

retinal fundus image collection. Our augmentation pipeline applied deterministic geometric transformations, specifically 90°, 180°, and 360° rotations combined with flipping, generating six distinct variants per original image. Additionally, all augmented images underwent standardized partitioning into 48×48 patches prior to model input, mitigating overfitting risks.

3.2. Network architecture

Figure 3 presents the structure of the proposed model, which is a U-shaped network. Considering the problem that the dataset images are small, the data features are also relatively small. In order to fully compute the image edge feature information, we perform padding around the image before the convolution calculation. In this way, the edge features will not be under-computed due to the positional problem even after the convolutional computation. However, as the network becomes more complex, the problem of vanishing gradients often arises, worsening the learning process. Therefore, we add a normalization operation Batch Normalization (BN) after the convolution operation, so that the distribution of features is brought back to the sensitive region of the function. The BN operation can turn the original distribution of eigenvalues of the data after each layer of convolution operation into a standard normal distribution, so that the distribution of features of the data can always be in the sensitive region of the activation function, and the loss function can also have a large change in the process of calculation [33]. In the decoder, we also add padding operation and BN operation. In addition, to capture global context information through pooling operations at different scales, we add a PSP structure after each up-sampling layer. The acquired information is spliced with the features of the current layer to improve image performance. In our model, a DAB module is added before the output layer. DAB is used to dynamically adjust the weights of the network so that it focuses on fine vessels. It can adjust its attention to changes in fine vessels based on the spatial location of the pixels. This can help to improve segmentation results.

3.3. PSP structure

The PSP is shown in Figure 4. Using average pooling kernels such as 1×1, 2×2, 3×3 and 6×6, we get pooling feature maps of different scales. Then, we up-sample the low-dimensional feature map by bilinear interpolation to make it the same scale as the originals. Finally, we splice it together to form the final global feature. The merge operation improves the global utilization information. The whole process is fusing the detailed features of retinal vessels with the global features containing contextual information.

3.4. DAB module

To enhance the extraction of fine vessel features, this paper proposes a DAB to enhance fine vessel segmentation in retinal images. The feature map is decoded using DAB and the image size is gradually adjusted to produce the results. The structure of DAB is shown in Figure 5. The channel attention mechanism enhances the feature response of the important channels, operating independently of input modalities by dynamically recalibrating feature channel responses in the input tensor. This enables the network to amplify critical features while suppressing less informative ones through learned inter-channel dependencies. Within the DAB module, single-channel features are processed identically to multi-channel inputs, applying channel-wise re-

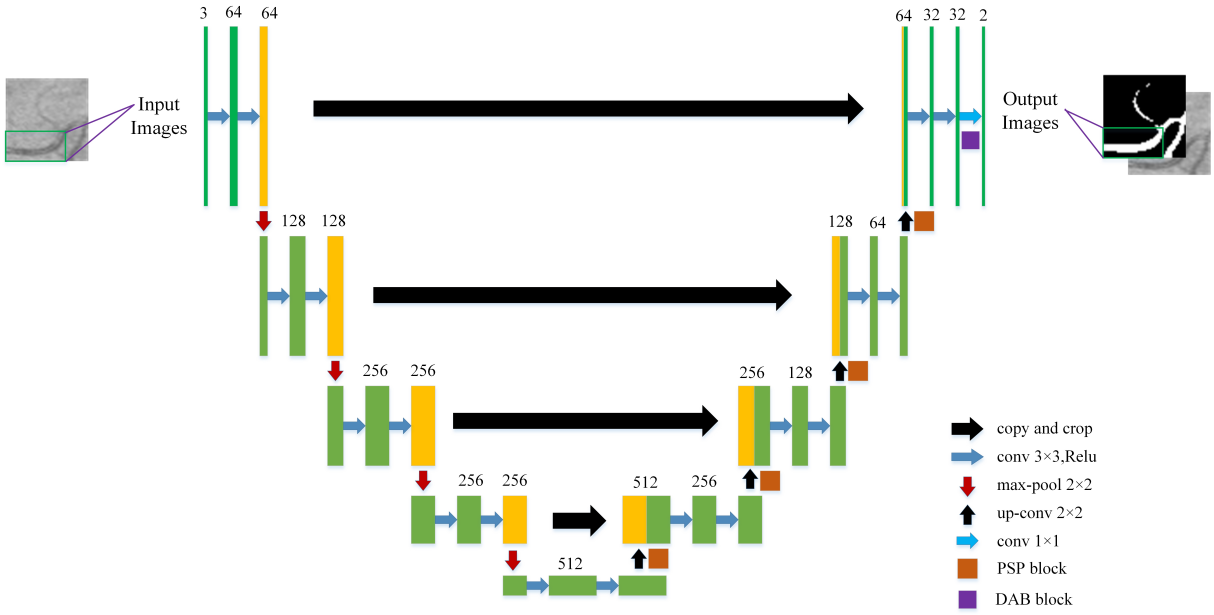

Figure 3. Network structure of the proposed model.

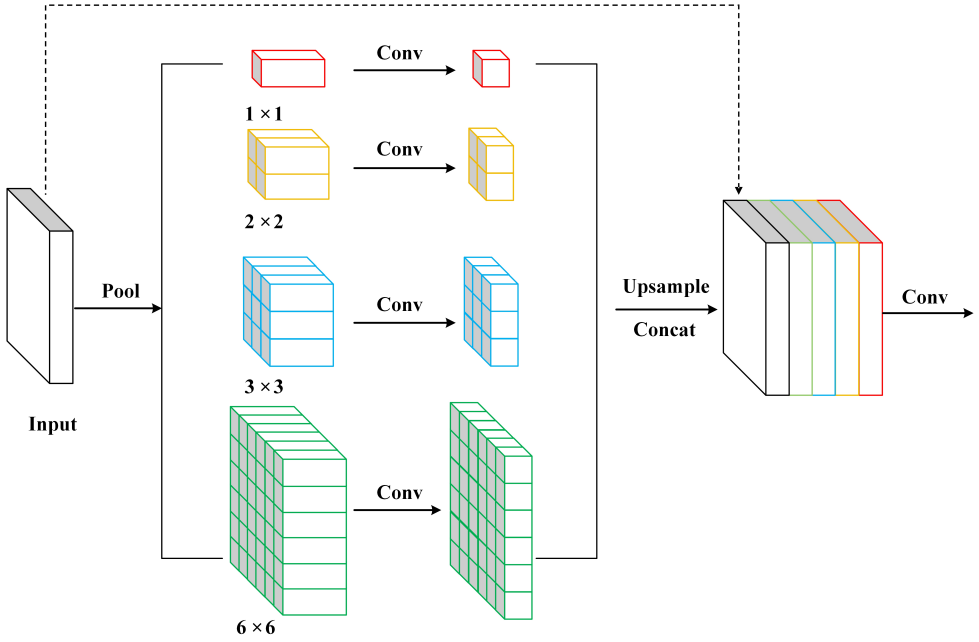

Figure 4. The structure of the PSP module. PSP: Pyramid pooling structure.

calibration to enhance discriminative feature representation, allowing the model to distinguish the fine vessels from the background more effectively. The spatial attention mechanism captures local features of fine vessels by enhancing the feature responses of important spatial locations, and more accurately localizes and segments fine blood vessels.

In Figure 5, we use 1×1, 3×3 and 1×1 convolution for feature extraction. On this basis, the channel attention

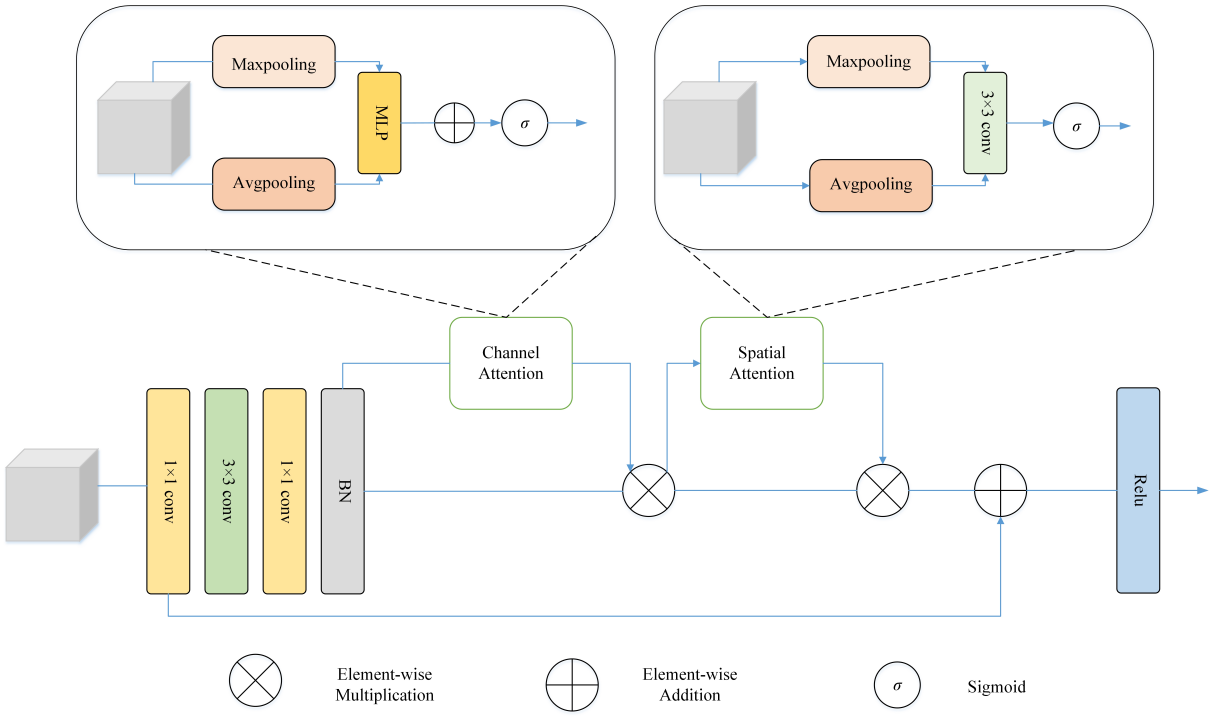

Figure 5. The structure of the DAB module. DAB: Dual attention block.

module flexibly adjusts the weights and suppresses irrelevant information, which can be given as:

$$I_c = N_c(I) = I * \sigma\left(LP\left(I^c_{\text{Avg}}\right) + LP\left(I^c_{\text{Max}}\right)\right). \tag{1}$$

where I is the input feature image, $*$ denotes element-wise multiplication operation, σ represents the Sigmoid function, LP indicates the weighted perceptron, I^c_{Avg} and I^c_{Max} are the average pooling and maximum pooling feature map outputs of I in spatial dimension, respectively, and I_c stands for the output.

At last, the enhanced results are fed into the spatial attention module to improve feature representation and suppress interference information, which can be given as:

$$I_s = N_s\left(I_r\right) = I_r * \sigma\left(f^{3\times3}\left(\left[I^s_{\text{Avg}}; I^s_{\text{Max}}\right]\right)\right). \tag{2}$$

where I^s_{Avg} and I^s_{Max} are the outputs of I_r in the channel dimension, respectively. $\left[I^s_{\text{Avg}}; I^s_{\text{Max}}\right]$ is the output of the channel dimension. I_s will get the final output.

3.5. Loss function

Since the binary cross entropy loss function can better optimize the model in the case of class imbalance, we choose it as the loss function of our model.

$$L = \sum_{i}^{N} \left(y_i \log\left(p_i\right) + \left(1 - y_i\right) \log\left(1 - p_i\right)\right) \tag{3}$$

where y_i represents the true probability and p_i denotes the probability that pixels i is predicted. N signifies the total number of pixels.

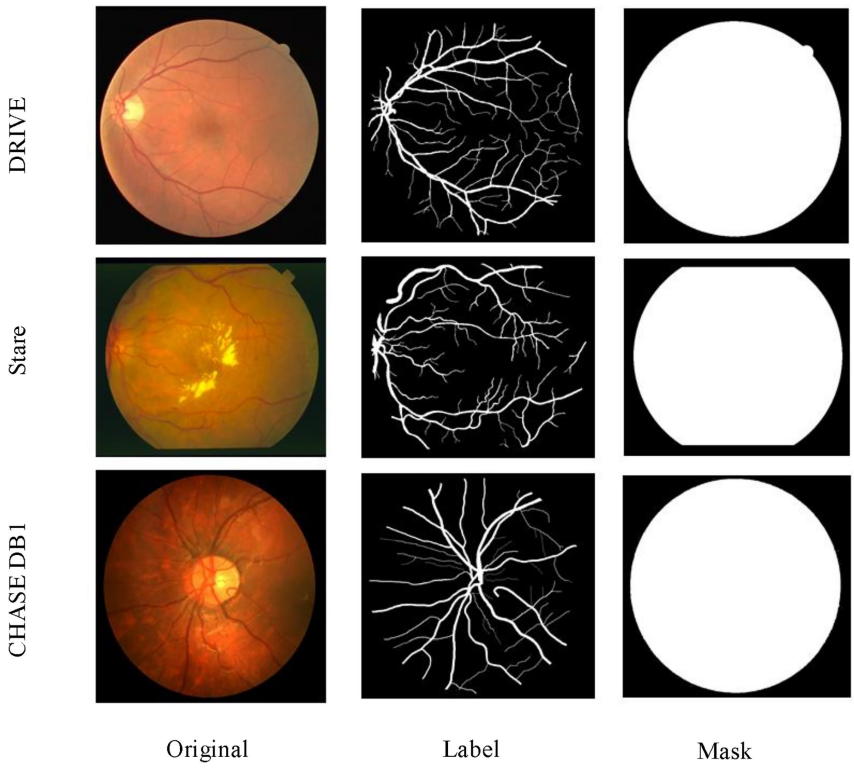

Figure 6. Original image, label image and mask image of three datasets.

4. EXPERIMENTS

4.1. Datasets and experimental configuration

The model proposed in this paper is verified on three public datasets, namely DRIVE, CHASE-DB1, and STARE[34,35]. (1) DRIVE: DRIVE is a widely used dataset. It includes 40 images, each with a size of 584 × 565 pixels and is detailed and manually labeled. Of these images, 20 are used for training and 20 for testing; (2) CHASE-DB1: CHASE-DB1 includes 28 images, which include expert-labeled vessel segmentation results. Each image has a size of 999 × 960 pixels, with 16 used for training and 12 for testing; (3) STARE: STARE contains 20 images from 10 different patients. Each image is a size of 700 × 605 pixels. Of these images, 14 are selected for training and 6 are used for testing. Some examples of these datasets are shown in Figure 6. Mask in Figure 6 refers to binary matrices that designate pixel-wise spatial localization of Regions of Interest, where foreground pixels are encoded as 1 and background regions as 0. Specifically for vasculature segmentation tasks, these masks usually indicate the true location of blood vessels. Through paired utilization with original fundus images, masks enable the model to learn discriminative feature representations for identifying and segmenting targeted specified structures or pathological regions, such as retinal vessels and lesion areas.

Our model is built based on the Tensorflow. Experiments are implemented on a NVIDIA GeForce GTX 1050 Ti platform using Keras (Tensorflow as the backend). During hyperparameter optimization, we conducted systematic experiments across varying epoch counts (50, 100, 150) and batch sizes (16, 32, 64), aimed at methodically assessing training duration effects. Based on comprehensive comparisons of the critical relationships between hyperparameters (such as segmentation accuracy and convergence rate) and model efficacy, we selected a batch size of 16 with 200 epochs and a 0.0001 learning rate using the Adam optimizer to balance accuracy and efficiency. To ensure equitable evaluation across all compared methods, we maintained rigorous consistency in experimental conditions: identical dataset partitions, standardized evaluation protocols, and uniform hardware environments. Each method underwent hyperparameter optimization within identical computational constraints. This framework explicitly minimizes confounding factors, guaranteeing that performance differentials solely reflect algorithmic capabilities rather than experimental variability.

4.2. Evaluation metrics

In order to more accurately evaluate our model, accuracy (ACC), sensitivity (SE), specificity (SP), F1 score ($F1$), and area under the ROC (AUC) metrics were used as most of the references in this field.

ACC: ACC indicates the proportion that the model correctly predicts, as given in

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TP (True Positive) means that the prediction algorithm segmentation result is the same as the total target vessel pixels manually annotated by experts; otherwise, it is FP (False Positive); for the background pixels, TN (True Negative) denotes that the total prediction algorithm segmentation result is the same as that manually annotated by experts; otherwise, it is FN (False Negative).

SE: SE indicates that the model can correctly identify vessels, as given in

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

SP: SP has the effect of correctly identifying negative class samples, as given in

$$SP = \frac{TN}{FP + TN} \tag{6}$$

$F1$: $F1$ score can combine the accuracy and completeness, as given in

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \tag{7}$$

AUC: AUC is used to evaluate the performance of the model in classifying vessels and non-vessels.

4.3. Experimental results

In the detection process, we input the pictures and masks of the test set into the model, and the size of each picture is 48×48. For each pixel point of the prediction result picture, there is also a pixel point corresponding to a pixel point in the real-value mask picture. By comparing these two pixel points, we utilize the confusion matrix to determine whether the pixel point of the prediction value is in the results. At the same time, the confusion matrix of various types of pixel points is counted and calculated, and finally the evaluation index results are given. The segmentation results on the three public datasets are shown in Tables 1-3.

First, we evaluated the model on the DRIVE dataset. There are many vessel ends in the low-contrast region of the image. It is critical to accurately segment the end of vessels. In Table 1, our model achieves 96.78%, 81.33%, 98.14%, 82.44% and 98.62% on ACC, $F1$, SP, SE and AUC, respectively. The observed $F1$ score of the proposed model reflects a deliberate design choice, which prioritizes detection rate and structural continuity of fine vasculature, compromising precision to mitigate vascular fragmentation and enhance clinical utility. Among them, our method obtained the best in ACC, SP, SE and AUC. This indicates that even in low-contrast regions, our proposed model can correctly segment the vessel ends. By introducing the PSP structure and DAB module, our model can effectively fuse multiscale features to capture relatively complete topology. It can also improve the segmentation of fine vessels.

Secondly, results conducted on the CHASE-DB1 dataset are presented in Table 2. Our model is more prominent in ACC and SP, reaching 97.17% for ACC and 99.02% for SP. SP of the proposed model is an im-

Table 1. Comparative experiments on the DRIVE datasets

Method	Year	*ACC*	*F1*	*SP*	*SE*	*AUC*
U-Net [36]	2016	95.57	81.26	98.11	76.15	97.06
R2U-Net [37]	2018	95.56	79.83	98.08	76.06	88.62
CE-Net [38]	2019	96.16	82.69	98.09	80.23	98.10
CPFNet [39]	2020	95.51	81.88	97.83	79.60	97.14
MHSU-Net [40]	2021	95.74	82.86	97.91	80.84	98.13
DEF-Net [41]	2022	94.57	78.51	96.99	77.99	97.19
AAU-Net [42]	2022	95.58	82.11	97.84	79.53	97.64
Wave-Net [43]	2023	95.61	82.54	97.64	81.64	-
IterMiU-Net [44]	2023	95.75	82.77	98.05	80.06	98.11
PGU-Net [45]	2024	95.72	82.50	98.14	79.15	98.12
nnU-Net [46]	2024	96.52	78.58	97.90	81.62	98.40
Ours	-	96.78	81.33	98.14	82.44	98.62

Table 2. Comparative experiments on the CHASE-DB1 datasets

Method	Year	*ACC*	*F1*	*SP*	*SE*	*AUC*
U-Net [36]	2016	96.37	80.25	98.21	78.38	97.10
R2U-Net [37]	2018	96.44	79.28	98.30	78.39	84.95
CPFNet [39]	2020	96.42	82.14	98.40	79.94	97.67
DEF-Net [41]	2022	94.41	77.73	97.01	76.61	95.80
RCAR-UNet [47]	2023	95.66	74.70	97.98	74.75	-
SDDC-Net [48]	2023	96.69	79.65	97.89	82.68	98.45
ETAU-Net [49]	2023	96.99	-	98.26	78.03	97.79
PGU-Net [45]	2024	96.58	80.99	98.17	80.51	98.57
nnU-Net [46]	2024	96.89	81.53	98.64	80.55	98.34
Ours	-	97.17	78.50	99.02	78.07	97.74

Table 3. Comparative experiments on the STARE datasets

Method	Year	*ACC*	*F1*	*SP*	*SE*	*AUC*
U-Net [36]	2016	96.17	80.78	98.67	75.40	97.02
LadderNet [50]	2018	96.08	80.75	98.75	75.55	98.17
AttU-Net [51]	2019	96.36	82.35	98.60	78.65	98.19
CPFNet [39]	2020	96.18	81.56	98.52	77.83	97.96
LightWeight [52]	2021	96.31	82.48	98.51	79.03	97.36
AAU-Net [42]	2022	96.32	82.43	98.68	78.07	97.61
IterMiU-Net [44]	2023	96.32	82.33	98.68	78.19	98.51
DASENet [53]	2023	97.28	82.31	98.39	84.02	99.02
IMFF-Net [54]	2024	97.07	83.47	98.69	86.34	-
nnU-Net [46]	2024	97.33	82.37	98.49	85.06	99.13
Ours	-	97.50	80.98	98.09	87.75	99.32

provement of 1.13%, 0.76% and 0.0.85% compared to the latest SDDC-Net, ETAU-Net and PGU-Net, respectively. The *ACC* of the proposed model improves by 0.48%, 0.18% and 0.59% compared to the latest SDDC-Net, ETAU-Net and PGU-Net, respectively. When compared to nnU-Net, our model performs excellently in

Table 4. Experimental time analysis on the DRIVE datasets

Model	U-Net	nnU-Net	Ours
Experimental time (s)	0.1225	0.6834	0.3445

global structure recognition and noise resistance (the ACC and SP of our proposed model increase by 0.28% and 0.38%, respectively). There is a certain gap in $F1$ and SE between our proposed model and some other models. This is primarily due to the attention mechanism failing to adequately focus on low-contrast vessels, and morphological processing potentially damaging weak-signal vessels, leading to reduced sensitivity.

Finally, Table 3 describes comparative results based on the STARE dataset, which includes many lesions, including hemorrhages and exudative inflammation, and half of the pathologic images[26]. However, at the same time, the STARE dataset has a relatively high image quality and clear vascular structures. From Table 3, compared with the models in recent papers (see[53] and[54]), our method outperforms them in ACC, SE, and AUC metrics, and SE of our model is 3.73% higher than that of the model in[53] (the second best model in SE). The main reason is that our model is allowed to focus on different regions at various spatial scales to enhance the detailed information in specific regions, due to the proposed PSP structure and DAB module. This helps the model to better distinguish different structures in fundus images and improves segmentation accuracy.

4.4. Analysis on visualization and experimental time

In order to visualize the segmentation performance, this study was analyzed through visualization. Figure 7 shows the comparison image of our model, the U-Net and the nnU-Net. Although U-Net can segment the main vessels, it performs poorly in the fine vessels, resulting in poor vessel connectivity. nnU-Net achieves marginally superior overall performance compared to U-Net, yet exhibits consistently suboptimal preservation of morphological details in fine vessel segmentation. In contrast, our proposed new model performs better. Segmentation accuracy is primarily augmented through two interconnected mechanisms. Firstly, DAB's dual-attention framework adaptively prioritizes capillary signatures via channel-wise feature recalibration while spatially modulating focus to accommodate morphological heterogeneity. Secondly, PSP's hierarchical fusion concurrently processes macrovascular context and microstructural details, establishing comprehensive multiscale representations essential for complex vasculature. The interaction between the PSP structure and the DAB module preserves more vessel details. To sum up, the proposed model performs exceptionally.

To show the computation efficiency, the experimental time of the three models (our proposed model, the U-Net and the nnU-Net) on the DRIVE dataset is given out, which is listed in Table 4. The results show that our proposed model demonstrates significant efficiency advantages. U-Net requires 0.1225 seconds per image. While nnU-Net's adaptive pre-processing and cross-validation strategies enhance segmentation accuracy, these improvements incur substantially increased computational costs, yielding 0.6834 s per image. Comparatively, though introducing additional model complexity relative to U-Net, our proposed novel architecture achieves only 0.3445 s per image while maintaining competitive segmentation performance.

4.5. Ablation experiments

Among three datasets used, DRIVE is the most generally applicable dataset, and we use it as an example for ablation experiments. The proposed method has three main components, including morphological pre-processing (PRE-Pro), PSP and DAB modules. U-Net was selected as the baseline and three important components were incorporated in sequence. Table 5 presents the results.

Firstly, morphological pre-processing was applied to the baseline (U-Net + PRE-Pro). It is clear that the performance of the U-Net + PRE-Pro model exceeds that of the baseline network. The information obtained from Table 5 allows us to determine that the evaluation metrics were all improved to some extent after adding the

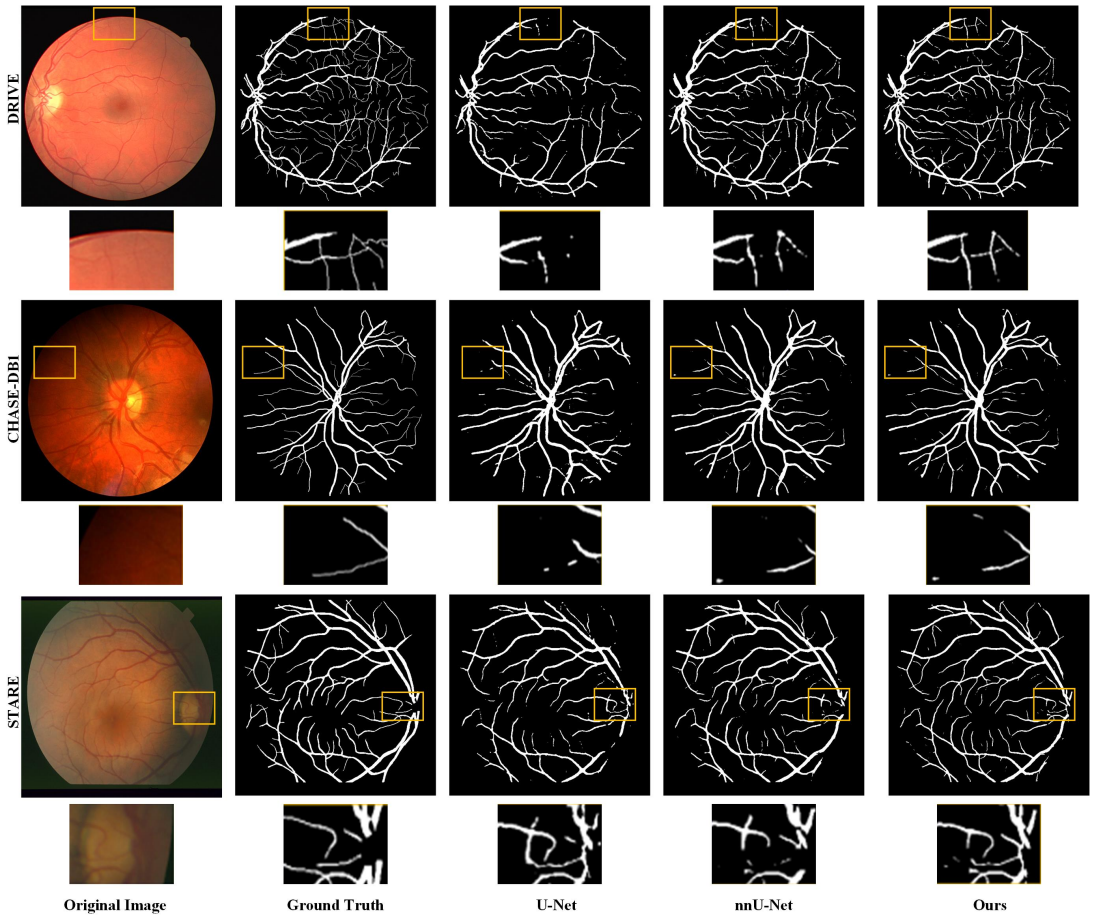

Figure 7. Comparison of the visualization results between our model, the U-Net and the nnU-Net.

Table 5. Ablation experiments on the DRIVE datasets

Method	*ACC*	*F1*	*SP*	*SE*	*AUC*
U-Net	96.39	79.01	97.38	80.53	97.16
U-Net + PRE-Pro	96.48	79.63	98.00	80.49	98.40
U-Net + PRE-Pro + PSP	96.55	79.92	97.99	81.28	98.51
U-Net + PRE-Pro + DAB	96.69	81.25	98.30	80.52	98.50
U-Net + PRE-Pro + PSP + DAB	96.78	81.33	98.14	82.44	98.62

pre-processing optimization method morphological treatment to the baseline (U-Net + PRE-Pro). Compared with U-Net, $F1$ score is improved by 0.62% from 79.01% to 79.63%, and the AUC is improved by 1.24%. The main reason was that the morphological manipulation resulted in smoother and clearer edges of the vessel region, which helped to reduce edge fragments and noise in the segmentation results. In addition, the morphological processing enhanced the connectivity and morphological features of the vessel structure, which made it easier for the segmentation network to recognize and segment the vessel region.

Second, we added the PSP structure (U-Net + PRE-Pro + PSP). In Table 5, the performance is improved in terms of SE. It can be seen that since the PSP structure is able to capture and integrate feature information, it helps the network to understand blood vessels and facilitate effective feature fusion. The capacity to extract features is enhanced as compared to U-Net and superior segmentation performance can be obtained.

We import the DAB module into the model (U-Net + PRE-Pro) called (U-Net + PRE-Pro + DAB) to validate its effectiveness. The network is assisted in learning the relationships between different channels by the channel attention module, which enables it to better extract useful feature information. By enhancing the weights of key channels, the network is able to better recognize vessel structures. Thereby, $F1$ score is improved by 2.24% compared to U-Net. In addition, the spatial attention module concentrates more attention on the vessel region, which enhances the recognition of vessel structures. This helps to reduce missed detections and false negatives, which in turn improves the SP by 0.92% compared to U-Net.

Finally, to efficiently capture contextual information and enhance segmentation of fine vessel structures, we designed a new model referred to as (U-Net + PRE-Pro + PSP + DAB). Compared with U-Net, our model can improve the segmentation of fine blood vessels due to the DAB module. In Table 5, compared with U-Net, our method shows that all evaluation metrics are improved. ACC, $F1$, SP, SE is improved by 0.39%, 2.32%, 0.76%, 1.91%, and the AUC is improved by 1.46%. Meanwhile, compared with U-Net + PRE-Pro + DAB, the SP of our proposed model dropped a little bit. The main reason is that PSP aggregates too much information at different scales, resulting in fuzzy or confusing feature representations, which cause redundancy and conflict of information, thus affecting the segmentation performance.

Overall, our model shows excellent performance in retinal vessel segmentation. Specifically, the results show that each component in the model has a certain improvement on the segmentation results. After integrating the components, we also obtain the best segmentation results.

5. CONCLUSION

Considering that retinal vessels are characterized by complex structure, fragility of blood vessels, and blurred boundaries of fine vessels, we proposed a new retinal vessel segmentation model. Specifically, we add an optimization method of morphological processing in the image pre-processing stage to reduce noise points and preserve as much vessel structure as possible. The PSP structure and DAB module added in the decoder section help to capture global and local information, enhance the focus on crucial aspects while suppressing irrelevant features. Our method was experimented with and analyzed on three publicly available datasets, showing improved accuracy in blood vessel segmentation.

Although our experimental results show promise and excellence of the model, it can still be further improved. On CHASE-DB1 dataset, which has poor fundus image quality and less vascular connectivity, our model's segmentation results are not well balanced and have limited validity. In addition, the $F1$ scores of our model on all three publicly available datasets were not very good, due to the complexity of the model, its recognition of minority category samples has not been significantly improved, thus affecting the scores. In the future, we will achieve model light-weighting while addressing the issue of uneven contrast and vessel discontinuities, while further improving the performance of the model.

DECLARATIONS

Authors' contributions

Funding acquisition: Ni, J.
Project administration: Ni, J.; Gu, Y.
Writing - original draft: Shi, J.
Writing - review and editing: Zhan, Q.; Zhang, Z.; Gu, Y.

Availability of data and materials

Publicly available datasets were analyzed in this study. These data can be found here: https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction/data, https://www.kaggle.com/datasets/buffyhridoy/chase-db1, and https://www.kaggle.com/datasets/vidheeshnacode/stare-dataset.

Financial support and sponsorship

This work was partly supported by the National Natural Science Foundation of China (61873086).

Conflicts of interest

Ni, J. is an Associate Editor of the journal *Intelligence & Robotics*. Ni, J. was not involved in any steps of the editorial processing, notably including reviewer selection, manuscript handling, and decision-making. The other authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Liu, Y.; Shen, J.; Yang, L.; Bian, G.; Yu, H. ResDO-UNet: a deep residual network for accurate retinal vessel segmentation from fundus images. *Biomed. Signal Process. Control* **2023**, *79*, 104087. DOI
2. Rayachoti, E.; Gundabatini, S. G.; Vedantham, R. Recurrent Residual Puzzle based Encoder Decoder Network (R2-PED) model for retinal vessel segmentation. *Multimed. Tools Appl.* **2024**, *83*, 39621 – 45. DOI
3. Pan, J.; Gong, J.; Yu, M.; Zhang, J.; Guo, Y.; Zhang, G. A multilevel remote relational modeling network for accurate segmentation of fundus blood cessels. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–14. DOI
4. Li, Z.; Guan, J.; Wang, H. A novel dual-supervised convolutional network for retinal vessel segmentation. In *2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, Xi'an, China. Oct 28-30, 2022. IEEE; 2022. pp. 567–71. DOI
5. Chen, Y.; Ge, P.; Wang, G.; Weng, G.; Chen, H. An overview of intelligent image segmentation using active contour models. *Intell. Robot.* **2023**, *3*, 23-55. DOI
6. Wang, H.; Zuo, W.; Li, B.; Pan, X.; Liu, Z.; Lan, R. Dilation-supervised learning: a novel strategy for scale difference in retinal vessel segmentation. *IEEE Trans. Artif. Intell.* **2024**, *5*, 1693–707. DOI
7. Kumar, K. S.; Singh, N. P. Analysis of retinal blood vessel segmentation techniques: a systematic survey. *Multimed. Tools Appl.* **2023**, *82*, 7679 – 733. DOI
8. Singh, L. K.; Khanna, M.; Thawkar, S.; Singh, R. Deep-learning based system for effective and automatic blood vessel segmentation from Retinal fundus images. *Multimed. Tools Appl.* **2024**, *83*, 6005 – 49. DOI
9. Ni, J.; Chen, Y.; Chen, Y.; Zhu, J.; Ali, D.; Cao, W. A survey on theories and applications for self-driving cars based on deep learning methods. *Appl. Sci.* **2020**, *10*, 2749. DOI
10. Ye, C.; Che, K.; Yao, Y.; et al. A deep learning-based system for accurate detection of anatomical landmarks in colon environment. *Intell. Robot.* **2024**, *4*, 164-78. DOI
11. Ni, J.; Chen, Y.; Tang, G.; Shi, J.; Cao, W.; Shi, P. Deep learning-based scene understanding for autonomous robots: a survey. *Intell. Robot.* **2023**, *3*, 374-401. DOI
12. Wang. H.; Xu, G.; Pan, X.; et al. Attention-inception-based U-Net for retinal vessel segmentation with advanced residual. *Comput. Electr. Eng.* **2022**, *98*, 107670. DOI
13. Zhou, Z.; Rahman Siddiquee, M. M.; 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*, Granada, Spain. Springer, Cham; 2018. pp. 3–11. DOI
14. Li, L.; Qin, J.; Lv, L.; et al. ICUnet++: an Inception-CBAM network based on Unet++ for MR spine image segmentation. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 3671 – 83. DOI
15. Tang, G.; Ni, J.; Chen, Y.; Cao, W.; Yang, S. X. An improved CycleGAN based model For low-light image enhancement. *IEEE Sens. J.* **2024**, *24*, 21879-92. DOI

16. Yang, X.; Li, Z.; Guo, Y.; Zhou, D. DCU-net: a deformable convolutional neural network based on cascade U-net for retinal vessel segmentation. *Multimed. Tools Appl.* **2022**, *81*, 15593 – 607. DOI
17. Kang, N.; Wang, M.; Pang, C.; et al. Cross-patch feature interactive net with edge refinement for retinal vessel segmentation. *Comput. Biol. Med.* **2024**, *174*, 108443. DOI
18. Ahmad, P.; Jin, H.; Alroobaea, R.; et al. MH UNet: a multi-scale hierarchical based architecture for medical image segmentation. *IEEE Access* **2021**, *9*, 148384–408. DOI
19. Deng, Y.; Hou, Y.; Yan, J.; Zeng, D. ELU-Net: an efficient and lightweight U-Net for medical image segmentation. *IEEE Access* **2022**, *10*, 35932–41. DOI
20. Nguyen, T. C.; Nguyen, T. P.; Diep, G. H.; Tran-Dinh, A. H.; Nguyen, T. V.; Tran, M. T. CCBANet: cascading context and balancing attention for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2021*, Strasbourg, France. Springer, Cham; 2021. pp. 633–43. DOI
21. Xiang, Z.; Ning, C.; Li, M.; et al. AFFD-Net: a dual-decoder network based on attention-enhancing and feature fusion for retinal vessel segmentation. *IEEE Access* **2023**, *11*, 45871–87. DOI
22. Liskowski, P.; Krawiec, K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* **2016**, *35*, 2369–80. DOI
23. Li, J.; Gao, G.; Yang, L.; Liu, Y. A retinal vessel segmentation network with multiple-dimension attention and adaptive feature fusion. *Comput. Biol. Med.* **2024**, *172*, 108315. DOI
24. Ye, Y.; Pan, C.; Wu, Y.; Wang, S.; Xia, Y. MFI-Net: multiscale feature interaction network for retinal vessel segmentation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4551–62. DOI
25. Shi, T.; Ding, X.; Zhou, W.; et al. Affinity feature strengthening for accurate, complete and robust vessel segmentation. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 4006–17. DOI
26. Zhou, W.; Bai, W.; Ji, J.; Yi, Y.; Zhang, N.; Cui, W. Dual-path multi-scale context dense aggregation network for retinal vessel segmentation. *Comput. Biol. Med.* **2023**, *164*, 107269. DOI
27. Guo, C.; Szemenyei, M.; Yi, Y.; Wang, W.; Chen, B.; Fan, C. SA-UNet: spatial attention U-Net for retinal vessel segmentation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy. Jan 10-15, 2021. IEEE; 2021. pp. 1236–42. DOI
28. Su, H.; Gao, L.; Wang, Z.; Yu, Y.; Hong, J.; Gao, Y. A hierarchical full-resolution fusion network and topology-aware connectivity booster for retinal vessel segmentation. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–16. DOI
29. Dong, F.; Wu, D.; Guo, C.; Zhang, S.; Yang, B.; Gong, X. CRAUNet: a cascaded residual attention U-Net for retinal vessel segmentation. *Comput. Biol. Med.* **2022**, *147*, 105651. DOI
30. Mahmoud, M.; Mansour, M.; Elrefai, H. M.; Hamed, A. J.; Rashed, E. A. Enhanced retinal arteries and veins segmentation through deep learning with conditional random fields. *Biomed. Signal Process. Control* **2025**, *106*, 107747. DOI
31. Wang, G.; Huang, Y.; Ma, K.; et al. Automatic vessel crossing and bifurcation detection based on multi-attention network vessel segmentation and directed graph search. *Comput. Biol. Med.* **2023**, *155*, 106647. DOI
32. Mishra, A. Contrast Limited Adaptive Histogram Equalization (CLAHE) approach for enhancement of the microstructures of friction stir welded joints. *arXiv* **2021**, arXiv:2109.00886. https://doi.org/10.48550/arXiv.2109.00886. (accessed 21 Jul 2025)
33. Ni, J.; Shen, K.; Chen, Y.; Yang, S. X. An improved SSD-like deep network-based object detection method for indoor scenes. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1-15. DOI
34. Yang, D.; Zhao, H.; Yu, K.; Geng, L. NAUNet: lightweight retinal vessel segmentation network with nested connections and efficient attention. *Multimed. Tools Appl.* **2023**, *82*, 25357 – 79. DOI
35. Du, X. F.; Wang, J. S.; Sun, W. Z.; Zhang, Z. H.; Zhang, Y. H. Bi-directional ConvLSTM residual U-Net retinal vessel segmentation algorithm with improved focal loss function. *J. Intell. Fuzzy Syst.* **2024**, *46*, 1-20. https://www.researchgate.net/publication/379530297_Bi-directional_ConvLSTM_residual_U-Net_retinal_vessel_segmentation_algorithm_with_improved_focal_loss_function. (accessed 21 Jul 2025)
36. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, Munich, Germany. Springer, Cham; 2015. pp. 234–41. DOI
37. Zahangir Alom, M.; Yakopcic, C.; Taha, T. M.; Asari, V. K. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). In *NAECON 2018 - IEEE National Aerospace and Electronics Conference*, Dayton, USA. Jul 23-26, 2018. IEEE; 2018. pp. 228-33. DOI
38. Gu, Z.; Cheng, J.; Fu, H.; et al. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–92. DOI
39. Feng, S.; Zhao, H.; Shi, F.; et al. CPFNet: context pyramid fusion network for medical image segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 3008–18. DOI
40. Ma, H.; Zou, Y.; Liu, P. X. MHSU-Net: a more versatile neural network for medical image segmentation. *Comput. Methods Programs Biomed.* **2021**, *208*, 106230. DOI
41. Li, J.; Gao, G.; Yang, L.; Liu, Y.; Yu, H. DEF-Net: a dual-encoder fusion network for fundus retinal vessel segmentation. *Electronics* **2022**, *11*, 3810. DOI
42. Chen, G.; Li, L.; Dai, Y.; Zhang, J.; Yap, M. H. AAU-Net: an adaptive attention U-Net for breast lesions segmentation in ultrasound images. *IEEE Trans. Med. Imaging* **2023**, *42*, 1289–300. DOI
43. Liu, Y.; Shen, J.; Yang, L.; Yu, H.; Bian, G. Wave-Net: a lightweight deep network for retinal vessel segmentation from fundus images. *Comput. Biol. Med.* **2023**, *152*, 106341. DOI

44. Kumar, A.; Agrawal, R. K.; Joseph, L. IterMiUnet: a lightweight architecture for automatic blood vessel segmentation. *Multimed. Tools Appl.* **2023**, *82*, 43207–31. DOI
45. Wang, Z.; Jia, L. V.; Liang, H. Partial class activation mapping guided graph convolution cascaded U-Net for retinal vessel segmentation. *Comput. Biol. Med.* **2024**, *178*, 108736. DOI
46. Isensee, F.; Wald, T.; Ulrich, C.; et al. nnU-Net revisited: a call for rigorous validation in 3D medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. Springer, Cham; 2024. pp. 488–98. DOI
47. Ding, W.; Sun, Y.; Huang, J.; et al. RCAR-UNet: retinal vessel segmentation network algorithm via novel rough attention mechanism. *Inform. Sci.* **2024**, *657*, 120007. DOI
48. Yang, B.; Qin, L.; Peng, H.; Guo, C.; Luo, X.; Wang, J. SDDC-Net: a U-shaped deep spiking neural P convolutional network for retinal vessel segmentation. *Digit. Signal Process.* **2023**, *136*, 104002. DOI
49. Kato, S.; Hotta, K. Expanded tube attention for tubular structure segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2024**, *19*, 2187-93. DOI
50. Zhuang, J. LadderNet: multi-path networks based on U-Net for medical image segmentation. *arXiv* **2018**, arXiv:1810.07810. https://doi.org/10.48550/arXiv.1810.07810. (accesssed 21 Jul 2025)
51. Schlemper, J.; Oktay, O.; Schaap, M.; et al. Attention gated networks: learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. DOI
52. Li, X.; Jiang, Y.; Li, M.; Yin, S. Lightweight attention convolutional neural network for retinal vessel image segmentation. *IEEE Trans. Ind. Inform.* **2021**, *17*, 1958–67. DOI
53. Zhu, J.; Lan, Z.; Wen, X.; Cai, S.; Xu, Y. DASENet: a detail aware U-Net with shuffle excitation for retinal vessel segmentation. In *2023 6th International Conference on Software Engineering and Computer Science (CSECS)*, Chengdu, China. Dec 22-24, 2023. IEEE; 2023. p. 1–7. DOI
54. Liu, M.; Wang, Y.; Wang, L.; Hu, S.; Wang, X.; Ge, Q. IMFF-Net: an integrated multi-scale feature fusion network for accurate retinal vessel segmentation from fundus images. *Biomed. Signal Process. Control* **2024**, *91*, 105980. DOI