IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 September 2023, accepted 14 September 2023, date of publication 19 September 2023,
date of current version 25 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3317176

RESEARCH ARTICLE

Modified Depthwise Parallel Attention UNet for Retinal Vessel Segmentation

K. RADHA AND YEPUGANTI KARUNA
School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
Corresponding author: Yepuganti Karuna (karun@vit.ac.in)

ABSTRACT Retinal fundus images contain highly informative geometrical features for detecting diabetic retinopathy (DR), including vessels, especially thin and low-contrast vessels, which are predominant features for accurately diagnosing diabetic retinopathy. Automatic segmentation methods have been developed based on deep convolutional neural networks to replace manual labeling. These methods have shown acceptable performance in fundus vessel segmentation. The UNet model is a well-known architecture of deep neural networks often used for vessel segmentation tasks and has achieved significant performance. However, segmentation tasks remain challenging due to multiple convolutions, down-sampling operations, and inadequate feature fusion in the encoder-decoder architecture. Also, traditional convolution increases the number of multiplications while performing convolution operations. These challenges lead to the loss of information related to thin and low-contrast vessels, eventually affecting the segmentation performance. To tackle this issue, we propose incorporating depthwise parallel attention in the existing UNet framework (DPA-UNet) to achieve accurate vessel segmentation. This approach entails the integration of a depthwise convolution block in the downsampling path and a parallel attention mechanism in the upsampling path of UNet. The primary benefit of depthwise convolution and global information embedding (GIE) is the ability to capture intricate information characteristics across channels. This helps to minimize the information degradation caused by conventional convolution and downsampling techniques. A parallel attention network is proposed in the upsampling path of the existing UNet to optimize the channel and spatial information acquired from the encoder-decoder. Extensive experiments are conducted on three publicly available datasets, namely DRIVE, STARE, and CHASE_DB1, to validate the performance of the proposed model. The findings indicate that the UNET model with depthwise parallel attention achieved a competitive performance with fewer network parameters in segmenting retinal vessels.

INDEX TERMS Diabetic retinopathy, vessel segmentation, depth-wise separable convolution, UNet, attention mechanism, deep learning.

I. INTRODUCTION

Diabetes is a widespread disease worldwide caused by insulin resistance or insufficient insulin production. Uncontrolled continuing diabetes can cause damage to the retina. As the disease progresses, the blood vessels inside the retina start to bulge, leak, and close, eventually developing vessel-like regions within the eye. The duration of diabetes will cause risk factors for developing DR, and monitoring blood glucose levels has a high impact on controlling the progression of DR [1]. Fundus imaging is a non-invasive popular imaging modality for diagnosing diabetic retinopathy as well as it is used in various security applications [2], [3], [4]. The segmentation of retinal blood vessels is a fundamental step in diagnosing DR. This process involves depicting various morphological characteristics of the retinal blood vessels, including bifurcation, branching points, length and width of the segmented vascular tree. These features act as a diagnostic source for identifying diabetic retinopathy. An ophthalmologist can mark these retinal vessels to diagnose DR, but it is time-consuming, and we need a skilled ophthalmologist. Automated systems have been developed

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

to assist ophthalmologists in DR detection due to the surge in diabetes cases and the lack of trained ophthalmologists. The Uniqueness of each retinal vascular structure discloses the significance of blood vessel segmentation [5]. Various factors, including noise, low contrast, asymmetrical shape, and multiscale properties of the vessels, may degrade the vessel segmentation results. Therefore, it is necessary to have a standard and automated segmentation of retinal vessels to diagnose diabetic retinopathy. However, it is a continuing challenge due to the following characteristics of retinal vessels,

(1) Low resolution, high noise, and poor contrast of retinal fundus images increase the complexity of the segmentation task.
(2) The retinal image comprises bifurcations, branching points, centreline reflex, and boundary points.
(3) The size and width of the vessels vary when they travel radially outward from the optic disc; likewise, intensity values differ significantly within the fundus image.

Diverse approaches have been proposed to address these problems, which can be broadly categorized into traditional and automated segmentation methodologies [6]. Traditional segmentation methods can be subdivided into matched filter [7], [8], [9] and morphology [10], [11], multiscale filtering [12], [13], and kernel-based [14], [15] segmentation methods. Matched filters perform feature extraction in retinal images by rotating the kernel in different directions, dimensions, and locations. These unsupervised methods require feature extractors and depend on additional preprocessing techniques. When the background of retinal images is even and simple, these extractors can perform excellently. However, retinal images have tremendous differences in illumination, and contrast might result in poor performance.

The increased computing resources of computers have encouraged the development of automatic retinal vessel segmentation techniques. It can achieve better results than unsupervised traditional segmentation techniques since they use annotated images. Automated retinal vessel segmentation techniques are supervised segmentation techniques that use convolutional neural network (CNN) algorithms, and they learn the features automatically. A vast number of methods practice convolutional neural networks to segment retinal vessels [16], [17], [18]. CNNs can be effective for retinal vessel segmentation, but they also have potential drawbacks that should be considered when designing a fundus retinal vessel segmentation pipeline. A limited vessel dataset for training the model can make it challenging to train the CNN. This challenge can be overcome by utilizing UNet effectively [19].

Retinal fundus images often have considerable resolution, contrast, and noise variability. The presence of variability in images can challenge Convolutional Neural Networks (CNNs) in accurately performing segmentation of regions of interest across fundus images. Researchers have experimented with CNN with cascaded convolutions [20], dilated convolutions [21], and cascading models [22], [23]. The utilization of dilated convolution presents certain advantages. The network's receptive field can be increased without an increase in parameters by incorporating an increased dilation rate. However, it aggregates the multiscale information from different receptive fields instead of contextual information. Contextual information regions exhibit a significant relationship with vessel data, particularly for small vessels. Dilated convolution is prone to overfitting when training the network with a limited dataset. Depthwise convolution gives an effective replacement for traditional convolution, it keeps the intricate contextual details of vessels. The utilization of the attention mechanism is a convincing computational framework for preserving long-range contextual information.

This paper is organized into six sections: Section I covers an introduction, Section II deals with a brief overview of the recent methods in the literature, and Section III presents a detailed description of the proposed work for segmentation of retinal vessels based on DPA-UNet. Dataset details are presented and discussed in Section IV. Section V covers simulation Results. Finally, the discussion and conclusion are presented in Sections VI and VII respectively.

II. RELATED WORKS

Previously, conventional image processing methods were employed by researchers for the segmentation of retinal vessels. These methods included thresholding segmentation [24], [25] and specific morphological operations [26]. Subsequently, various learning-based methodologies incorporating handcrafted features were employed to tackle this task [27]. Nonetheless, the scope of these handcrafted features was insufficient to generalize the varied characteristics inherent in complex fundus images, impeding their ability to accurately depict such features. Consequently, these methodologies were occasionally subject to erroneous interpretation, particularly in extreme cases such as low-contrast microvessels and lesion areas. The use of DNNs-based methods shows superior performance in handling complexity in fundus images as it learns features directly from the training data [28], [29]. There are many methods available for diagnosing diabetic retinopathy through retinal vessel segmentation [30].

Numerous studies have been conducted on retinal vessel segmentation utilizing the UNet architecture. One of the main challenges in diabetic retinopathy detection is segmenting vessel information with a limited dataset. It was developed by Olaf Ronneberger [19] and inspired by the performance of a fully connected neural network [31]. The resolution of fundus retinal images exhibits significant variability due to vessels with varying widths, an optic disc [32], a bright retinal spot, and the darkest region known as the macula. To address the issue of significant scale variation, the authors have proposed a three-part network to segment vessels. The vessel features are identified using an enhanced encoder network with a decoder network. The decoder network is a dual-decoder network, effectively segmenting thick and thin vessels. Segmentation output was produced by adaptive feature fusion, which helps to tackle the multiscale information [33].

The segmentation task is challenging since the diameter of the retinal vessel is approximately two pixels wide. To address this issue, a Gaussian filter is implemented to amplify low-contrast vessels. The filter is utilized as an input to the multipath convolutional neural network to extract low and high-contrast vessels separately. The downsampling and upsampling pathways are used to extract the low-frequency vessel map. The regions responsible for coding and decoding facilitate the extraction of a segmentation map with high-frequency components. The dilated convolution is employed to extract high-frequency vessel maps, while the fusion network facilitates the acquisition of the ultimate segmentation map for vessels [34]. The Weighted UNet and Residual UNet architectures are employed in a cascaded manner to process globally enhanced and locally enhanced patches effectively. The weighted UNet model receives locally enhanced patches as input and generates a coarse segmentation output. The Residual UNet model utilizes locally enhanced patches and the coarse segmentation output generated by the same model as its input, and subsequently generates the segmentation output [35].

The authors Wei et al. [36] introduced an enhanced version of the UNet architecture that utilizes a genetic algorithm to decrease the computational complexity of the model. In contrast to concatenation employed in the upsampling pathway of UNet, elementwise addition is employed; it requires a lower computation. The process of concatenation results in the production of feature maps of greater size, thereby increasing the computational complexity involved. The authors have achieved significant reductions in computational complexity and parameters. Jin et al. [37] introduced an integrated deep-learning architecture in response to the scale changes observed in retinal fundus photographs and geometrical shape changes. A novel architecture was proposed by integrating UNet and Deformable-ConvNet. The deformable convolution has replaced the conventional convolution layer in the standard UNet architecture. This enhancement resulted in an improved ability to capture the variable shape and scale of the vessel.

According to Lv et al. [38], the accuracy of retinal vessel segmentation plays a crucial role in the diagnosis of diabetes and diabetic retinopathy. The technique of atrous convolution and attention modules is employed to segregate vessel and nonvessel components. The attention module employs a circular bounding box to estimate the maximum anticipated area of retinal vessels. Further the attention module facilitates the acquisition of circular bounding box coordinates by the network. The authors Wang et al. [39] presented an algorithm for vessel segmentation that uses a UNet architecture as a base network, with a focus on capturing the intricate microvascular features of vessels. The UNet architecture was enhanced by incorporating residual blocks and dilated convolutions. The upsampling process in UNet employs bilinear interpolation and transpose convolution techniques to perform weight updates.

Segmenting vessels from fundus images can be challenging due to low contrast, uneven background, and complex structural information. To address this issue, Luo et al. [40] proposed an improved UNet that includes a densely connected network and attention mechanism. The addition of a dense network in the encoder-decoder framework maximizes the utilization of network features. Moreover, the attention mechanism helps to identify vessel structures amidst uneven background and other abnormalities. However, their method still struggles with segmenting retinal vessels from unsmooth edges in the fundus image background.

Thin vessels have low segmentation accuracy due to imbalance with thick vessels, which dominate pixel-wise loss. Learning discriminative features can separate thick and thin vessels during segmentation, reducing the negative effects of their imbalanced ratio [41].

Many deep convolutional neural network methods have limited ability to capture global context information of larger regions due to their small receptive fields. Jiang et al. [42] suggested using parallel convolution layers with various dilation rates to enhance the model's ability to gather dense feature information and accurately capture retinal vessel information of different sizes.

Currently, the designs of deep networks mostly concentrate on vessels that are easy to segment. However, they tend to ignore vessels that are more challenging to segment, like thin vessels or those with unclear boundaries. Wang et al. [43] introduced a shallow UNet with three decoders. There is one decoder that is utilized to identify regions that are easy or difficult to segment, while the other two are responsible for segmenting the vessels.

Many current retinal vessel segmentation models in deep learning treat each pixel equally. However, the multi-scale vessel structure is a crucial factor that significantly impacts the segmentation results, especially in thin vessels. A New attention block models global dependencies and minimizes inconsistencies by considering pixel correlations [44].

Motivated by the existing methods, depthwise parallel attention UNet is proposed as illustrated in Fig. 1 to automatically segment retinal vessels. Experiments on well-known public datasets resulted a good performance among existing methods. This paper presents a lightweight retinal vessel segmentation model based on a deep neural network. The main contributions of our work are as follows,

1. In order to address the issue of information loss resulting from the implementation of multiple convolutions and downsampling operations, our proposed method involves the use of a depthwise convolution instead of a traditional convolution in the downsampling path of the UNet model.
2. To capture the intricate informations from multiple channels we have used Global Information Embedding Block(GIE) in the Downsampling layer of the UNet.

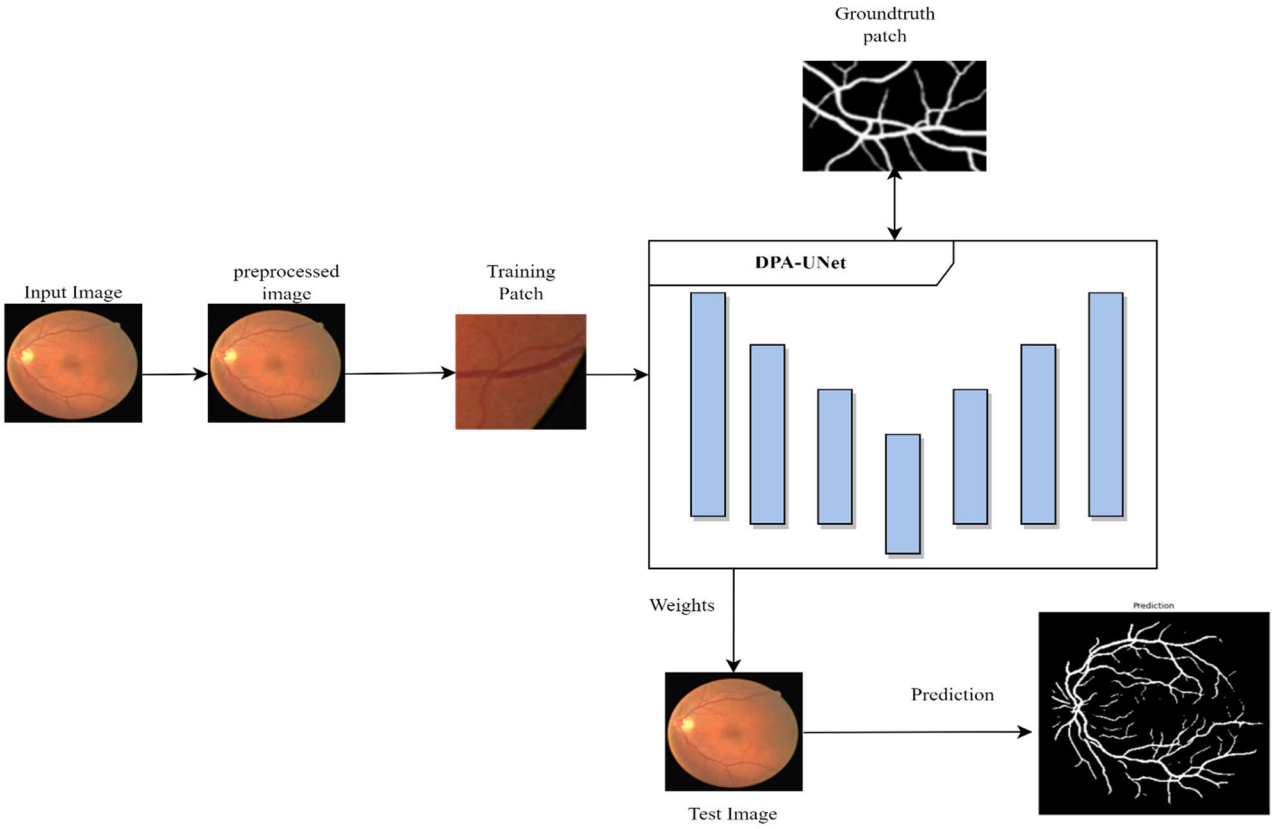


FIGURE 1. Overview of the proposed DPA-UNet.

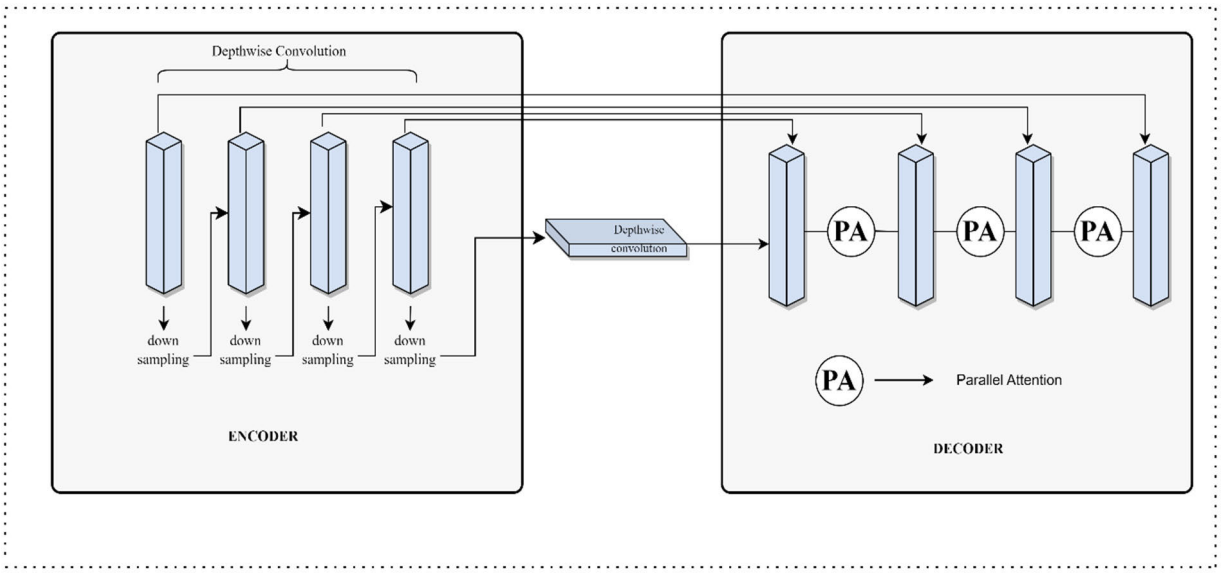


FIGURE 2. Proposed depthwise parallel attention UNet.

3. To effectively fuse the features from the encoder and decoder blocks, a parallel attention mechanism is proposed to efficiently utilize the channel and spatial information.

In summary, deep learning methods can work with limited ground truth. Hence it is difficult to get ground truth images for medical images. The main challenge in working with deep learning models is making a more efficient and lightweight

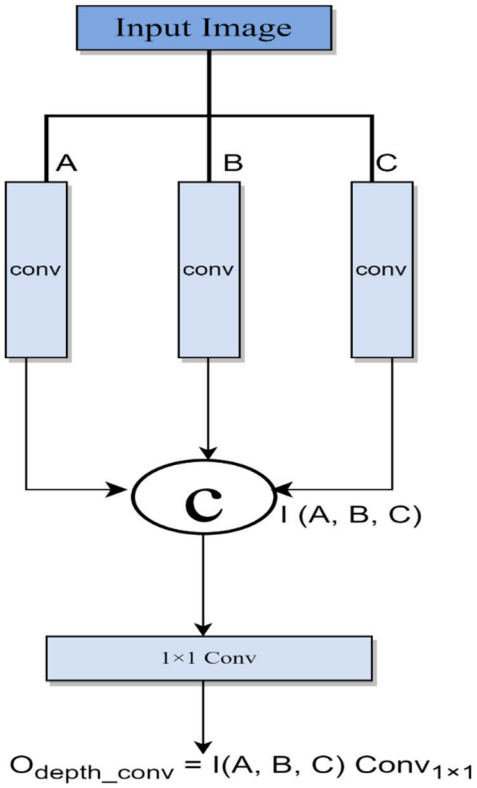


FIGURE 3. Depthwise convolution block (DCB).

network. Therefore, we propose DPA-UNet to segment retinal vessels efficiently with fewer network parameters.

III. PROPOSED METHODOLOGY

This section explains the channel and spatial attention UNet designed to segment the retinal vessels efficiently. We present the proposed architecture design and a detailed explanation of each block.

A. WORK FLOW OF THE PROPOSED DPA-UNet

We utilize UNet as a backbone network, a depthwise convolution block, GIE block, and a parallel attention block integrated into UNet to increase the segmentation performance of the model. Depth-wise convolution block in the proposed model performs convolution, which applies a single filter to each input image channel, resulting in a set of output features. This operation reduces the number multiplications needed for the convolutional layers and increases the computation efficiency of the model. A parallel attention block considers channel and spatial relationships so that long-rang dependencies could benefit the understanding of thin and low-contrast vessels. The overall architectural design is shown in Fig.1.

B. BASE NETWORK

The rich representation capabilities of many well-designed deep neural networks show excellent performance in semantic segmentation tasks. One such architecture is a fully convolutional neural network (FCN) [31], features are perceived using an encoder and decoder structure and a fully convolution classification. Aggregating low-level features into high-level features using skip connections is an added contribution to semantic segmentation to recover the reduced details. Inspired by skip connection and FCN, UNet architecture is developed which modifies and extends the structure of FCN with a U-shaped encoder-decoder architecture. UNet is a well-known deep-learning architecture for medical image segmentation, especially for retinal vessel segmentation. Many variants of UNet architectures have been explored by the researchers. A typical UNet has a 3×3 convolution, which limits the feature representation ability. Traditional convolution increases the parameters, forcing the network to memorize rather than learning. It leads to overfitting. To overcome this, we propose an enhanced base network composed of three key blocks: a depth Convolution Block (DCB) with a Global Information Embedding (GIE) block and a parallel Attention Block (PAB).

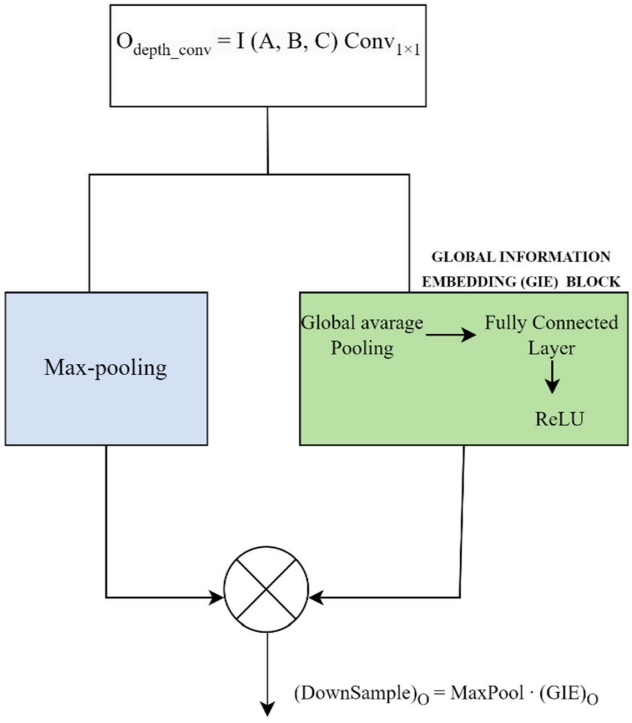


FIGURE 4. Modified downsampling block.

1) DCB BLOCK

Traditional convolution increases the parameters and shrinks the spatial dimension of the input image. Minimizing the number of parameters can effectively mitigate the complexity of the network. Protecting the spatial extent of the image also plays a vital role in information loss, which is important to effectively segment the features with different sizes. Motivated by [45] traditional convolution layer is replaced with depth wise convolution in base network. A DCB block is proposed in UNet to build an efficient base network shown in Fig.3.

A depthwise convolution preserves the spatial dimension and channel dimension of the image with a smaller network

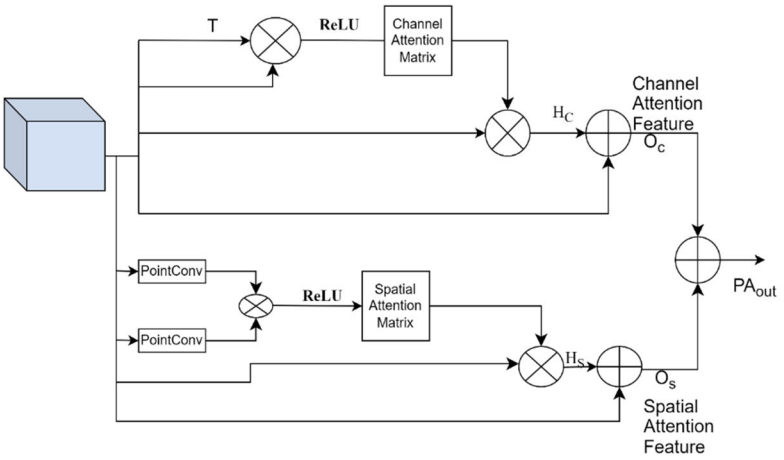


FIGURE 5. Parallel attention block.

TABLE 1. Dataset description.

Dataset	Number of images	Image Size	Training -Testing
DRIVE	40	565×584	20/20
STARE	20	700×605	10/10
CHASE_DB1	28	999×960	14/14

complexity. The convolution output without any loss in the channel and spatial dimension is given to the upsampling path in the UNet since feature representation in the upsampling path is vital in the segmentation output. Applying a typical 3×3 convolution filter to each channel of the input image could reduce the complexity of the network compared to standard convolution. Each layer in the UNet has two convolutional layers with a kernel size of 3×3 that are applied to each channel of the input image and cascaded together to obtain the output feature map without losing any channel or spatial dimension. After performing convolution in each channel separately, those features are concatenated, and 1×1 kernel is applied across all channels simultaneously to capture the complex relationship between thin and low contrast vessels mentioned in (1). It could help to increase the feature representation capability of the model with the limited dataset.

$$O_{depth\_conv} = I(A, B, C)Conv_{1\times1} \tag{1}$$

2) MODIFIED DOWNSAMPLING (GIE) BLOCK

To explicitly model the interdependencies between the channels of the convolution features, we embed the GIE block inspired by the squeeze and eexcitation [46] block after convolution, which showed promising pperformance in modeling interdependencies between the channels as shown in Fig.4. Feature recalibration is performed to selectively emphasize the features, allowing the network to learn and utilize the global information. The re-calibrated feature is next exposed to a squeeze operation, which combines the feature maps across spatial dimensions to generate a channel descriptor. An additional activation function is applied to control how excited each channel is. The output feature map's channels are each given a signal as part of the squeeze process, which makes the best use of channel dependencies possible. Learned filters are limited to the local receptive field and are unable to capture data outside the receptive field. To solve this problem, we used the squeeze operation, which performs global average pooling to generate channel wise information. Since we are aggregating multichannel features, the network should learn a non-mutually exclusive relationship between channels. To achieve this, we used the ReLU activation function, allowing the network to introduce nonlinearities into the channel interaction. This nonlinear interaction between the channels allows the network to combine features from multiple channels, which helps the network to produce more complex and abstract feature. Incorporating a fully connected layer between the global average pooling and activation function has been shown to facilitate generalization and mitigate model complexity. Finally, max pooling and GIE block outputs are combined to get the final output of downsampling mentioned in (2). Each layer of the UNet follows the above-mentioned downsampling operation.

$$(DownSample)_O = MaxPool \cdot (GIE)_O \tag{2}$$

3) PARALLEL ATTENTION BLOCK

The upsampling path of UNet plays a significant role in feature representation and segmentation. Motivated by the dual attention mechanism [47], we propose a parallel attention block to model a network that produces long range contextual information. Each upsampling layer has a parallel attention module. The concatenation process involves merging of features obtained from the upsampling layer with those obtained from the skip connection in the downsampling path. The UNet architecture utilizes skip connections to transfer

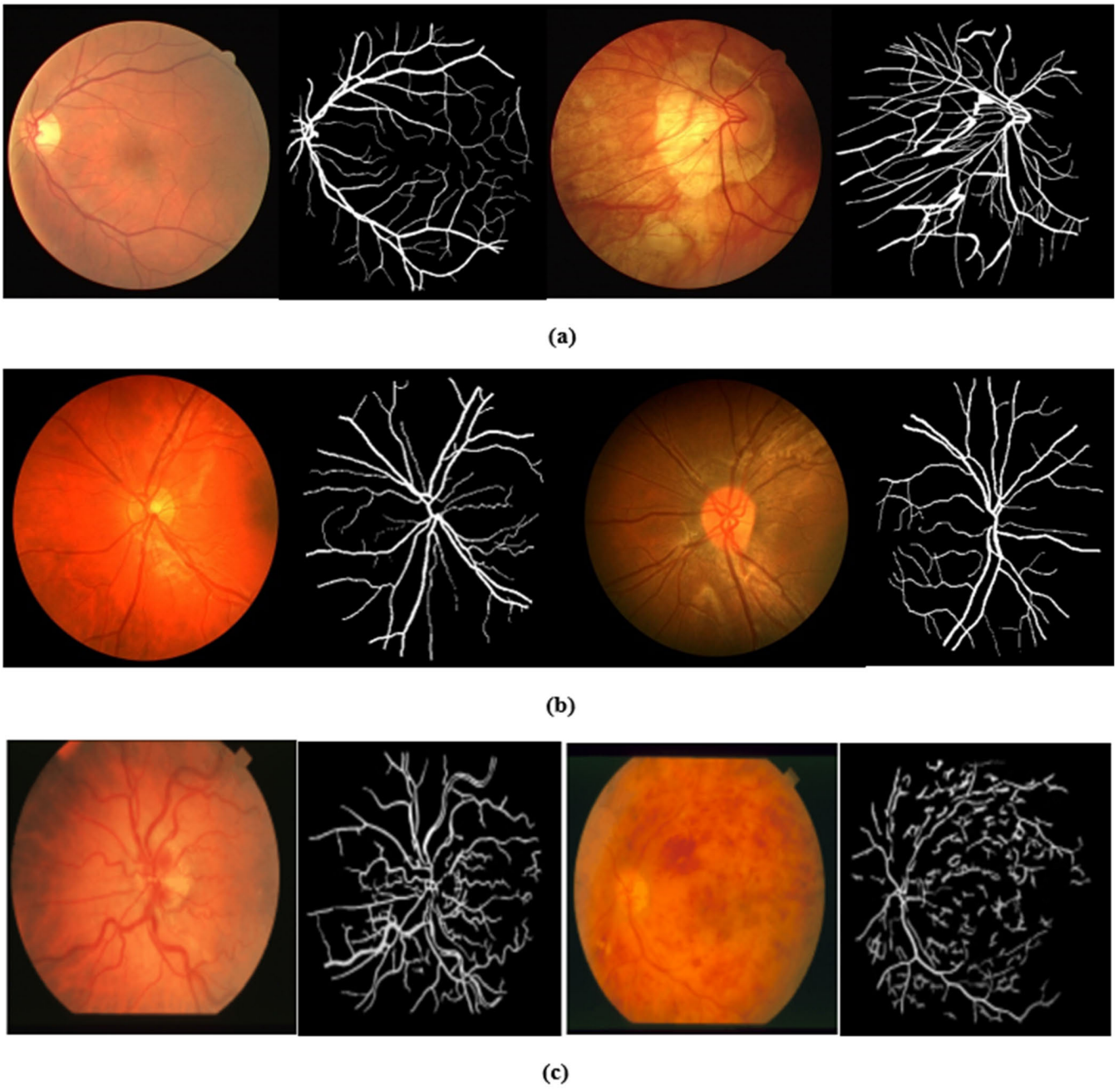


FIGURE 6. Original images from the dataset and corresponding ground truth images. (a) DRIVE dataset, (b) CHASE_DB1 dataset (c). STARE dataset.

convolutional features from the downsampling path. The downsampling path consists of layers that perform depthwise convolution. The outcome of the concatenation process is then fed into a parallel attention mechanism that comprises spatial and channel attention modules.

The spatial and channel attention module is responsible for producing distinct long-range spatial contextual features and channel dimension features achieved by performing following steps. Initially, a channel attention matrix mentioned in (3) is produced to examine the relationship among the pixels within the features. The multiplication operation given in equation (4) is applied to the original feature and the attention matrix. Following this, the resultant matrix obtained from the multiplication of the matrix and the initial feature is merged using element-wise addition, thereby producing a new feature that reflects channel information over a significant range, as indicated by equation (5).

$$CA_{Matrix} = ReLU(F_C^T \cdot F_C) \tag{3}$$

$$Hc = (CA_{Matrix} \cdot F_C) \tag{4}$$

$$O_c = Hc + F_C \tag{5}$$

Fig 5. Shows the structure of the parallel attention module. F_C^T represents the transpose of the input feature map. F_C

TABLE 2. Performance of attention block in the proposed model DPA-UNet.

Position of Attention block	Sensitivity	Specificity	Accuracy	AUC
Upsampling path	0.8193	0.9863	0.9882	0.9856
Downsampling path	0.8054	0.9621	0.9676	0.9749

TABLE 3. Computational complexity comparison between Traditional and Depth wise convolution.

Depth wise convolution:	**Traditional convolution**
Input Image = 256×256×3 Filter= Three 3×3×1 filters Required (256×256) × (3×3×1) × (3) = 1,769,472	Input Image = 256×256×3 Filter= 3×3×3 (256×256) × (3×3×3) × (3) = 5,308,416

```
1/1 [==============================] - 0s 364ms/step
1/1 [==============================] - 0s 86ms/step
Traditional Convolution Result Shape: (1, 256, 256, 1)
Depthwise Convolution Result Shape: (1, 256, 256, 3)
```

TABLE 4. Algorithm of the proposed depth wise separable attention UNet model.

Algorithm of the Proposed Model

- **Image preprocessing**

  -2D Gaussian function for base intensity estimation

  -intensity transfer is performed to achieve color balance

- **Network Training**

  -Initialize Weights (W_{DSA_UNet}=0) of DSA-UNet

  -for epochs 0 to 2000,

  -compute results (W_{DSA_UNet}, Input);

  -compute loss L=Cross Entropy (Target Mask, Results);

  - compute gradient (Adam optimizer with momentum of 0.9 and weight decay of 0.0001);

  -update weights

  -end

  -prediction DSA-UNet (W_{DSA_UNet}, Input)

is the input feature map given to parallel attention module. CA_{Matrix} is the channel attention matrix generated after applying ReLU activation to the product of the input feature and the transpose of the input feature. The final channel output feature map is produced by adding the product of the channel attention matrix and input feature represented as H_C and the original input feature. This will make the network to refine the channel features. The spatial attention module computes channel dimension features in an instance similar to channel attention module, except the channel attention matrix. where 1×1 convolution is performed on the input feature map before generating the spatial attention matrix mentioned in (6), (7), and (8).

$$SA_{Matrix} = ReLU(F_{Conv1\times1} \cdot F^{T}_{Cov1\times1}) \tag{6}$$

$$Hs = SA_{Matrix} \cdot F_S^{Conv1\times1} \tag{7}$$

$$O_S = Hs + F_S \tag{8}$$

where SA_{Matrix} represents spatial attention matrix. Hs represents the product of spatial attention matrix and input feature map (F_S − *spatial attiention module input*). Finally, outputs of attention modules are summed up to get the features, where O_C and O_S represents output of channel and spatial attention module respectively. Each layer in the upsampling path of UNet adopts this parallel attention mechanism to aggregate the long-range contextual information of both spatial and channel dimensions, as shown in (9).

$$PA_{out} = O_C + O_S \tag{9}$$

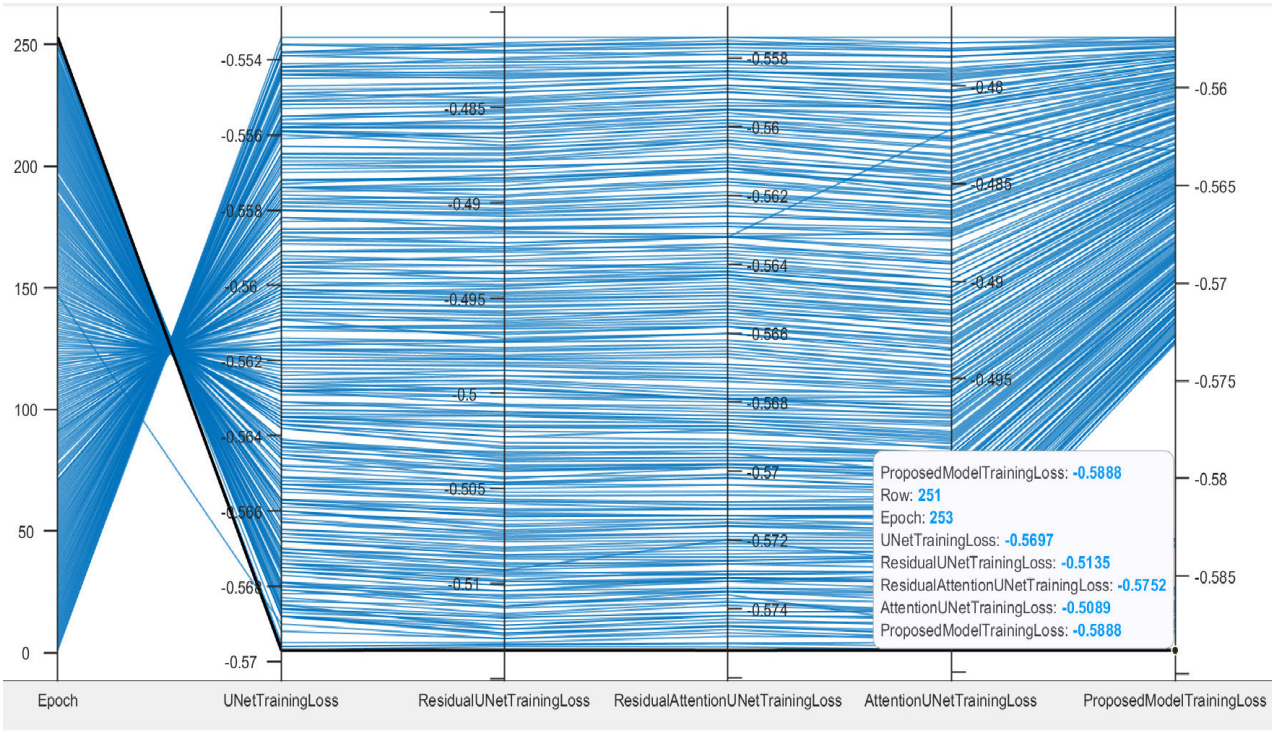


FIGURE 7. Training loss comparison.

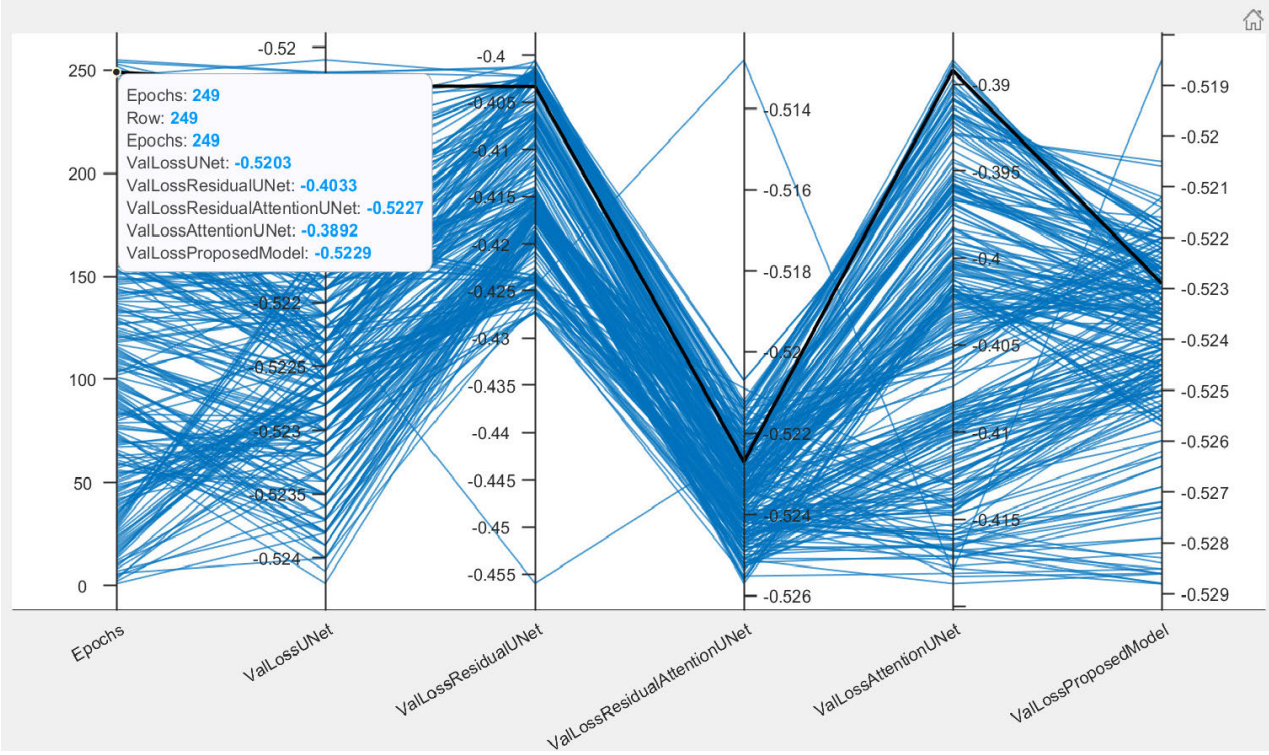


FIGURE 8. Validation loss comparison when performing the validation with same type of dataset which is used in training.

IV. DATASETS

The performance of DPA-UNet is evaluated on the STARE, DRIVE and CHASE_DB1 datasets which are distinctive datasets available for the vessel segmentation task mentioned in Table.1. Original fundus images and their corresponding ground truth images are shown in Fig.6.

The DRIVE dataset consists of 40 images divided into a training and test containing 20 images each. Each image has

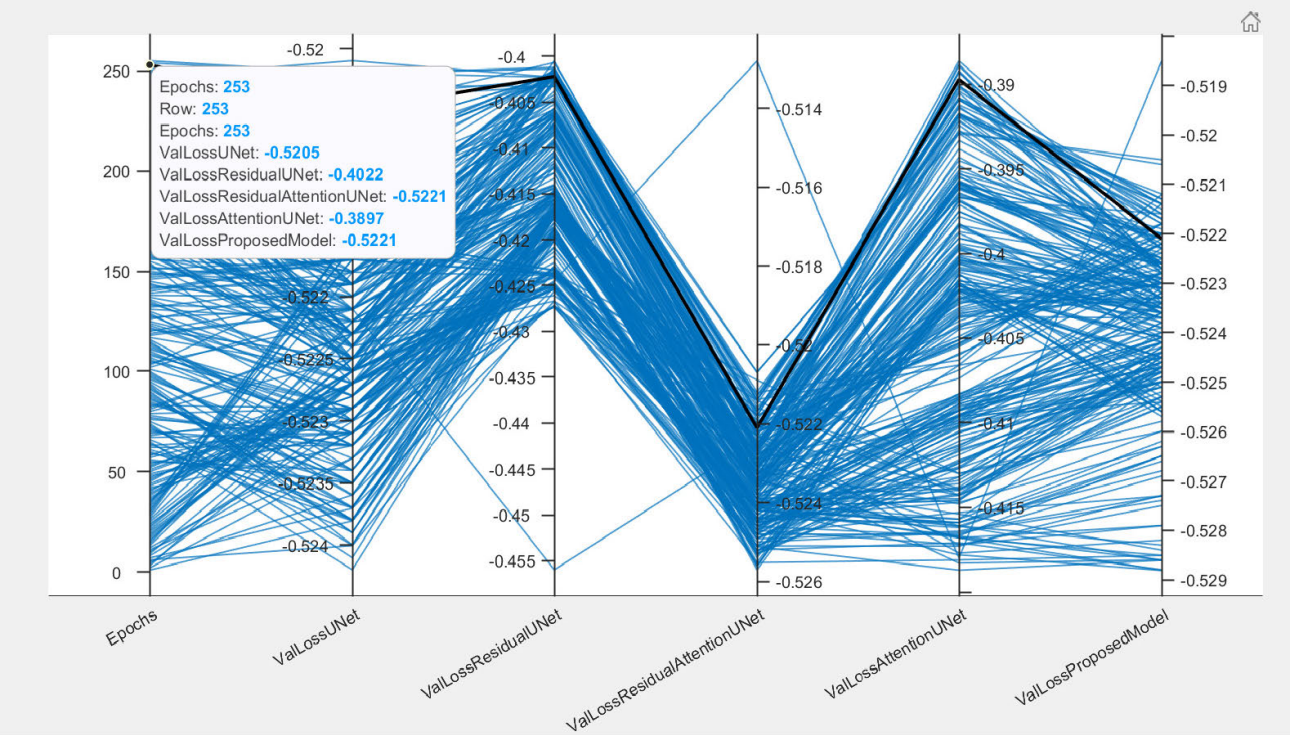


FIGURE 9. Cross validation loss comparison.

a size of 565 × 584. The canon CR5 nonmydriatic 3CCD camera was used to acquire the image with a 45-degree field of view [48].

The STARE dataset consists of 20 fundus images with a size of 605 × 700 divided into training and testing sets containing 10 images each. A 35V TopCon camera was used to acquire the images [25].

The CHASE_DB1 [49] dataset is taken from 14 children and consists of 28 images with a size of 960×999, and images were acquired using a Nidek camera at 30V.

Since we have a limited dataset patch based training is used, Table 1 shows the dataset details used to evaluate the proposed algorithm.

A. MULTISCALE PATCH EXTRACTION

We extracted the patches of size 256 × 256 from the images in the dataset. For both original and resized image we have assigned the window dimension $W_k \times H_k \times C$ with stride of S=2.

We used 960 patches for training. The patch extraction from the data with a size of H×W is denoted as,

$$P = ([\frac{W - W_k}{S}] + 1) \times ([\frac{H - H_k}{S}] + 1)$$

V. SIMULATION RESULTS

The proposed model incorporates both an upsampling path and a downsampling path. To evaluate the model's performance, we trained and tested it by integrating the attention module separately in both the upsampling and downsampling paths. The performance of the model was evaluated individually, and upon combining the module in the upsampling path, the segmentation results were found to be appropriate as presented in Table 2.

We attempted fusing the multiple modules like Residual Net, attention gate, depth wise convolution in UNet. We Choose the depth wise separable attention UNet since it segmented the vessels efficiently with less network parameters while maintaining accuracy compared to existing methods. Finally, the Adam optimizer and binary cross entropy are used to calculate the gradient and loss respectively. Table 4. Shows the algorithm of the proposed model. The experiment is trained and tested in the TensorFlow framework. Training loss and validation loss curve was plotted for last 250 epochs shown in Fig. 7. and Fig. 8. It shows that, compared to existing state of art methods proposed model achieved less training and validation loss respectively. Since we have not used more data for training to test the generalization of the model, we performed cross validation and calculated the validation loss. Residual Attention UNet and the Proposed model achieved the acceptable validation loss compared to other existing state art methods shown in Fig.9.

The main advantage of performing depth wise convolution is shown in table.3. depth wise convolution has the ability to significantly reduce the number of parameters and calculations. Depth wise convolution decomposes the convolution operation into separate convolutions for each input channel, using a separate set of filters for each channel. This reduces the number of parameters compared to traditional convolutional layers, making the model more efficient. Depth wise convolution is a highly effective way to capture localized

TABLE 5. Average performance analysis of module fusion on the DRIVE, STARE and CHASE_DB datasets.

Modules	Specificity	Sensitivity	Accuracy	AUC
UNet	0.8032	0.9865	0.9647	0.9857
Residual-UNet	0.8035	0.9877	0.9652	0.9862
Residual-Attention-UNet	0.8142	0.9865	0.9857	0.9885
Attention UNet	0.8063	0.9869	0.9762	0.9859
Depth wise Separable- Attention UNet	**0.8142**	**0.9872**	**0.9933**	**0.9887**

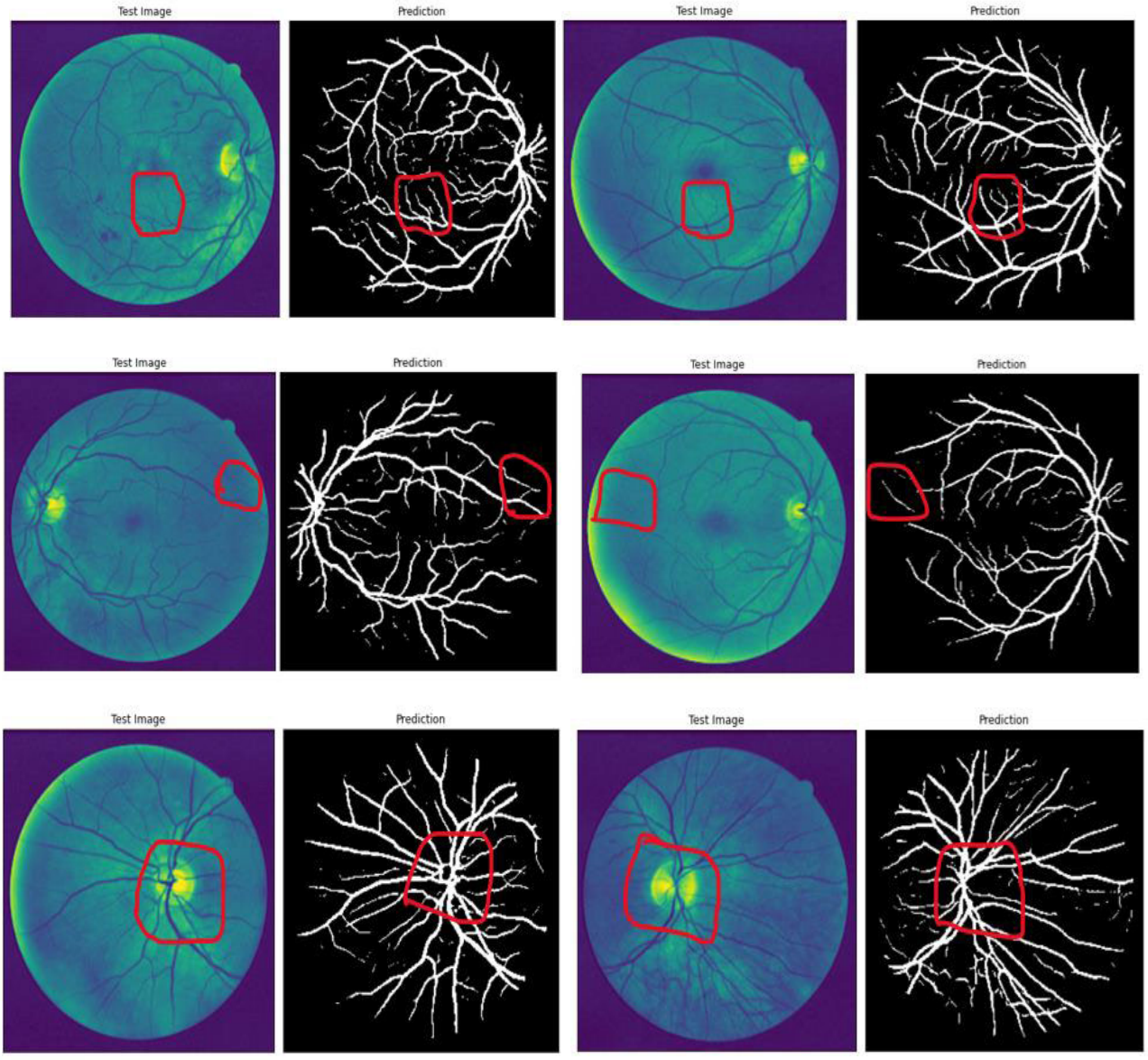


FIGURE 10. Vessel segmentation results of proposed work in the green channel, Row 1 and 2 shows that the model predicts low contrast vessels also, Row 3 shows that even in strong central reflex vessel struture is predicted appropriately.

features within individual input channels. By allowing for feature separation, it enhances the separation of features, enabling each channel to focus on different aspects of the input. Since each input channel has its own set of filters, the number of multiplications required per channel is reduced, leading to faster inference times. After analyzing the table.3. provided, it is evident that conventional convolution filters demand 364ms for processing a single image. On the other

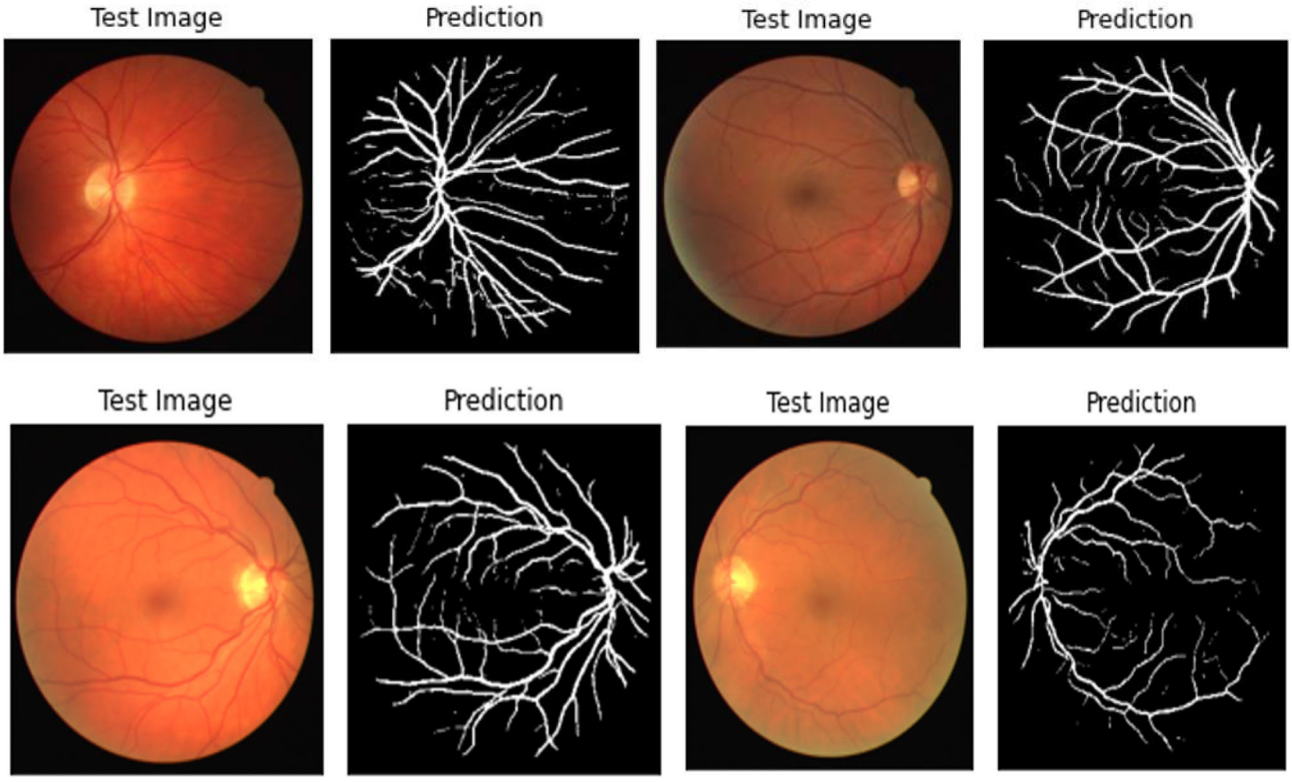


FIGURE 11. Vessel segmentation results of the proposed model in RGB image.

TABLE 6. Computational complexity comparision of proposed model with existing models.

Model	UNet	Residual UNet	Attention UNet	Residual Attention UNet	Proposed Model (DPA-UNet)
Number of Parameters (In Millions)	1.96 M	2.08 M	2.34 M	2.45 M	**1.57 M**

hand, a depth-wise convolution filter only requires 86ms to process a single image.

The comparison of different module fusion performances on STARE, DRIVE, and CHASE_DB are mentioned in Table.5. The depthwise separable attention UNet was selected due to its efficient vessel segmentation capabilities with fewer network parameters, while still maintaining accuracy compared to other methods.

Evaluation of proposed method is mentioned in Table.7. A, Table.7. B, Table.7. C. proposed methodology is compared with existing vessel segmentation algorithms. It shows proposed methodology achieved a competitive performance with exiting algorithms with less parameters mentioned in Table.6. Fig. 10 shows algorithms segmentation result on green channel. Fig.11. shows the segmentation result of the proposed algorithm on RGB image. We have compared our segmented results with ground truth images shown in Fig. 12.

We conducted a thorough evaluation of vessel segmentation using the DPA-UNet model and compared its performance against both unsupervised and supervised approaches across three separate datasets. Our assessment utilized important metrics, including Sensitivity, Specificity, Accuracy, and AUC, resulting in clear and compelling outcomes. Our findings demonstrate the superior capabilities of the DPA-UNet model in vessel segmentation, which could be of great value in various medical applications. Table 7, A, B, C presents the findings of a study that examined various segmentation methods applied to the DRIVE, STARE, and CHASE datasets. The data clearly indicates that supervised methods tend to outperform their unsupervised counterparts. Additionally, deep learning techniques perform exceptionally well in terms of Accuracy (Acc) and AUC. The STARE dataset and CHASE_DB1 dataset contain a large number of lesion abnormal images with uneven illumination, low contrast, and similar morphology of blood vessels and some pathological sites, which lead to difficulty in retinal blood vessel segmentation. When we observe the table, our proposed approach achieved a good accuracy on all three datasets compared to existing method.

(a)DRIVE

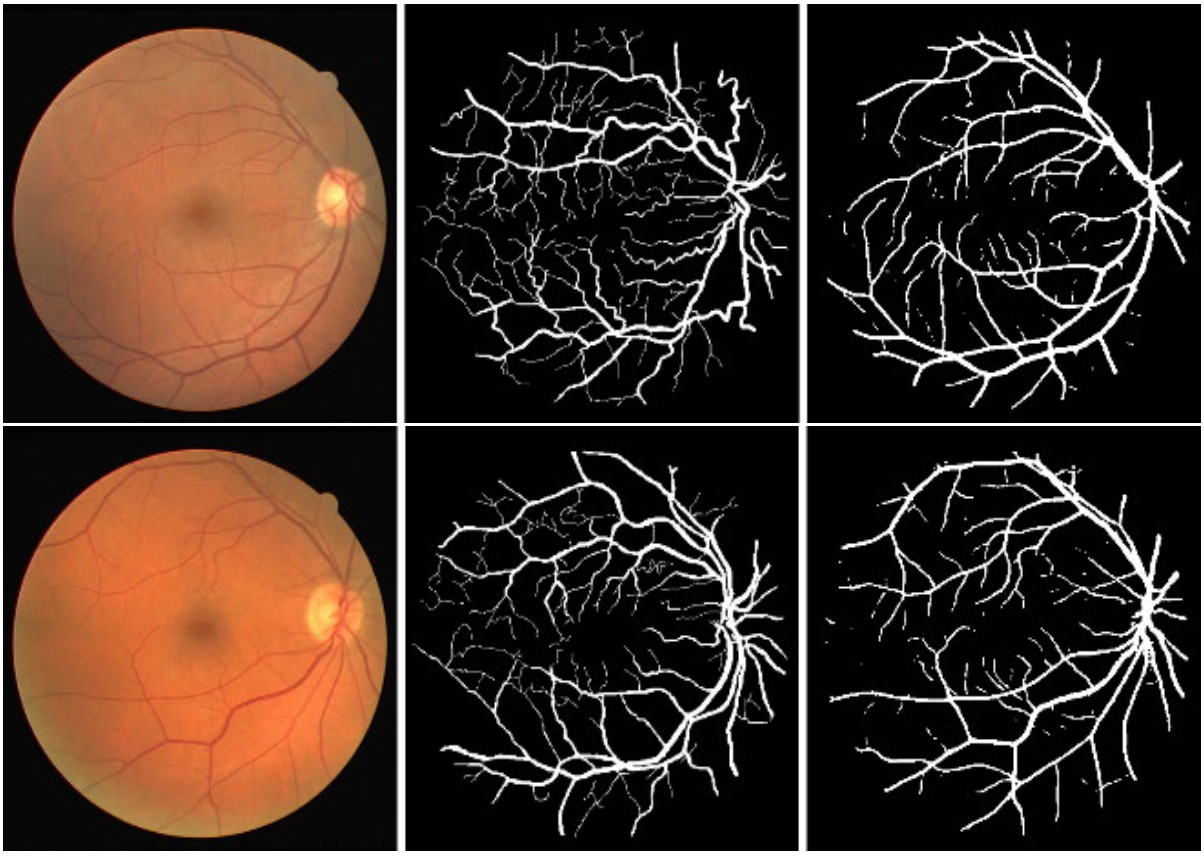
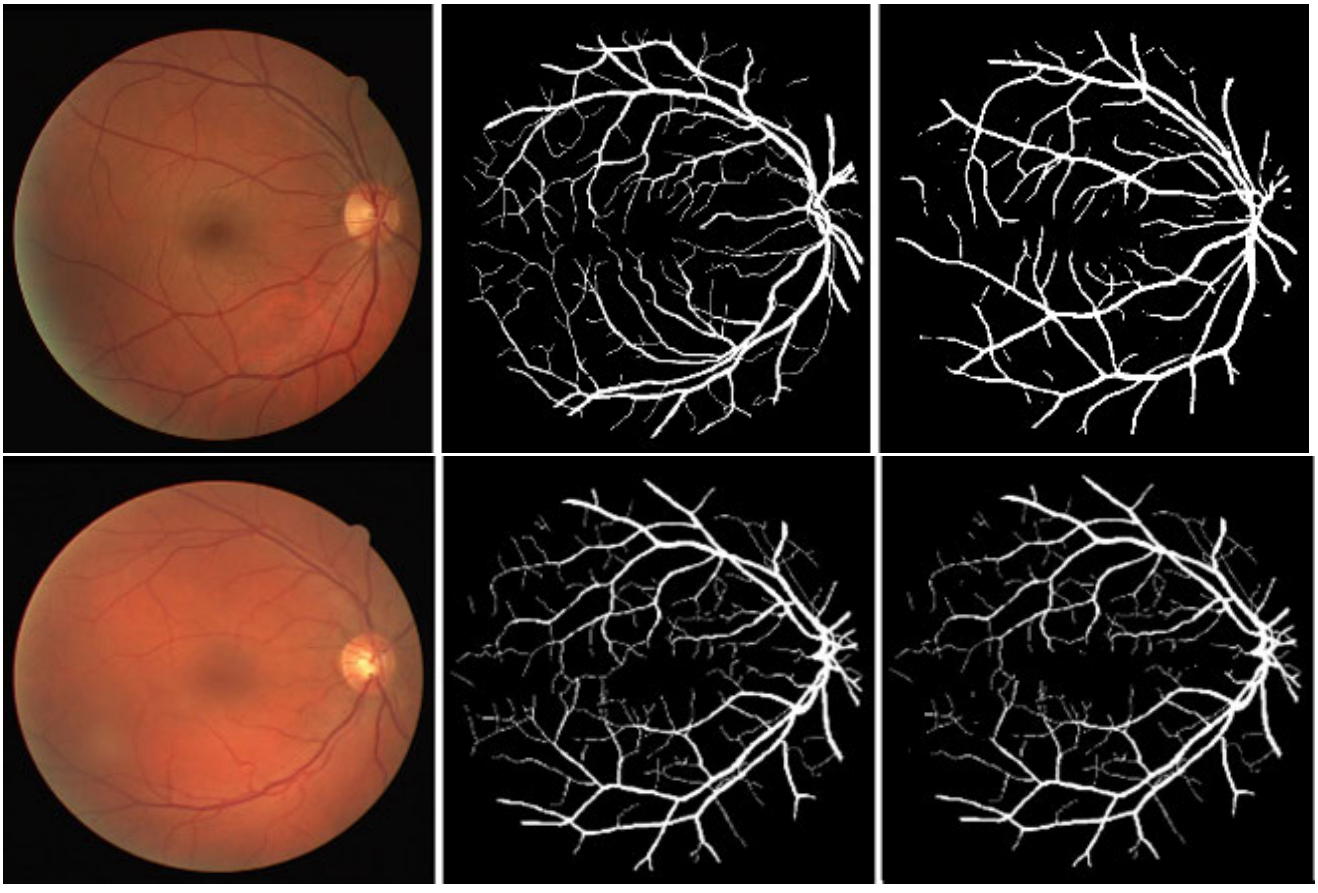


FIGURE 12. Comparison of segmentation result of a proposed model with ground truth.

(b)STARE



(c) CHASE_DB1

FIGURE 12. ***(Continued.)*** **Comparison of segmentation result of a proposed model with ground truth.**

VI. DISCUSSION

This article introduces a depthwise parallel attention UNet (DPA-UNet) for automatically segmenting retinal vessels in color fundus images. The illumination changes, centreline reflexes, bifurcation points, and reduced size and width of the vessels from the optic disc brought challenges in vessel segmentation. Our model integrates depth wise convolution to deal with spatial and channel dimension reduction due to standard convolution operation. In the downsampling process we introduced two pooling operations, max pooling and the GIE block, to address interdependencies between multiple channels. Therefore, the problem of modeling channel- wise dependencies is solved, and selective feature emphasizing is achieved. Table 2 shows the performance of the parallel attention block. We used it in the upsampling path of UNet to extract the long-range contextual information. We tested three datasets, and the proposed architecture showed good performance in segmenting thin vessels even in low contrast

TABLE 7. A. Analysis of the proposed method over various existing methods on DRIVE datasets. B. Analysis of the proposed method with various existing methods on STARE datasets. C. Analysis of the proposed method with various existing methods on CHASE_DB1 datasets.

(a)

Methodology		Year	DRIVE Dataset			
			Sp	Se	Acc	AUC
Traditional Methods	Hoover et.al.[25]	2000	-	-	0.9275	
	Staal et.al.[24]	2014	-	-	0.9516	-
	Rodrigues et.al.[26]	2017	0.9801	0.7165	0.9472	-
	FC-CRF[27]	2017	0.9684	0.7897	-	-
	Liskowski et.al.[28]	2016	0.9241	0.9160	0.9251	0.9738
	Li et.al.[29]	2016	0.9816	0.7569	0.9527	0.9738
Deep learning Methods	Yang et.al.[33]	2021	0.9751	-	0.9579	-
	Tian et.al.[34]	2020	0.9690	0.8639	0.9580	0.9560
	Lian et.al.[35]	2021	0.9861	0.8278	0.9692	-
	Wei et.al.[36]	2022	0.9758	0.8300	0.9577	0.9823
	Jin et.al.[37]	2018	-	-	0.9481	0.9718
	Lv et.al.[38]	2020	-	-	0.9558	0.9847
	Wang et.al.[39]	2022	0.9867	0.8385	0.9611	0.9829
	Luo et.al. [40]	2019	0.9814	0.8075	0.9663	0.9846
	Yan et.al. [41]	2019	0.9820	0.7631	0.9538	0.9750
	Jiang et.al.[42]	2018	0.9890	0.7839	0.9709	0.9864
	Wang [43]	2020	0.9813	0.7991	0.9581	0.9823
	Li et.al.[44]	2021	0.9883	0.8145	0.9769	0.9895
	Proposed (DSA_UNET)		0.9882	0.8144	**0.9882**	**0.9897**

('- 'represents authors have not evaluated the particular parameter for the given dataset)

(b)

Methodology		Year	STARE Dataset			
			Sp	Se	Acc	AUC
Traditional methods	FC-CRF[27]	2017	0.9738	0.7680	-	-
	Liskowski et.al.[28]	2016	0.9304	0.9307	0.9620	0.9820
	Li et.al.[29]	2016	0.9844	0.7726	0.9628	0.9879
Deep learning Methods	Yang et.al.[33]	2021	0.9821	-	0.9626	-
	Lian et.al.[35]	2021	0.9916	0.8342	0.9740	-
	Wei et.al.[36]	2022	0.9846	0.8658	0.9719	0.9921
	Jin et.al.[37]	2018	-	-	0.9445	0.9690
	Lv et.al.[38]	2020	-	-	0.9640	0.9824
	Wang et.al.[39]	2022	0.9866	0.8006	0.9796	0.9902
	Luo et.al. [40]	2019	0.9762	0.8437	0.9684	0.9765
	Yan et.al. [41]	2019	0.9857	0.7735	0.9638	0.9833
	Jiang et.al.[42]	2018	0.9904	0.8249	0.9781	0.9927
	Wang [43]	2020	0.9844	0.8186	0.9673	0.9881
	Li et.al.[44]	2021	0.9889	0.8505	0.9797	0.9924
	Proposed (DSA_UNET)		0.9856	**0.9816**	**0.9865**	0.9902

(c)

Methodology		Year	CHASE_DB1 Dataset			
			Sp	Se	Acc	AUC
Traditional methods	FC-CRF[27]	2017	0.9712	0.7277	-	-
	Li et.al.[29]	2016	0.9793	0.7507	0.9581	0.9716
Deep learning Methods	Yang et.al.[33]	2021	0.9776	-	0.9632	-
	Tian et.al.[34]	2020	0.9680	0.8778	0.9601	0.9577
	Wei et.al.[36]	2022	0.9818	0.8463	0.9667	0.9880
	Lv et.al.[38]	2020	-	-	0.9608	0.9865
	Wang et.al.[39]	2022	0.9659	0.7958	0.9662	0.9873
	Yan et.al. [41]	2019	0.9806	0.7641	0.9607	0.9776

TABLE 7. *(Continued.)* **A. Analysis of the proposed method over various existing methods on DRIVE datasets. B. Analysis of the proposed method with various existing methods on STARE datasets. C. Analysis of the proposed method with various existing methods on CHASE_DB1 datasets.**

	Jiang et.al.[42]	**2018**	0.9894	0.7839	0.9721	0.9866
	Wang [43]	**2020**	0.9813	0.8239	0.9670	0.9871
	Li et.al.[44]	**2021**	0.9862	0.8334	0.9802	0.9912
	Proposed (DSA_UNET)		0.9879	0.8226	**0.9803**	0.9900

and strong central reflex. Furthermore, compared to [44], proposed method can distinguish thin vessels better. Hence our method is consistent and robust with less parameters. Further generalization ability of the model is tested by performing cross validation. Even though depth wise convolution is computationally efficient, there are some potential drawbacks when we implement this in UNet. It is sensitive to channel correlations in the input data, so a proper preprocessing technique should be employed before training. Each input channel is processed independently, it leads to reduced information sharing among the channels while this reduces the computation complexity. At the same time, it limits the model's ability to learn and represent the high-level features that rely on interactions between different channels, to overcome this we used GIE block. When implementing depthwise convolution the main challenge is capturing multichannel information, distinct efforts need to be taken to introduce inter-channel dependency.

VII. CONCLUSION

In this article we propose a depthwise parallel attention UNet for vessel segmentation in fundus images. Depth wise separable convolution is performed instead of regular convolution to sort the channel features as well as to lessen the parameters. To lower the overfitting problem and preserve the maximum vessel information between the convolution layers, dropout technique is adopted in the proposed model. Channel interdependency is addressed by global information embedding block. It also helps to lessen the information loss due to pooling operation. The upsampling path is responsible for segmented feature maps, so we used a parallel attention block after concatenation, which helps to preserve and fully explore features that are more important. We conducted experiments on the DRIVE, STARE and CHASE_DB1 datasets to evaluate the proposed algorithm. We have achieved acceptable performance with less parameters compared to existing methods. Though the proposed model has achieved a competitive performance compared to existing state of art methods while maintaining less parameter, we observed lack of ground truth for fundus dataset for vessel segmentation prevented a better validation performance of the network, specifically fundus images with abnormalities.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Dr. D. Anusha, at the Vijaya Diagnostic Center in Hyderabad, Telangana, for her insightful comments on the findings and ideas for the article's standards.

REFERENCES

[1] M. J. Mayer and J. E. Dowling, "The retina: An approachable part of the brain," *Amer. J. Psychol.*, vol. 101, no. 4, p. 602, 1988, doi: 10.2307/1423238.

[2] V. Conti, L. Rundo, C. Militello, V. M. Salerno, S. Vitabile, and S. M. Siniscalchi, "A multimodal retina-iris biometric system using the levenshtein distance for spatial feature comparison," *IET Biometrics*, vol. 10, no. 1, pp. 44–64, Jan. 2021, doi: 10.1049/bme2.12001.

[3] S. M. Lajevardi, A. Arakala, S. A. Davis, and K. J. Horadam, "Retina verification system based on biometric graph matching," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3625–3635, Sep. 2013, doi: 10.1109/TIP.2013.2266257.

[4] Z. Waheed, A. Waheed, and M. U. Akram, "A robust non-vascular retina recognition system using structural features of retinal image," in *Proc. 13th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2016, pp. 101–105, doi: 10.1109/IBCAST.2016.7429862.

[5] J. Krause, V. Gulshan, E. Rahimy, P. Karth, K. Widner, G. S. Corrado, L. Peng, and D. R. Webster, "Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy," *Ophthalmology*, vol. 125, no. 8, pp. 1264–1272, Aug. 2018, doi: 10.1016/j.ophtha.2018.01.034.

[6] O. O. Sule, "A survey of deep learning for retinal blood vessel segmentation methods: Taxonomy, trends, challenges and future directions," *IEEE Access*, vol. 10, pp. 38202–38236, 2022, doi: 10.1109/ACCESS.2022.3163247.

[7] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Med. Imag.*, vol. 8, no. 3, pp. 263–269, Sep. 1989, doi: 10.1109/42.34715.

[8] M. Al-Rawi, M. Qutaishat, and M. Arrar, "An improved matched filter for blood vessel detection of digital retinal images," *Comput. Biol. Med.*, vol. 37, no. 2, pp. 262–267, Feb. 2007, doi: 10.1016/j.compbiomed.2006.03.003.

[9] B. Zhang, L. Zhang, L. Zhang, and F. Karray, "Retinal vessel extraction by matched filter with first-order derivative of Gaussian," *Comput. Biol. Med.*, vol. 40, no. 4, pp. 438–445, Apr. 2010, doi: 10.1016/j.compbiomed.2010.02.008.

[10] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1010–1019, Jul. 2001, doi: 10.1109/83.931095.

[11] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1200–1213, Sep. 2006, doi: 10.1109/TMI.2006.879955.

[12] E. Moghimirad, S. H. Rezatofighi, and H. Soltanian-Zadeh, "Retinal vessel segmentation using a multi-scale medialness function," *Comput. Biol. Med.*, vol. 42, no. 1, pp. 50–60, Jan. 2012, doi: 10.1016/j.compbiomed.2011.10.008.

[13] Q. Li, J. You, and D. Zhang, "Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 7600–7610, Jul. 2012, doi: 10.1016/j.eswa.2011.12.046.

[14] N. P. Singh and R. Srivastava, "Retinal blood vessels segmentation by using Gumbel probability distribution function based matched filter," *Comput. Methods Programs Biomed.*, vol. 129, pp. 40–50, Jun. 2016, doi: 10.1016/j.cmpb.2016.03.001.

[15] H. Zolfagharnasab and A. Naghsh-Nilchi, "Cauchy based matched filter for retinal vessels detection," *J. Med. Signals Sensors*, vol. 4, no. 1, p. 1, 2014, doi: 10.4103/2228-7477.128432.

[16] A. F. Khalaf, I. A. Yassine, and A. S. Fahmy, "Convolutional neural networks for deep feature learning in retinal vessel segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 385–388, doi: 10.1109/ICIP.2016.7532384.

[17] S. K. Vengalil, N. Sinha, S. S. S. Kruthiventi, and R. V. Babu, "Customizing CNNs for blood vessel segmentation from fundus images," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Jun. 2016, pp. 1–4, doi: 10.1109/SPCOM.2016.7746702.
[18] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1912–1923, Sep. 2018, doi: 10.1109/TBME.2018.2828137.
[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[21] P. Msonda, S. A. Uymaz, and S. S. Karaağaç, "Spatial pyramid pooling in deep convolutional networks for automatic tuberculosis diagnosis," *Traitement du Signal*, vol. 37, no. 6, pp. 1075–1084, Dec. 2020, doi: 10.18280/TS.370620.
[22] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
[23] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 636–644, doi: 10.1109/CVPR.2017.75.
[24] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004, doi: 10.1109/TMI.2004.825627.
[25] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000, doi: 10.1109/42.845178.
[26] L. C. Rodrigues and M. Marengoni, "Segmentation of optic disc and blood vessels in retinal images using wavelets, mathematical morphology and hessian-based multi-scale filtering," *Biomed. Signal Process. Control*, vol. 36, pp. 39–49, Jul. 2017, doi: 10.1016/j.bspc.2017.03.014.
[27] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, Jan. 2017, doi: 10.1109/TBME.2016.2535311.
[28] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016, doi: 10.1109/TMI.2016.2546227.
[29] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 109–118, Jan. 2016, doi: 10.1109/TMI.2015.2457891.
[30] K. Radha and Y. Karuna, "Retinal vessel segmentation to diagnose diabetic retinopathy using fundus images: A survey," *Int. J. Imag. Syst. Technol.*, pp. 1–31, May 2023, doi: 10.1002/ima.22945.
[31] F. Li, Z. Long, P. He, P. Feng, X. Guo, X. Ren, B. Wei, M. Zhao, and B. Tang, "Fully convolutional pyramidal networks for semantic segmentation," *IEEE Access*, vol. 8, pp. 229132–229140, 2020, doi: 10.1109/ACCESS.2020.3045280.
[32] G. Mittal and J. Sivaswamy, "Optic disk and macula detection from retinal images using generalized motion pattern," in *Proc. 5th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Dec. 2015, pp. 1–4, doi: 10.1109/NCVPRIPG.2015.7490071.
[33] L. Yang, H. Wang, Q. Zeng, Y. Liu, and G. Bian, "A hybrid deep segmentation network for fundus vessels via deep-learning framework," *Neurocomputing*, vol. 448, pp. 168–178, Aug. 2021, doi: 10.1016/j.neucom.2021.03.085.
[34] C. Tian, T. Fang, Y. Fan, and W. Wu, "Multi-path convolutional neural network in fundus segmentation of blood vessels," *Biocybern. Biomed. Eng.*, vol. 40, no. 2, pp. 583–595, Apr. 2020, doi: 10.1016/j.bbe.2020.01.011.
[35] S. Lian, L. Li, G. Lian, X. Xiao, Z. Luo, and S. Li, "A global and local enhanced residual U-Net for accurate retinal vessel segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 3, pp. 852–862, May 2021, doi: 10.1109/TCBB.2019.2917188.
[36] J. Wei, G. Zhu, Z. Fan, J. Liu, Y. Rong, J. Mo, W. Li, and X. Chen, "Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 292–307, Feb. 2022, doi: 10.1109/TMI.2021.3111679.
[37] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019, doi: 10.1016/j.knosys.2019.04.025.
[38] Y. Lv, H. Ma, J. Li, and S. Liu, "Attention guided U-Net with atrous convolution for accurate retinal vessels segmentation," *IEEE Access*, vol. 8, pp. 32826–32839, 2020, doi: 10.1109/ACCESS.2020.2974027.
[39] H. Wang, G. Xu, X. Pan, Z. Liu, N. Tang, R. Lan, and X. Luo, "Attention-inception-based U-Net for retinal vessel segmentation with advanced residual," *Comput. Electr. Eng.*, vol. 98, Mar. 2022, Art. no. 107670, doi: 10.1016/j.compeleceng.2021.107670.
[40] Z. Luo, Y. Zhang, L. Zhou, B. Zhang, J. Luo, and H. Wu, "Micro-vessel image segmentation based on the AD-UNet model," *IEEE Access*, vol. 7, pp. 143402–143411, 2019, doi: 10.1109/ACCESS.2019.2945556.
[41] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1427–1436, Jul. 2019, doi: 10.1109/JBHI.2018.2872813.
[42] Y. Jiang, N. Tan, T. Peng, and H. Zhang, "Retinal vessels segmentation based on dilated multi-scale convolutional neural network," *IEEE Access*, vol. 7, pp. 76342–76352, 2019, doi: 10.1109/ACCESS.2019.2922365.
[43] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard Attention Net for Automatic Retinal Vessel Segmentation," *IEEE J. Biomed. Heal. Informat.*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020, doi: 10.1109/JBHI.2020.3002985.
[44] K. Li, X. Qi, Y. Luo, Z. Yao, X. Zhou, and M. Sun, "Accurate retinal vessel segmentation in color fundus images via fully attention-based networks," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2071–2081, Jun. 2021, doi: 10.1109/JBHI.2020.3028180.
[45] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *IEEE Xplore*, vol. 7, no. 3, pp. 560–566, Apr. 2014, doi: 10.4271/2014-01-0975.
[46] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. CVPR*, Jun. 2018, pp. 7132–7141. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html
[47] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149, doi: 10.1109/CVPR.2019.00326.
[48] J. J. Staal, M. D. Abramoff, M. A. Viergever, B. van Ginneken, and M. Niemeijer, "DRIVE: Digital retinal images for vessel extraction," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
[49] M. M. Fraz, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Delineation of blood vessels in pediatric retinal images using decision trees-based ensemble classification," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 5, pp. 795–811, Sep. 2014, doi: 10.1007/s11548-013-0965-9.

K. RADHA received the B.E. degree from SVS institutions, Coimbatore, in 2017, and the M.Tech. degree from VIT University, Chennai campus, in 2020. She is currently pursuing the Ph.D. degree with VIT University, Vellore, Tamil Nadu. She has published five international book chapters and conference publications. Her area of research includes bio-medical image segmentation, diabetic retinopathy, artificial intelligence, and deep learning.

YEPUGANTI KARUNA received the B.E. degree from the CRR College of Engineering, Eluru, AP, USA, in 2007, and the M. Tech. degree from the National Institute of Technology, Tiruchirappalli, Tamil Nadu, in 2009, and the Ph.D. degree from the Vellore Institute of Technology, Vellore, India, in 2019. He has been working at VIT University as an Assistant. Professor (Sr Grade 2) at the School of Electronics Engineering since March 2010. He has published a total of 40 international journals and conference publications. His reseach interests include biomedical signal processing, medical image processing, artificial intelligence, machine learning, and deep learning.