



Research article

EAMR-Net: A multiscale effective spatial and cross-channel attention network for retinal vessel segmentation

G. Prethija^{1,*} and Jeevaa Katiravan²

¹ School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India

² Department of Information Technology, Velammal Engineering College, Chennai 600066, India

* **Correspondence:** Email: prethija.g@vit.ac.in.

Abstract: Delineation of retinal vessels in fundus images is essential for detecting a range of eye disorders. An automated technique for vessel segmentation can assist clinicians and enhance the efficiency of the diagnostic process. Traditional methods fail to extract multiscale information, discard unnecessary information, and delineate thin vessels. In this paper, a novel residual U-Net architecture that incorporates multi-scale feature learning and effective attention is proposed to delineate the retinal vessels precisely. Since drop block regularization performs better than drop out in preventing overfitting, drop block was used in this study. A multi-scale feature learning module was added instead of a skip connection to learn multi-scale features. A novel effective attention block was proposed and integrated with the decoder block to obtain precise spatial and channel information. Experimental findings indicated that the proposed model exhibited outstanding performance in retinal vessel delineation. The sensitivities achieved for DRIVE, STARE, and CHASE_DB datasets were 0.8293, 0.8151 and 0.8084, respectively.

Keywords: residual; attention; drop block; spatial pooling; U-Net

1. Introduction

Diabetic retinopathy is a complication of diabetes that affects a significant number of people worldwide. Diabetic retinopathy adversely impacts the retinal blood vessels, leading to eye complications. It may lead to vision problems and may cause blindness if left untreated. This condition

affects approximately one-third of people with diabetes, and its prevalence increases with the duration of diabetes [1]. According to the American Diabetes Association, around 28% of people with diabetes over the age of 40 have diabetic retinopathy [2].

Blood vessel segmentation is an important step in analyzing retinal images because it separates blood vessels from the surrounding tissue. Manual labeling of blood vessels in medical images is a labor-intensive process that involves experts manually delineating vessel structures. Also, it is time-consuming as well as highly challenging for manual experts to detect small or complex blood vessels.

There are a lot of challenges faced during blood vessel segmentation namely: (a) Blood vessels often have low contrast compared to the background, making it difficult to distinguish them accurately; (b) since blood vessels often overlap with each other, it is difficult to separate individual vessels; and (c) the presence of vessels with different widths and branching patterns adds complexity to the segmentation task.

Addressing these challenges requires the development of robust segmentation algorithms that can handle variations in vessel appearance, structure, and imaging conditions. Thresholding [3,4], edge-based techniques [5,6], mathematical morphology [7,8], graph-based techniques [9], machine learning [10], and deep learning [11] are a few methods used to accurately segment blood vessels. Machine learning techniques such as SVM [12], Random Forest [13], and CNN [14] have shown promising results in segmentation. However, the U-Net architecture and its variants have recently shown better performance in blood vessel segmentation. Hence, effective attention with multi-scale learning residual U-Net (EAMR-Net) is proposed to enhance the effectiveness of retinal blood vessel segmentation. The major contributions are:

- 1). Residual block is employed instead of convolution block to avoid vanishing gradients and drop block is added after the convolution block to prevent overfitting.

- 2). A novel attention mechanism is introduced in the decoder block to treat all features uniquely. An effective spatial and channel attention mechanism is employed to preserve dimensionality and learn important features and interactions between channels in a cross-channel manner. Effective attention focuses on more information regions and ignores the background information.

- 3). A multi-scale feature learning (MSFL) module is proposed instead of the skip connection to avoid the semantic gap issue and enhance the model's ability to capture intricate details. Extracting hierarchical information will help to identify both large and finer details. To obtain image-level features, we add average pooling and up-sampling blocks. In comparison with the existing methods, the performance of our work has shown better accuracy.

The remaining sections of the paper are organized as follows: In Section 2, various research studies on retinal vascular segmentation are discussed. Section 3 provides an in-depth description of the research methodology and experimental design, while Section 4 discusses the experimental results obtained and ablation studies. Last, Section 5 provides a comprehensive summary and final remarks on the study.

2. Related works

Based on the availability of label information, vessel segmentation in retinal images can be classified into unsupervised (unlabeled) and supervised learning techniques.

2.1. Unsupervised learning techniques

Unsupervised techniques do not use manual annotations for detecting blood vessels. Graph-based methods, clustering methods, and deep learning methods such as autoencoders and GAN (Generative Adversarial Networks) are a few unsupervised approaches used to segment the retinal vessels.

Padmapriya et al. [15] used basic pre-processing techniques, primary curvatures, ISODATA algorithm, and labeling to segment blood vessels with higher accuracy. Muzammil et al. [16] applied pre-processing techniques and used top-hat morphological operations to eliminate the noise. Matched and Gabor wavelet filters are applied to extract vessels. Then, binarization is applied using the human visual system. After applying a few post-processing procedures, the segmented results are produced. Qaiser et al. [17] suggested Otsu's thresholding method to create an automated retinal segmentation strategy that made use of pre-processing methods such as G-channel extraction, CLAHE, PCA, filtering, and segmentation. Upadhyay et al. [18] implemented two different varied scale transform approaches namely local directional wavelet and global curvelet. To recover boundary pixels, morphological thickness correction is proposed.

2.2. Supervised learning techniques

Initially, the U-Net architecture has shown a better performance. Few researchers have focused on building lightweight models and few others focused on performance improvement.

Ronneberger et al. [19] have developed a U-Net architecture that has shown better performance in biomedical segmentation. Laibacher et al. [20] proposed a new architecture that employs M2Unet as an encoder and contractive bottlenecks in the decoder part. Their focus is to reduce the parameter size used for training. This model is suitable for high-resolution image analysis. Boudegga et al. [21] proposed a model with convolution blocks of lightweight to reduce the computational complexity in terms of execution time. His model has shown better segmentation performance. Yang et al. [22] evaluated the performance of DCU-Net on three publicly available datasets and found that it had a lower number of parameters and a faster inference time.

Recently, an attention-based U-Net was developed to detect blood vessels accurately. Wang et al. [23] introduced a new deep learning architecture for retinal vessel segmentation called the Attention Inception-based U-Net- Advanced Residual (AIU-AR). The proposed model incorporates the attention mechanism and residual connections to enhance the accuracy of vessel segmentation in retinal images. The AIU-AR model has exhibited exceptional performance in comparison to several cutting-edge segmentation models when assessed using publicly accessible datasets. Through the incorporation of attention mechanisms and residual connections, researchers observed significant improvements in accuracy.

Wang et al. [24] developed a model, namely SAU-Net, that incorporates ResBlock along with inception blocks in the encoder part and, ResBlock with squeeze and excitation in the decoder part to eliminate the impact of hard exudate in fundus images. Dong et al. [25] proposed a model named CRAUnet which is a cascaded structure architecture that uses a residual block and an attention block. They also added a drop block to solve overfitting issues. Their contribution is multiscale feature extraction. Liu et al. [26] developed a ResDOU-Net that uses residual overparameterized convolution block, pooling fusion block, and atrous convolution to extract features in multiple scales, which demonstrated robustness to noise. Ren et al. [27] use Bi-FPN in U-Net, which uses depth-wise separable convolution for multiscale fusion. Li et al. [28] proposed GDFNet, which uses an

enhancement network block to detect thin vessels, a global segmentation network to extract global features, and an attention fusion network that combines both. They also addressed information loss issues.

Liu et al. [29] introduced a WAVE-Net architecture that uses a denoising module to acquire microstructures and fuse contexts. The model addressed semantic loss and limited receptive field issues. Yi et al. [30] proposed MRANet that uses a feature fusion block to collect useful information, an attention block for better feature extraction, and a drop block for overfitting. Researchers provided the solution for segmenting capillary vessels and weak anti-noise interference ability.

Kumar et al. [31] proposed IterMiU-Net, which has iterative modules of Iternet. It is helpful to construct a lightweight convolution model with fewer parameters. The proposed IterMiU-Net architecture is computationally efficient and can be used for real-time segmentation applications. Liu et al. [32] introduced DAREs2Unet, which uses spatial attention and a dual attention model to address multiscale information and avoid unnecessary information.

Sun et al. [33] proposed SDAU-Net, which uses a series of deformable convolution and dual attention modules. It detects small vessels and solves the uneven brightness of the background. Li et al. [34] use MAGF-Net, which uses a residual block, an attention-guided fusion block and addresses information loss issues.

2.3. Motivation

U-Net model and its variations have shown better performance when compared with convolutional neural networks. However, most of the architectures fail to address the sensitivity rate problem, extracting multiscale feature extraction, information loss issues and segmenting thin vessels. Moreover, the computational complexity for most of the models is high.

Due to the above reasons, we are motivated to use drop block regularization in the residual block to prevent overfitting and added a multi-scale feature learning module that can extract features in multiple scales which is expected to capture intricate details. Also, adding effective attention in the decoder block can help to focus on more information regions and ignore the background information. This approach is expected to give a more powerful and context-aware learning model.

3. Research methodology

Accurate segmentation of blood vessels is a highly challenging endeavor. The proposed architecture uses Residual U-Net as its baseline since it has demonstrated notable performance in various image segmentation processes [19]. The skip connection in U-Net reduces the information loss issue but introduces a semantic gap issue when the low-level features are combined with the high-level features. Therefore, an atrous spatial pyramid pooling block is proposed instead of the skip connection.

3.1. Architecture

Each encoder block includes two residual blocks and max pooling operations. In the encoder phase, the extraction of features is carried out iteratively using pairs of consecutive residual blocks. Each pooling layer diminishes the spatial dimension. To expand the receptive field and capture broader contextual information, down sampling becomes essential. This downscaling operation is achieved by employing a pooling operation and a stride of 2 is used. During each down sampling process, the size

of the feature maps is reduced by half, while the number of channels is doubled. Starting from initial feature maps of size 48×48 , after two down sampling stages, they become 12×12 in dimensions. Further down sampling could lead to the loss of critical spatial information. Consequently, the network is designed with only two down sampling stages, featuring channel counts of 32, 64 and 128 in each stage.

The decoder block includes an up-sampling layer followed by a dual attention block and two residual blocks. For each up-sampling layer, bilinear interpolation is used to maintain the size of the feature map. The convolutional layers are configured with channel numbers 32, 64, and 128. The architecture of EAMR-Net is depicted in Figure 1.

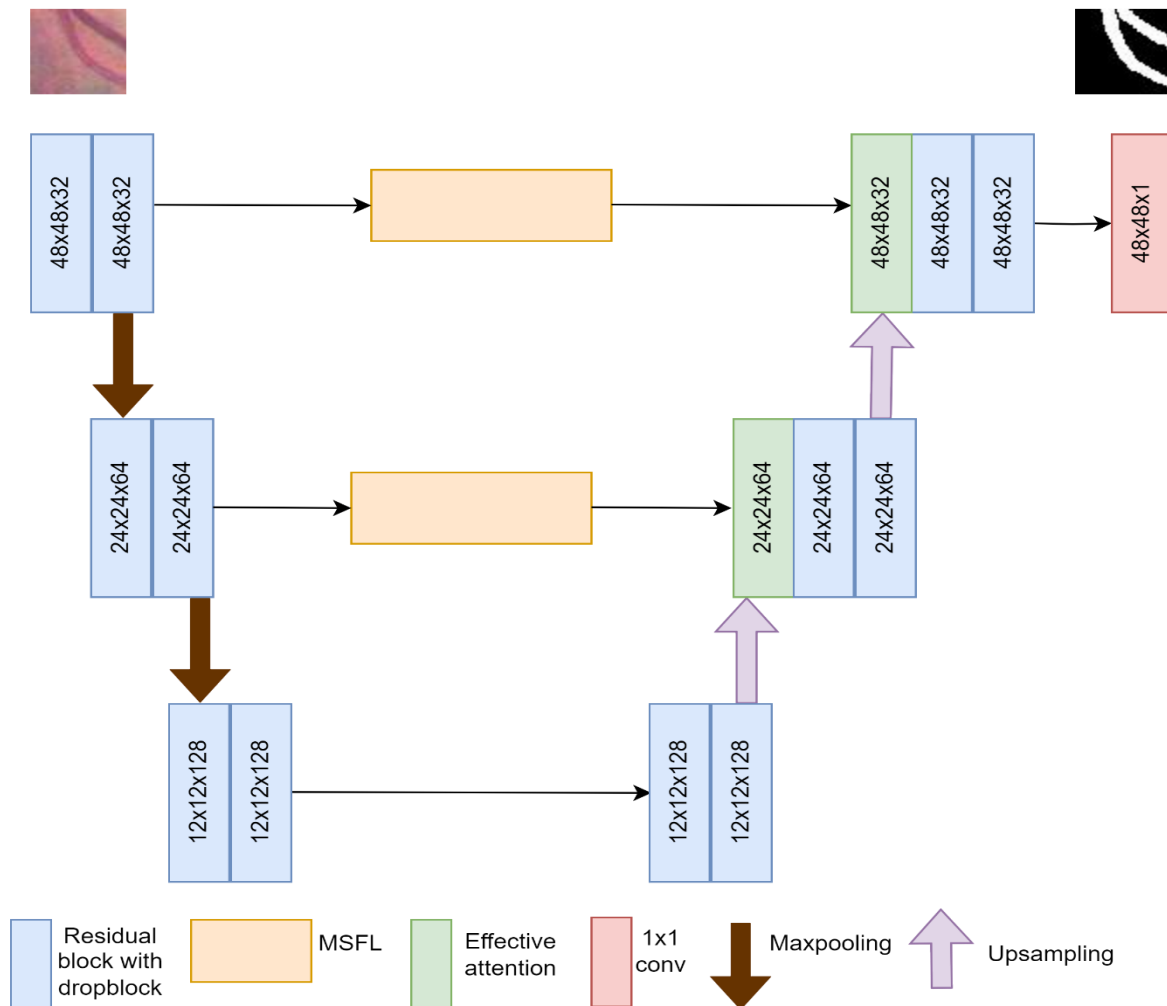


Figure 1. EAMR-Net architecture.

3.1.1. Residual block

The residual block [35] has shortcut connections that may skip a few layers and will perform identity mapping without increasing computational complexity. It facilitates faster convergence, reduces the impact of vanishing gradients, and reduces the risk of losing important information by encouraging feature reuse. A residual block consists of multiple layers that are arranged sequentially, allowing them to pass their output to a subsequent layer positioned deeper within the block. Residual

blocks address the issue of gradient vanishing by introducing skip-connections. The residual block contains two sequences of convolution, drop block, batch normalization (BN), and ReLU. The convolution operation is carried out on a skip connection. Finally, the residual block is added with a skip connection, which does identity mapping as shown in Figure 2. Drop block is a technique reminiscent of dropout, differing in that it excludes entire contiguous regions from a layer's feature map, as opposed to independently removing random individual units. Drop block retains the spatial information and helps to alleviate issues like overfitting and vanishing gradients. In the residual module, it is proven that placing the drop block [36] after the convolution block improves accuracy.

Let x_r represent the input of the residual block, $S(x_r)$ represent the skip connection path output and $R(x_r)$ represent the residual path output, Then the output of residual block y_r is obtained as

$$y_r = R(R(x_r)) \oplus S(x_r). \quad (1)$$

The residual path output $R(x_r)$ is represented as

$$R(x_r) = \text{conv}(\text{Dropblock}(\text{BN}(\text{ReLU}(x_r)))). \quad (2)$$

The skip connection path output $S(x_r)$ is represented as

$$S(x_r) = \text{conv}(x_r). \quad (3)$$

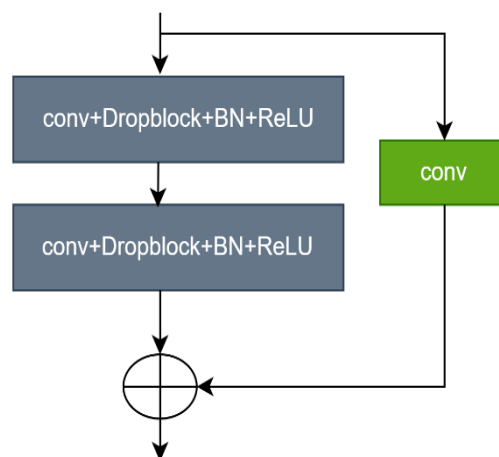


Figure 2. Residual block.

3.1.2. MSFL module

The MSFL module [37,38] employs multiple atrous convolutions to find deep features parallelly. To extract multiscale features [39] and large receptive fields [40], these convolution layers have different sampling (dilation) rates. This is a better approach for controlling the field of view without increasing parameters and computational time.

The input feature map is passed to the 1×1 convolution layer, followed by three atrous 3×3 convolution layers with different dilation rates (6, 12 and 18). In addition, we add average pooling and upsampling blocks to obtain image-level features. Finally, the output of these parallel branches is

combined using the concatenation operator. Then, 1x1 convolution is applied to the final feature map by fusing the feature maps of MSFL as depicted in Figure 3. The major shortcoming of MSFL is that it will treat all feature scales equally [41]. To tackle this issue, an attention block is introduced after the MSFL blocks.

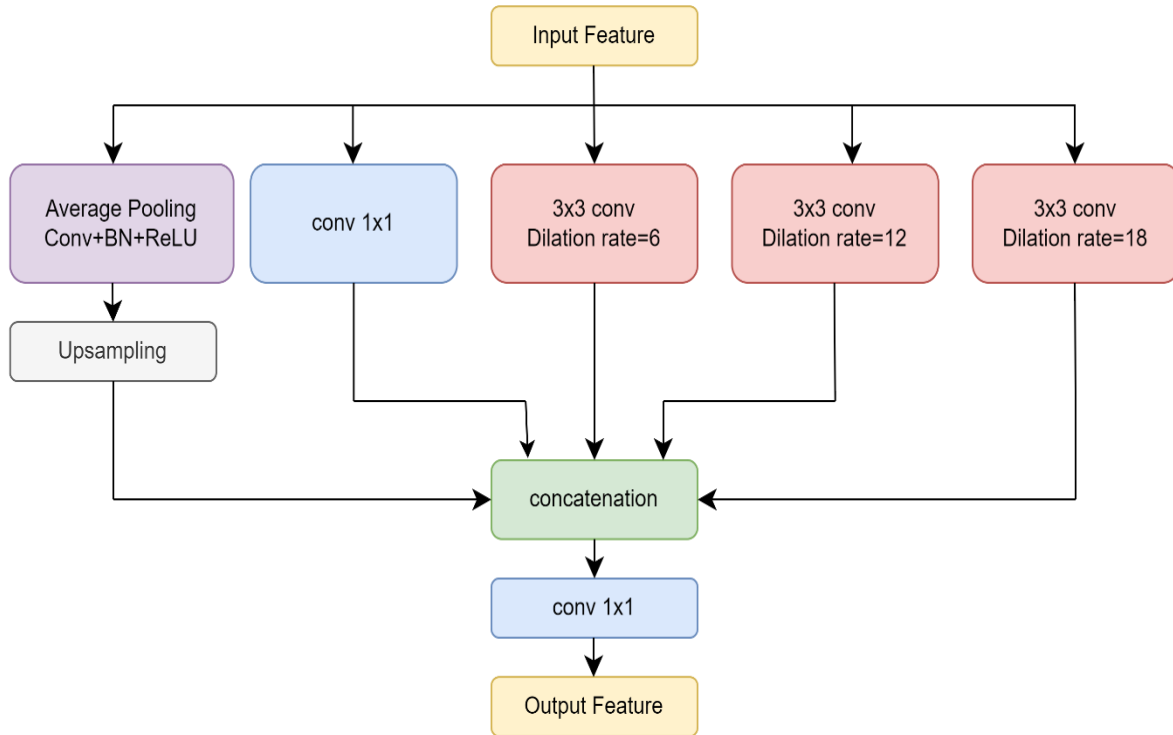


Figure 3. Multi-scale feature learning (MSFL).

3.1.3. Effective attention block

The effective attention block (Figure 4) incorporates both spatial and effective channel attention. The spatial attention block assigns higher weights to important features and diminishes the weights of redundant features, thereby enabling effective learning [42]. To the input feature map (F), both max pooling and average pooling operations are performed, followed by a 7×7 convolution and a sigmoid operation. This attention map obtained gives more precise spatial information.

$$A_{spatial}(F) = \sigma(\text{conv}_{7 \times 7}(\text{maxpool}(F); \text{avgpool}(F))), \quad (4)$$

where σ represents a sigmoid activation function.

The spatial attention map is multiplied by the input feature to produce the output feature.

$$F_{spatial} = A_{spatial}(F) \otimes F, \quad (5)$$

where \otimes represents element-wise multiplication.

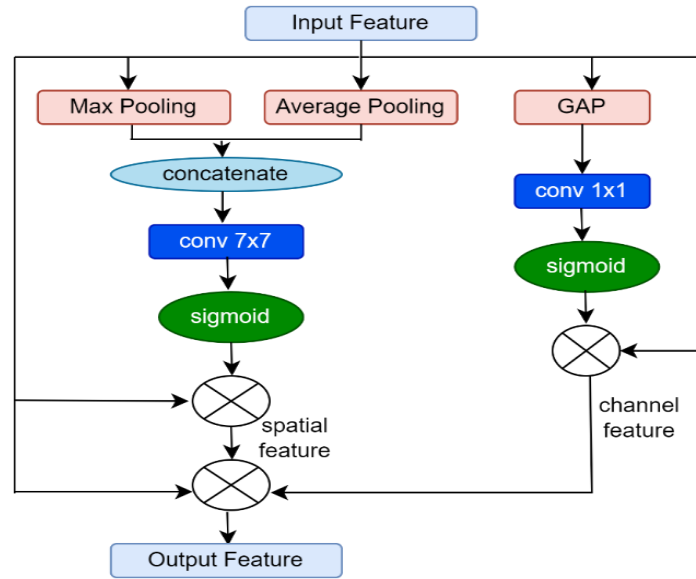


Figure 4. Effective attention module.

Effective channel attention involves dynamically adjusting the characteristics of individual channels in a feature representation to adaptively recalibrate their relevance. Squeeze and excitation (SE) [43] block use a dimensionality reduction mechanism to obtain channel-wise descriptors that diminish the complexity of the model but break the direct relationship between a channel and its corresponding weight. To mitigate this issue, ECANet [44] is employed for channel attention, which preserves the dimensionality to learn important features and does interactions between channels in a cross-channel manner.

Therefore, we apply Global Average Pooling (GAP) to capture global information by taking average values across spatial dimensions. To obtain the channel descriptor space, forward convolution is applied with ReLU activation. Then, channel attention is done by fully sigmoid activation, which produces a channel attention weight that describes the significance of each channel relative to the other. Finally, recalibration is done to rescale each channel's activation adaptively.

$$A_{channel}(F) = \sigma(\text{conv}_{k \times k}(\text{GAP}(F))), \quad (6)$$

where k represents the size of the kernel of 1D convolution which is determined adaptively and is proportional to the dimension of the channel.

The new channel feature maps are procured by reassigning weights:

$$F_{channel} = A_{channel}(F) \otimes F, \quad (7)$$

where \otimes represents channel-wise multiplication.

The final output feature of effective attention is obtained by multiplying the channel feature with the spatial feature.

$$F_{output} = F_{spatial} \otimes F_{channel}. \quad (8)$$

The EAMR-Net architecture summary is tabulated in Table 1. The table displays the name of the layer and image size of each block used in the architecture.

Table 1. Summary of EAMR-Net architecture.

S.No.	Block name	Layer Name	Image size
1		Input $48 \times 48 \times 3$	$48 \times 48 \times 3$
2	Residual block 1	Conv, DB, BN, ReLU \oplus conv	$48 \times 48 \times 32$
3	Residual block 2	Conv, DB, BN, ReLU \oplus conv	$48 \times 48 \times 32$
4	MaxPooling 1		$24 \times 24 \times 32$
5	Residual block 3	Conv, DB, BN, ReLU \oplus conv	$24 \times 24 \times 64$
6	Residual block 4	Conv, DB, BN, ReLU \oplus conv	$24 \times 24 \times 64$
7	MaxPooling 2		$12 \times 12 \times 64$
8	Residual block 5	Conv, DB, BN, ReLU \oplus conv	$12 \times 12 \times 128$
9	Residual block 6	Conv, DB, BN, ReLU \oplus conv	$12 \times 12 \times 128$
10	Residual block 7		$12 \times 12 \times 128$
11	Residual block 8		$12 \times 12 \times 128$
12	UpSampling 1		$24 \times 24 \times 128$
13	ASPP 1		$24 \times 24 \times 64$
14	Concatenate (ASPP1 + Up1)		$24 \times 24 \times 192$
15	Effective Attention		$24 \times 24 \times 192$
16	Residual block 9	Conv, DB, BN, ReLU \oplus conv	$24 \times 24 \times 64$
17	Residual block 10	Conv, DB, BN, ReLU \oplus conv	$24 \times 24 \times 64$
18	UpSampling 2		$48 \times 48 \times 64$
19	ASPP 2		$48 \times 48 \times 32$
20	Concatenate (ASPP2 + Up2)		$48 \times 48 \times 96$
21	Effective Attention		$48 \times 48 \times 96$
22	Residual block 11	Conv, DB, BN, ReLU \oplus conv	$48 \times 48 \times 32$
23	Residual block 12	Conv, DB, BN, ReLU \oplus conv	$48 \times 48 \times 32$
24	Conv 2D		$48 \times 48 \times 1$

3.2. Experimental setup

3.2.1. Datasets used

1). DRIVE: The DRIVE (<https://drive.grand-challenge.org/>) dataset consists of 40 images collected from 400 patients with a manual annotation equally split for training and testing. Out of the total dataset, consisting of 40 images, 33 are categorized as normal cases, while the remaining 7 exhibit indications of diabetic retinopathy. Each image is of 565×584 resolution. The images are captured with a 3CCD camera with a 45° field of view (FOV). The images are in TIFF format and the manual annotations are stored as GIF files.

2). CHASE_DB1: CHASE_DB1 (<https://blogs.kingston.ac.uk/retinal/chasedb1/>) has 28 images captured from 14 children with 999×960 resolution. The images are captured with a Nidek NM-200-D fundus camera, 30-degree FOV. Since images need to be categorized as train and test, the initial set of 20 images will be utilized for training, while the remaining images will be reserved for testing. Two manual annotations are given. We have considered the first manual labelling as ground truth. The images are in JPEG format and manually labelled images are in PNG format.

3). STARE: The STARE (<https://cecas.clemson.edu/~ahoover/stare/>) dataset contains 20 color

fundus images and a manually labeled segmentation mask. Out of a total of 20 images, 10 exhibit retinal abnormalities associated with diabetic retinopathy, while the remaining 10 images show no indications of this condition. Each image has a resolution of 700×605 . The images are captured with Top con TRV-50 fundus camera 35-degree FOV. Both the image and labelled image are in JPEG format. In the training phase, the initial set of 15 images is utilized, while during the testing phase, the remaining set of 5 images is employed for evaluation.

3.2.2. Preprocessing

Preprocessing is a fundamental aspect of image processing that significantly improves image quality, reduces noise, enhances contrast, standardizes image properties, and extracts pertinent features. These preprocessing steps are critical for ensuring precise and dependable analysis, interpretation, and decision-making across diverse fields like computer vision, medical imaging, remote sensing, and other relevant domains. During the preprocessing phase, techniques such as normalization, contrast adjustment (CLAHE), and gamma transformation are carried out. Figure 5 depicts the result of a sample fundus image after each preprocessing stage.

To alleviate the problem of overfitting due to the limited dataset, the images are divided into 48×48 patches using a random cropping technique. Image resolution and size of the input patch influence the level of detail the model can capture. Smaller patch sizes can capture fine details. Moreover, the computational resources and memory constraints have a greater influence in determining the image size. Then, we applied data augmentation techniques such as horizontal flipping, vertical flipping, elastic transform, grid distortion, rotation at an angle of 30° , 60° , random brightness, and random contrast.



(a) Original image.

(b) Normalization.

(c) CLAHE.

(d) Gamma correction.

Figure 5. Preprocessing stages of a fundus image.

3.2.3. Evaluation metrics

For quantitative analysis of segmentation results, several metrics used are Accuracy (ACC), Specificity (SPEC), Sensitivity (SEN), F1 score (F1), and Area under Receiver Operating Characteristic (AUC) curve.

ACC: It tells us how exactly the vessels and non-vessels are predicted.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

SEN: It specifies the proportion of correctly predicted blood vessels to actual blood vessels.

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

SPEC: It represents the ratio between the accurately detected non-blood vessels and the total number of non-blood vessels.

$$\text{SPEC} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (11)$$

F1 score (F1): It considers both the ability to correctly identify blood vessels (precision) and its ability to capture all actual blood vessels (recall).

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (12)$$

AUC: It tells how well the blood vessels and non-blood vessel segmentation is done.

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

3.3. Implementation details

The algorithm utilizes the Keras library from the TensorFlow backend and is executed on an NVIDIA RTX3060, equipped with 12 GB of memory. Pre-trained model is not used in our experimentation. The initialization of the algorithm is performed using Glorot uniform weights.

The algorithm employs the Adam optimizer with an initial learning rate of 0.001. It utilizes a binary cross-entropy loss function for training. The number of epochs used during the training process is 100. Due to limited memory, a batch size of 8 is used.

3.3.1. Loss function

The binary cross-entropy loss function quantifies how well a model's predicted probabilities align with the actual binary labels (0 or 1) for each example, particularly focusing on cases where the true label is 1.

Mathematically, the binary cross-entropy loss function is defined as:

$$\text{Loss}_{\text{BCE}} = -\frac{1}{N} [\sum_{i=1}^N g t_i * \log p_i + (1 - g t_i) * \log (1 - p_i)], \quad (13)$$

where, Loss_{BCE} is the loss function, N is the total number of pixels, $g t_i$ is the true binary label (ground truth), and p_i is the predicted probability that the example belongs to class 1.

3.3.2. Network hyperparameter analysis

The hyperparameters such as architecture depth, loss function, learning rate, and optimizer play a crucial role in improving the accuracy of segmentation. In the research studies, Adam optimizer and BCE loss function are employed. If the model is down-sampled further from 12×12 , there is a loss of more spatial information. So, the layer depth is been fixed to three.

The segmentation outcomes are significantly influenced by the choice of learning rate. Adjusting the learning rate has a pronounced impact on the final segmentation results. Table 2 lists the evaluation models with different learning rate on the dataset DRIVE.

Table 2. Comparison results of EAMR-Net with different learning rate on DRIVE dataset.

Learning rate	ACC	SPEC	SEN
0.00001	0.9671	0.9878	0.7912
0.0001	0.9675	0.9877	0.799
0.001	0.9685	0.9872	0.8219
0.003	0.9534	0.9783	0.7729

When the learning rate is 0.001, our proposed model obtained higher accuracy and sensitivity. With a learning rate set at 0.003, the model exhibits poor performance. At learning rates of both 0.0001 and 0.00001, the model has demonstrated relatively modest performance.

As shown in Figure 6, during the training stage, there is a gradual improvement in training and validation accuracy as the epoch increases.

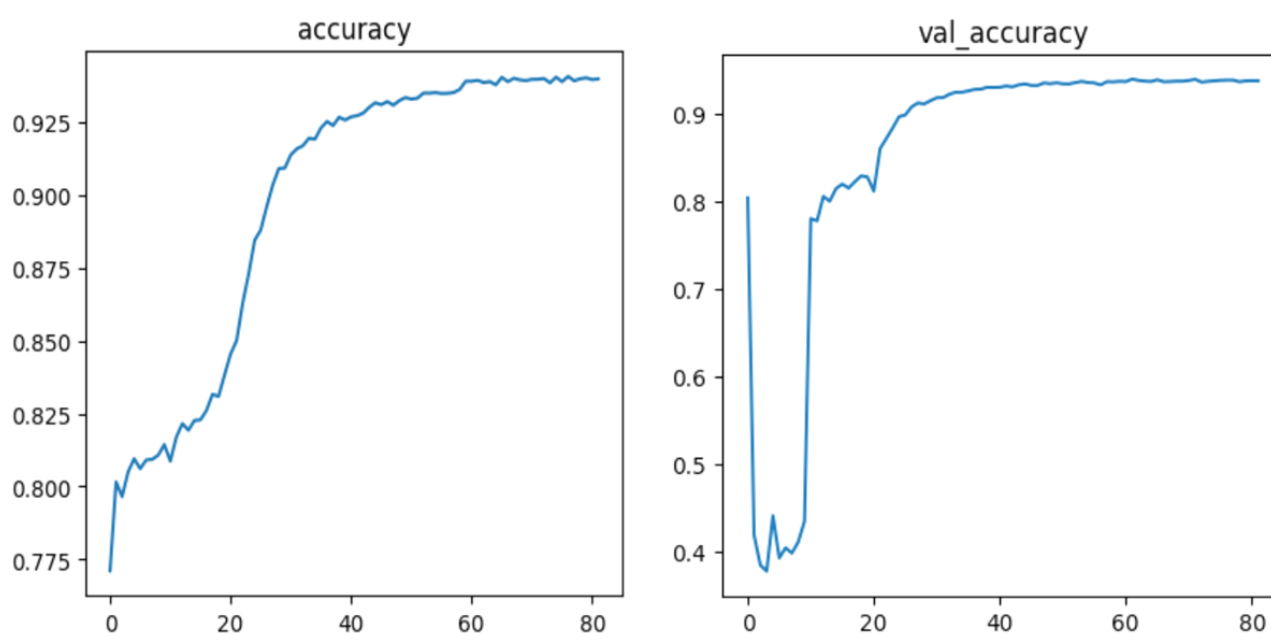


Figure 6. Training and validation accuracy curve for the DRIVE dataset.

4. Experimental results

4.1. Ablation studies

To track the model's effects, various ablation studies are conducted on three datasets by adding new blocks to an existing architecture. We used Residual U-Net as the baseline and compared the segmentation results with EA+ResU-Net and EAMR-Net. The proposed model has shown better performance when compared with the other models in ablation studies, as shown in Figure 7.

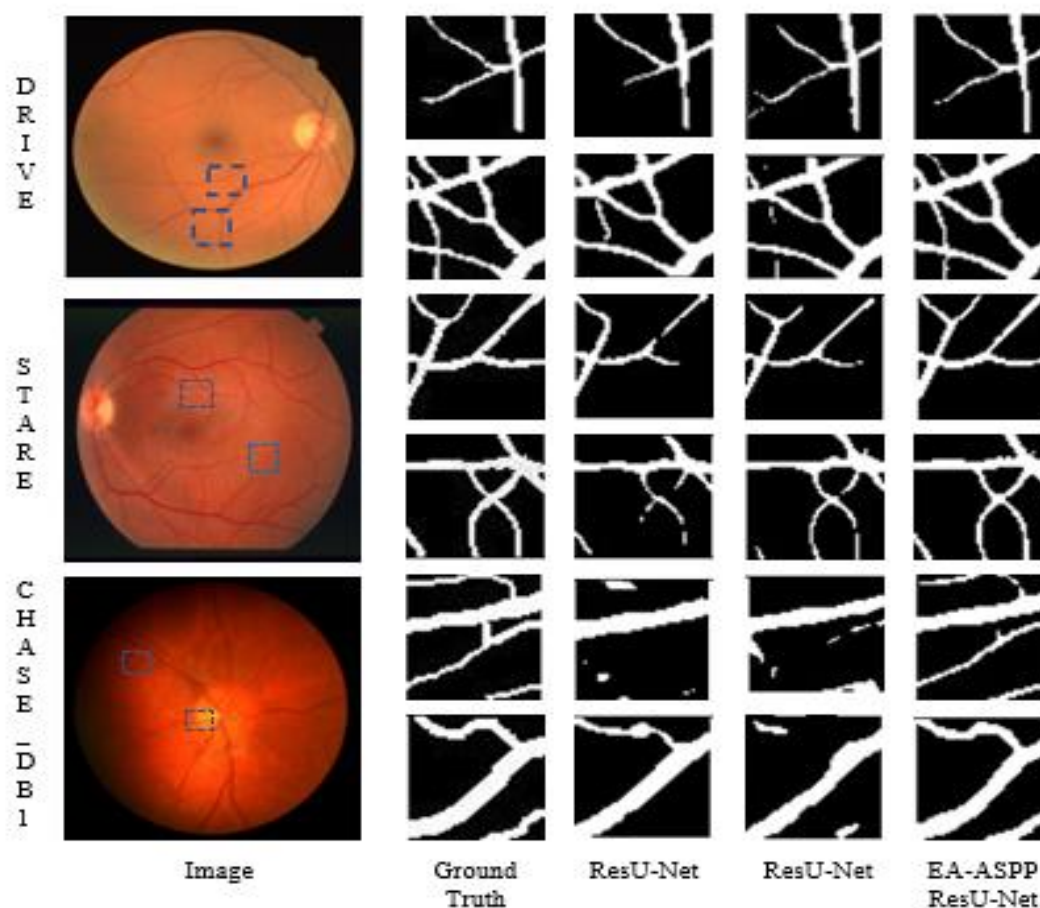


Figure 7. Segmentation results on different datasets.

We set the drop rate value as 0.1 and the block size as 7. When residual U-Net with a drop block is employed, we observed some discontinuous segmentation in overlapping areas of vessels. Then, an efficient dual attention block is added to the residual U-Net which shows improvement in the detection of vessels in overlapping areas. However, due to the low contrast in the background, it was unable to distinguish a few blood vessels. So, instead of using a skip connection, an MSFL block is added, and we found that the performance is superior to other models in detecting vessels in low contrast as well as overlapping areas.

The ablation study results for different datasets are tabulated in Table 3. In the DRIVE dataset, our model has shown a 1.98% increase in sensitivity, which reports that there is an improvement in blood vessel detection. Specificity, accuracy, and the F1 score have also shown improvement. However, in DRIVE and CHASE_DB1, there is a slight decrease in specificity when an effective attention block is added. This behavior is due to poor contrast in a few images in the dataset. In the STARE dataset, sensitivity has improved by 1.45%. Also, there is an improvement in results in ACC, SPEC, and AUC curves. The CHASE_DB1 dataset has also shown progress in detecting blood vessels. There is a 1.26% increase in sensitivity for the proposed model.

Table 3. Ablation study results on different datasets.

Dataset	Method	ACC	SPEC	SEN	AUC	F1
DRIVE	ResU-Net	0.9501	0.9847	0.8010	0.9780	0.8269
	EA+ResU-Net	0.9523	0.9840	0.8095	0.9821	0.8286
	EAMR-Net	0.9621	0.9853	0.8293	0.9859	0.8382
STARE	ResU-Net	0.9520	0.9864	0.7982	0.98	0.8251
	EA+ResU-Net	0.9665	0.9866	0.8006	0.9802	0.8394
	EAMR-Net	0.9713	0.9932	0.8151	0.9881	0.8254
CHASE_DB1	ResU-Net	0.9504	0.9720	0.7413	0.9721	0.8358
	EA+ResU-Net	0.9662	0.9659	0.7958	0.9873	0.8340
	EAMR-Net	0.9665	0.9848	0.8084	0.9875	0.8423

4.2. Comparative analysis with other models

We compared our model with some recent models like R2U-Net [45], DCU-Net [22], CRAU-Net [25], ResDOU-Net [26], GDF-Net [28], Wave-Net [29], MAGF-Net [34], IterMiU-Net [31] and SDAU-Net [33]. The EAMR-Net model has shown better performance in specificity, sensitivity, and the AUC compared with state-of-the-art models.

The findings of the experiment on DRIVE, STARE, and CHASE_DB1 are tabulated in Table 4. Our model has shown better performance on SPEC and AUC. Its sensitivity is the same as the SDAU-Net model. Our model reaches nearly 0.9853 on SPEC and 0.9861 on AUC. The SPEC is 0.0050 higher than R2U-Net and AUC is 0.0002 higher than ResDOU-Net.

On the STARE dataset, the proposed model has achieved 0.9713 ACC, 0.9932 SPEC, 0.8151 SEN, 0.9881 AUC, and 0.8254 F1 scores. It has achieved a 0.0004 increase in specificity when compared to RV-Net.

In CHASE_DB1, experimental results show that the ACC, SPEC, SEN, AUC, and F1 score are 0.9665, 0.9848, 0.8084, 0.9875 and 0.8423 respectively. In comparison with other methods, it has achieved a specificity of 0.0007 higher than DCU-Net.

The qualitative analysis is done on the three benchmark datasets. Figure 8 presents a visual comparison of our model, EAMR-Net with other state-of-art models. We have chosen ResU-Net, AttResU-Net, and RCARU-Net for qualitative analysis. Upon examining the detailed regions, it becomes evident that ResU-Net, AttResU-Net, and RCARU-Net [46] possess the capability to extract the retinal vessels from the original image. However, a closer inspection of the ground truth images reveals noticeable disconnections and mis-segmentations in the capillary region. In this comparison, EAMR-Net stands out for its proficiency in detecting capillary vessels, and the demonstration of excellent connectivity between vessels. It is observed that the area shown in the red rectangular box shows the segmentation of thin vessels. Compared to other techniques, the EAMR-Net method has performed better.

Table 4. Comparison of the DRIVE, STARE, and CHASE_DB1 datasets.

Dataset	Author	Year	Method	ACC	SPEC	SEN	AUC	F1
DRIVE	Alom et al. [45]	2018	R2U-Net	0.9556	0.9813	0.7792	0.9784	0.8171
	Yang et al. [22]	2022	DCU-Net	0.9568	0.9780	0.8115	0.9810	0.8272
	Dong et al. [25]	2022	CRAU-Net	0.9586	-	0.7954	0.9830	0.8302
	Liu et al.[26]	2022	ResDOU-Net	0.9561	0.9791	0.7985	-	0.8229
	Li et al. [28]	2022	GDF-Net	0.9622	0.9852	0.8291	0.9859	0.8302
	Liu et al. [29]	2023	Wave-Net	0.9561	0.9764	0.8164	-	0.8254
	Li et al.[34]	2022	MAGF-Net	0.9578	0.9783	0.8262	0.9819	0.8307
	Kumar et al.[31]	2023	IterMiU-Net	0.9568	0.9789	0.8053	0.9810	0.8262
	Sun et al.[33]	2023	SDAU-Net	0.9675	0.9807	0.8293	0.9832	-
	Proposed method	2023	EAMR-Net	0.9621	0.9853	0.8293	0.9861	0.8382
STARE	Alom et al. [45]	2018	R2Unet	0.9712	0.9862	0.8298	0.9914	0.8475
	Boudegga et al.[21]	2021	RVNet,LCM	0.9796	0.9928	0.806	-	-
	Wang et al.[23]	2021	Sa Unet	0.9611	0.9866	0.8006	0.9902	0.8394
	Liu et al. [26]	2022	ResDO Unet	0.9567	0.9792	0.7963	-	0.8172
	Liu et al.[29]	2023	Wavenet	0.9641	0.9836	0.7902	-	0.8140
	Kumar et al.[31]	2023	IterMiUnet	0.9649	0.9831	0.8069	0.9852	0.8231
	Proposed method	2023	EAMR-Net	0.9713	0.9932	0.8151	0.9881	0.8254
CHASE DB1	Alom et al. [45]	2018	R2U-Net	0.9634	0.9820	0.7756	0.9815	0.7928
	Wang et al.[23]	2021	SAU-Net	0.9662	0.9659	0.7958	0.9873	0.8340
	Yang et al. [22]	2022	DCU-Net	0.9664	0.9841	0.8075	0.9872	0.8278
	Dong et al. [25]	2022	CRAU-Net	0.9659	-	0.8259	0.9864	0.8156
	Liu et al. [26]	2022	ResDOU-Net	0.9672	0.9794	0.8020	-	0.8236
	Liu et al. [29]	2023	Wave-Net	0.9664	0.9821	0.7284	-	0.8349
	Kumar et al. [31]	2023	IterMiU-Net	0.9591	0.9704	0.8443	0.9812	0.7875
	Sun et al. [33]	2023	SDAU-Net	0.9732	0.9825	0.8321	0.9858	-
Proposed method	2023	EAMR-Net	0.9665	0.9848	0.8084	0.9875	0.8423	

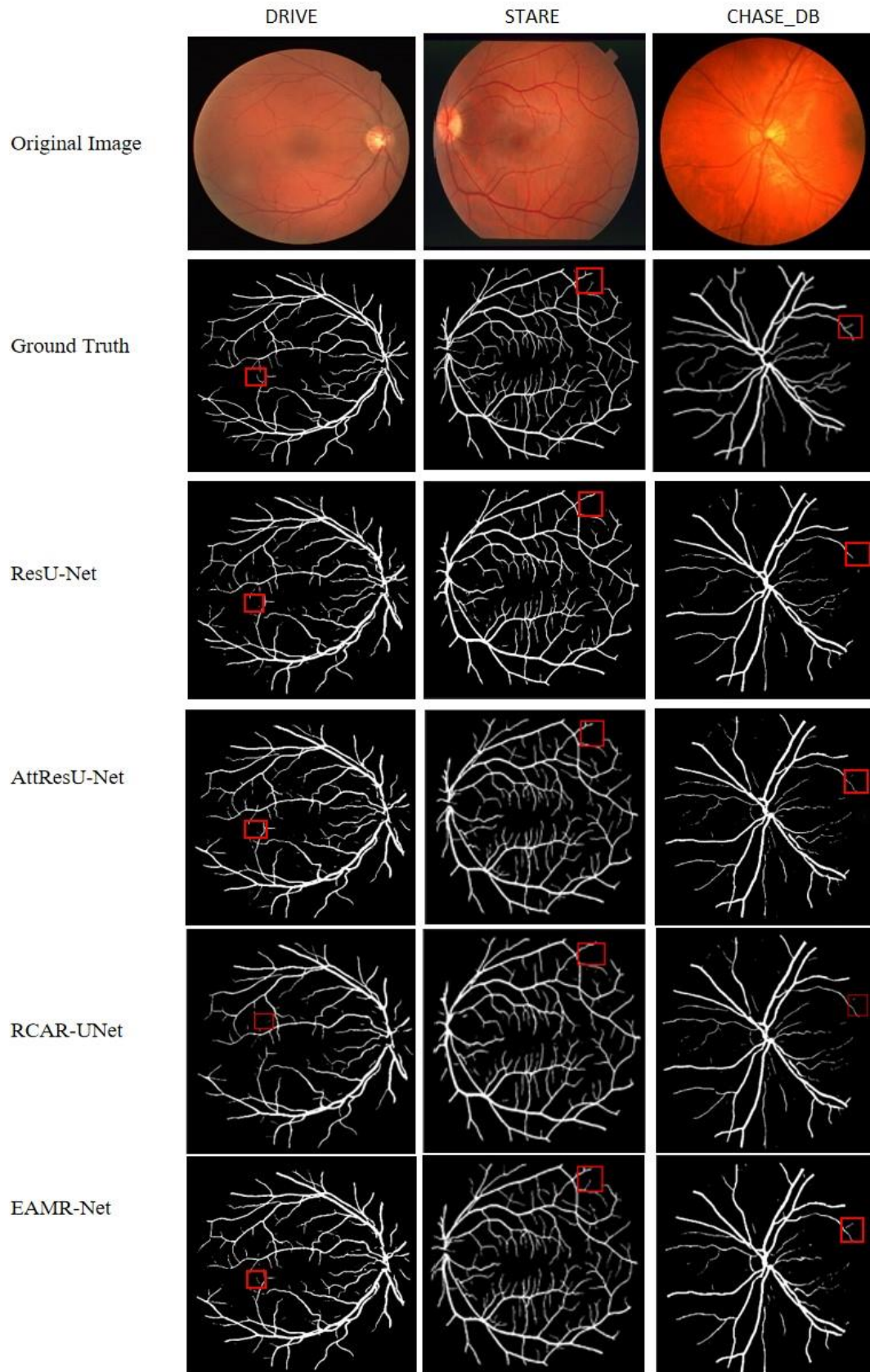


Figure 8. Segmentation results of three datasets using various approaches.

5. Conclusions

A novel architecture with effective attention and multiscale learning is proposed to detect blood vessels precisely. To prevent overfitting, a drop block is appended after every convolution block, and residual U-net is used to prevent vanishing gradients. To preserve the dimensionality of spatial and channel features, a dual attention block is used. To avoid the semantic gap issue, an MSFL block is added. The effectiveness of the suggested approach is compared with the existing method in three datasets namely, DRIVE, STARE and CHASE_DB1, and has achieved better results. The sensitivity of 0.8293, 0.8151, and 0.8084 is attained on the DRIVE, STARE and CHASE_DB1 datasets respectively. In future research and development, one exciting avenue to explore is the integration of vision transformers (ViTs) into model design. Also, Generative Adversarial Networks (GANs) can be employed for enhancing image augmentation techniques.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. Centers for Disease Control and Prevention, *National diabetes statistics report, 2020: estimates of diabetes and its burden in the United States*, CDC, 2020. Available from: <https://stacks.cdc.gov/view/cdc/85309>
2. *Eye Complications*, ADA, 2021. Available from: <https://diabetes.org/about-diabetes/complications/eye-complication>.
3. M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, et al., Blood vessel segmentation methodologies in retinal images—a survey, *Comput. Methods Programs Biomed.*, **108** (2012), 407–433. <https://doi.org/10.1016/j.cmpb.2012.03.009>
4. S. Dash, S. Verma, Kavita, S. Bevinakoppa, M. Wozniak, J. Shafi, et al., Guidance image-based enhanced matched filter with modified thresholding for blood vessel extraction, *Symmetry*, **14** (2022), 194. <https://doi.org/10.3390/sym14020194>
5. S. Chatterjee, A. Suman, R. Gaurav, S. Banerjee, A. K. Singh, B. K. Ghosh, et al., Retinal blood vessel segmentation using edge detection method, *J. Phys. Conf. Ser.*, **1717** (2021), 012008. <https://doi.org/10.1088/1742-6596/1717/1/012008>
6. P. Kuppusamy, M. M. Basha, C. L. Hung, Retinal blood vessel segmentation using random forest with Gabor and Canny edge features, in *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, Villupuram, India, (2022), 1–4. <https://doi.org/10.1109/ICSTSN53084.2022.9761339>

7. S. Roychowdhury, D. D. Koozekanani, K. K. Parhi, Blood vessel segmentation of fundus images by major vessel extraction and subimage classification, *IEEE J. Biomed. Health Inf.*, **19** (2014) 1118–1128. <https://doi.org/10.1109/JBHI.2014.2335617>
8. E. Chakour, Y. Mrad, A. Mansouri, Y. Elloumi, M. H. Bedoui, I. B. Andaloussi, et al., Blood vessel segmentation of retinal fundus images using dynamic preprocessing and mathematical morphology, in *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)*, Istanbul, Turkey, (2022), 1473–1478. <https://doi.org/10.1109/CoDIT55151.2022.9804004>
9. P. R. Wankhede, K. B. Khanchandani, Retinal blood vessel segmentation using graph cut analysis, in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, Pune, India, (2015), 1429–1432. <https://doi.org/10.1109/IIC.2015.7150973>
10. M. R. K. Mookiah, S. Hogg, T. J. MacGillivray, V. Prathiba, R. Pradeepa, V. Mohan, et al., A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification, *Med. Image Anal.*, **68** (2021), 101905. <https://doi.org/10.1016/j.media.2020.101905>
11. O. O. Sule, A survey of deep learning for retinal blood vessel segmentation methods: Taxonomy, trends, challenges and future directions, *IEEE Access*, **10** (2022), 38202–38236. <https://doi.org/10.1109/ACCESS.2022.3163247>
12. T. J. Jebaseeli, C. A. D. Durai, J. D. Peter, Retinal blood vessel segmentation from diabetic retinopathy images using tandem PCNN model and deep learning based SVM, *Optik*, **199** (2019), 163328. <https://doi.org/10.1016/j.ijleo.2019.163328>
13. X. Yang, Z. Li, Y. Guo, D. Zhou, Retinal vessel segmentation based on an improved deep forest, *Int. J. Imaging Syst. Technol.*, **31** (2021), 1792–1802. <https://doi.org/10.1002/ima.22610>
14. D. Yang, G. Liu, M. Ren, B. Xu, J. Wang, A multi-scale feature fusion method based on U-net for retinal vessel segmentation, *Entropy*, **22** (2020), 811. <https://doi.org/10.3390/E22080811>
15. M. Padmapriya, S. Pasupathy, V. Punitha, Early diagnosis of diabetic retinopathy using unsupervised learning, *Soft Comput.*, **27** (2023), 9093–9104. <https://doi.org/10.1007/s00500-023-08418-z>
16. N. Muzammil, S. A. A. Shah, A. Shahzad, M. A. Khan, R. M. Ghoniem, Multifilters-based unsupervised method for retinal blood vessel segmentation, *Appl. Sci.*, **12** (2022), 6393. <https://doi.org/10.3390/app12136393>
17. Z. Qaiser, W. Ahmad, M. Y. Umair, Z. Mahmood, Unsupervised vessel segmentation method in retinal images, in *2022 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, (2022), 65–70. <https://doi.org/10.1109/FIT57066.2022.00022>
18. K. Upadhyay, M. Agrawal, P. Vashist, Unsupervised multiscale retinal blood vessel segmentation using fundus images, *IET Image Proc.*, **14** (2020) 2616–2625. <https://doi.org/10.1049/iet-ipr.2019.0969>
19. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, Munich, Germany, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

20. T. Laibacher, S. Jalali, M2u-net: Effective and efficient retinal vessel segmentation for real-world applications, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, Long Beach, CA, USA, 2019. <https://doi.org/10.1109/CVPRW.2019.00020>
21. H. Boudegga, Y. Elloumi, M. Akil, M. H. Bedoui, R. Kachouri, A. B. Abdallah, Fast and efficient retinal blood vessel segmentation method based on deep learning network, *Comput. Med. Imaging Graphics*, **90** (2021). <https://doi.org/10.1016/j.compmedimag.2021.101902>
22. X. Yang, Z. Li, Y. Guo, D. Zhou, DCU-net: a deformable convolutional neural network based on cascade U-net for retinal vessel segmentation, *Multimed. Tools Appl.*, **81** (2022), 15593–15607. <https://doi.org/10.1007/s11042-022-12418-w>
23. H. Wang, G. Xu, X. Pan, Z. Liu, N. Tang, R. Lan, et al., Attention-inception-based U-Net for retinal vessel segmentation with advanced residual, *Comput. Electr. Eng.*, **98** (2022), 107670. <https://doi.org/10.1016/j.compeleceng.2021.107670>
24. X. Wang, X. Jiang, J. Ren, Blood vessel segmentation from fundus image by a cascade classification framework, *Pattern Recognit.*, **88** (2019), 331–341. <https://doi.org/10.1016/j.patcog.2018.11.030>
25. F. Dong, D. Wu, C. Guo, S. Zhang, B. Yang, X. Gong, CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation, *Comput. Biol. Med.*, **147** (2022), 105651. <https://doi.org/10.1016/j.compbiomed.2022.105651>
26. Y. Liu, J. Shen, L. Yang, G. Bian, H. Yu, ResDO-UNet: A deep residual network for accurate retinal vessel segmentation from fundus images, *Biomed. Signal Process. Control*, **79** (2023), 104087. <https://doi.org/10.1016/j.bspc.2022.104087>
27. K. Ren, L. Chang, M. Wan, G. Gu, Q. Chen, An improved U-net based retinal vessel image segmentation method, *Heliyon*, **8** (2022), e11187. <https://doi.org/10.1016/j.heliyon.2022.e11187>
28. J. Li, G. Gao, L. Yang, Y. Liu, GDF-Net: A multi-task symmetrical network for retinal vessel segmentation, *Biomed. Signal Process. Control*, **81** (2023), 104426. <https://doi.org/10.1016/j.bspc.2022.104426>
29. Y. Liu, J. Shen, L. Yang, H. Yu, G. Bian, Wave-Net: A lightweight deep network for retinal vessel segmentation from fundus images, *Comput. Biol. Med.*, **152** (2023), 106341. <https://doi.org/10.1016/j.compbiomed.2022.106341>
30. S. Yi, Y. Wei, G. Zhang, T. Wang, F. She, X. Yang, Segmentation of retinal vessels based on MRANet, *Heliyon*, **9** (2023). <https://doi.org/10.1016/j.heliyon.2022.e12361>
31. A. Kumar, R. K. Agrawal, L. Joseph, IterMiUnet: A lightweight architecture for automatic blood vessel segmentation, *Multimedia Tools Appl.*, **82** (2023), 1–25. <https://doi.org/10.1007/s11042-023-15433-7>
32. R. Liu, T. Wang, X. Zhang, X. Zhou, DA-Res2UNet: Explainable blood vessel segmentation from fundus images, *Alexandria Eng. J.*, **68** (2023) 539–549. <https://doi.org/10.1016/j.aej.2023.01.049>
33. K. Sun, Y. Chen, Y. Chao, J. Geng, Y. Chen, A retinal vessel segmentation method based improved U-Net model, *Biomed. Signal Process. Control*, **82** (2023), 104574. <https://doi.org/10.1016/j.bspc.2023.104574>
34. J. Li, G. Gao, Y. Liu, L. Yang, MAGF-Net: A multiscale attention-guided fusion network for retinal vessel segmentation, *Measurement*, **206** (2023), 112316. <https://doi.org/10.1016/j.measurement.2022.112316>

35. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, (2016), 770–778. <https://doi.org/10.1007/s11042-023-15433-7>
36. G. V. Ghiasi, T. Y. Lin, Q. Le, DropBlock: A regularization method for convolutional networks, *Adv. Neural Inf. Process. Syst.*, **31** (2018).
37. L. C. Chen, G. Papandreou, S. Member, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.*, **40** (2017), 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>
38. L. C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, preprint, arXiv:1706.05587. <https://doi.org/10.48550/arXiv.1706.05587>
39. F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, preprint arXiv:1511.07122. <https://doi.org/10.48550/arXiv.1511.07122>
40. R. Liu, F. Tao, X. Liu, J. Na, H. Leng, J. Wu, et al., RANet: a residual ASPP with attention framework for semantic segmentation of high-resolution remote sensing images, *Remote Sens.*, **14** (2022), 3109. <https://doi.org/10.3390/rs14133109>
41. Y. Qiu, Y. Liu, Y. Chen, J. Zhang, J. Zhu, J. Xu, A2SPPNet: Attentive atrous spatial pyramid pooling network for salient object detection, *IEEE Trans. Multimedia*, **25** (2023) 1991–2006. <https://doi.org/10.1109/TMM.2022.3141933>
42. G. Cao, S. Luo, Multimodal perception for dexterous manipulation, in *Tactile Sensing, Skill Learning, and Robotic Dexterous Manipulation*, Academic Press, (2022), 45–58. <https://doi.org/10.1016/B978-0-32-390445-2.00010-6>
43. J. Hu, Squeeze-and-Excitation networks, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, (2018), 7132–7141.
44. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Seattle, WA, USA, (2020), 11534–11542. <https://doi.org/10.1109/CVPR42600.2020.01155>
45. M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, V. K. Asari, Recurrent residual U-Net for medical image segmentation, *J. Med. Imaging*, **6** (2019), 014006. <https://doi.org/10.1117/1.JMI.6.1.014006>
46. W. Ding, Y. Sun, J. Huang, H. Ju, C. Zhang, G. Yang, et al., RCAR-UNet: Retinal vessel segmentation network algorithm via novel rough attention mechanism, *Inf. Sci.*, **657** (2024), 120007. <https://doi.org/10.1016/j.ins.2023.120007>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)