IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 March 2023, accepted 2 April 2023, date of publication 10 April 2023, date of current version 21 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3265729

RESEARCH ARTICLE

Block Attention and Switchable Normalization Based Deep Learning Framework for Segmentation of Retinal Vessels

SABRI DEARI[1], ILKAY OKSUZ[2], AND SEZER ULUKAYA[1]
[1]Department of Electrical and Electronics Engineering, Trakya University, 22030 Edirne, Turkey
[2]Department of Computer Engineering, Istanbul Technical University, 34467 Istanbul, Turkey

Corresponding author: Ilkay Oksuz (oksuzilkay@itu.edu.tr)

This paper has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of the Scientific and Technological Research Council of Türkiye (TUBITAK) (Project No: 118C353).

ABSTRACT The presence of high blood sugar levels damages blood vessels and causes an eye condition called diabetic retinopathy. The ophthalmologist can detect this disease by looking at the variations in retinal blood vasculature. Manual segmentation of vessels requires highly skilled specialists, and not possible for many patients to be done quickly in their daily routine. For these reasons, it is of great importance to isolate retinal vessels precisely, quickly, and accurately. The difficulty distinguishing the retinal vessels from the background, and the small number of samples in the databases make this segmentation problem difficult. In this study, we propose a novel network called Block Feature Map Distorted Switchable Normalization U-net with Global Context Informative Convolutional Block Attention Module (BFMD SN U-net with GCI-CBAM). We improve the traditional Fully Convolutional Segmentation Networks in multiple aspects with the proposed model as follows; The model converges in earlier epochs, adapts more flexibly to different data, is more robust against overfitting, and gets better feature refinement at different dilation rates to cope with different sizes of retinal vessels. We evaluate the proposed network on two reference retinal datasets, DRIVE and CHASE DB1, and achieve state-of-the-art performances with 97.00 % accuracy and 98.71 % AUC in DRIVE and 97.62 % accuracy and 99.11 % AUC on CHASE DB1 databases. Additionally, the convergence step of the model is reduced and it has fewer parameters than the baseline U-net. In summary, the proposed model surpasses the U-net-based approaches used for retinal vessel separation in the literature.

INDEX TERMS Retinal vessel, segmentation, disout, block attention, switchable normalization.

I. INTRODUCTION

Symptoms of many important diseases are encountered as a result of retinal scanning. These diseases include age-related macular degeneration (AMD), diabetes, diabetic retinopathy, and glaucoma. Hyperglycemia (the presence of high blood sugar) damages blood vessels and is known to damage the eyes and causes a retinal complication of diabetes called diabetic retinopathy [1]. Hypertensive retinopathy is originated from a high level of blood pressure and this condition is a complication of diabetes. It causes damage to the retina. Moderate situations of this disorder are connected with retinal arteriolar contraction and nicking of vasculature in some cases [2]. AMD is also among the most prevalent causes of visual impairment in the world [1] Therefore by looking at the retinal vessels, diseases can be diagnosed and tracked by experts. In a form of AMD called wet AMD, there is the expansion of new and uncommon blood vessels. The vessels can leak blood or other fluids, causing scarring of the macula. Therefore, thinning, thickening, growth and other changes in the retinal vessels may be signs of the above-mentioned important diseases, the ophthalmologist can detect the diseases by looking at the changes in these retinal blood vessels, and as a result of the early diagnosis, experts may prevent blindness [1]. Thus retinal vessel separation is an essential task, and ophthalmologists need to capture even the smallest

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

changes in retinal vessels to correctly assess the disease. Manual separation of vessels requires highly skilled specialists, and it is also very time-consuming and not possible for many patients to be done quickly. For these reasons, it is of great importance to automatically separate retinal vascular structures from fundus photographs automatically. In this manuscript, we propose an improved deep learning architecture based on regularization and normalization for retinal segmentation to achieve high segmentation accuracy and precision. Our main contributions can be listed as:

- We employ a Block Feature Map Distortion (BFMD) to avoid overfitting problem effectively for retinal vessel segmentation,
- We utilize 3 different types of normalization instead of one type of normalization in the network for earlier convergence and improving generalization ability,
- We employ Global Context Informative Convolutional Block Attention Module (GCI-CBAM) to deal with thin and thick vessels for improving segmentation performance further.

The remainder of this manuscript is split into sections as follows: Section II explains the studies in the literature and the points that limit the segmentation performance. Section III details the methods proposed in this study. Section IV details the datasets, ablation experiments, implementation, and training procedure. Section V compares the proposed network with the basic U-net and the separation results with other state-of-the-art works. It also explores chosen switchable normalization weights in the network. Section VI provides a discussion and concludes the paper.

II. RELATED STUDIES

In this section, a summary of previous studies on retinal vessel segmentation will be presented. For this segmentation problem, first Fully Convolutional Networks (FCN), encoder-decoder type networks, U-net, and then improved U-net networks will be introduced.

While classical image processing techniques were initially applied for retinal vessel segmentation, deep learning techniques have been used more frequently in recent years due to their success [3]. In an early attempt, [4] used a FCN network with 2 times downsample and upsample blocks. After FCNs, encoder-decoder networks especially U-net [5] became the most popular framework for segmentation tasks. The work of [6] employed a Dual Encoding U-net (DEU-net) and it has two encoding parts, one for obtaining spatial features and the other for obtaining more content knowledge. An Attention Guided Network (AG-Net) was proposed by [7] which can preserve edge and structural information. Another U-net-based study is proposed by [8] namely, SD U-net which adds structured dropout after convolution to prevent overfitting. Then, they extended their work and proposed SA-U-net [9] by adding the Spatial Attention module to the bottleneck part of SD U-net.

Although U-net and enhanced U-net models perform well in general, the fact that fundus images are often affected by noise, the interconnectedness of vessels, irregularly illuminated samples, small samples in the databases, and especially the difficulty to distinguishing retinal blood vessels from the background makes segmentation difficult. As illustrated in Fig.1 some retinal blood vessels are very thin, and distinguishing these vessels from the background is quite challenging even to the human eye. The inability to segment these thin vessels significantly affects the sensitivity metric as it decreases the true positive value, and this causes low sensitivity values. Considering the studies in the literature, we proposed a model for improved accuracy, sensitivity, and precision.

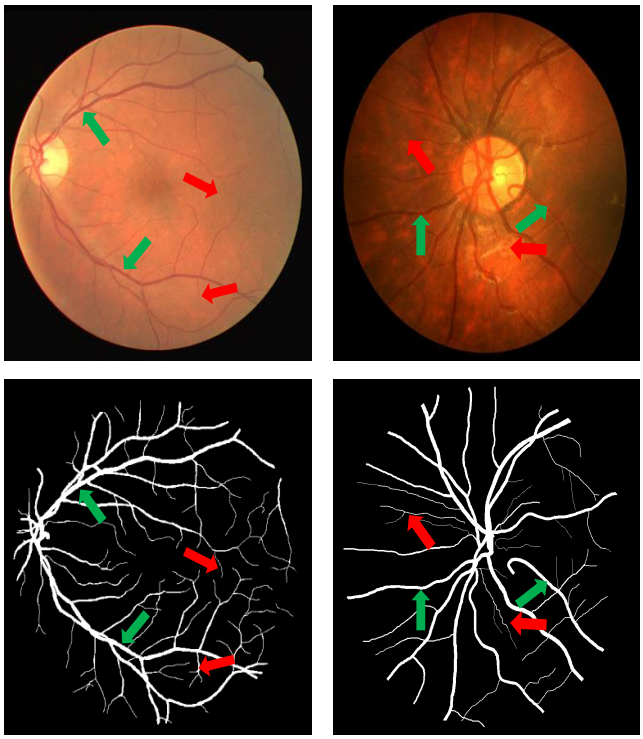

FIGURE 1. A color fundus image from the DRIVE training database indicated thin and thick vessels (left top) and corresponding ground truth image (left bottom), and a color fundus image from the CHASE DB1 training database, indicated thin and thick vessels (right top) and corresponding ground truth image (right bottom). Green arrows show thick vessels, red arrows show thin vessels.

III. METHODOLOGY

In this section, we give details on the proposed network for vascular structure separation. We propose 3 additions to the baseline U-net to perform this segmentation better:

- Inspired by the work of [10], Switchable Normalization (SN) is appended to each convolutional layer of the U-net, this way it accelerates both the convergence and generalization performance of the model and makes it adaptable for different data as it can choose different normalization types (batch, instance or layer) for different images.
- Secondly, inspired by [11], BFMD is appended to each convolutional layer part of the network. The addition

of BFMD ensures that the regularization of the model is improved. After this addition, the convolution block of the model becomes convolution-BFMD-SN-relu and this can be seen in Fig.3.

- Finally, we propose GCI-CBAM and add it between the decoder-encoder parts of the net to improve segmentation performance further. The proposed network and GCI-CBAM can be seen in Fig.2 and Fig.4, respectively.

Experimental studies show that adding SN to the model improves both the convergence of training and the performance of the model. BFMD, on the other hand, prevents overfitting, thus small sample size datasets can be well trained, and also by distorting feature maps it facilitates the model to learn better the features of the fundus, therefore, the segmentation of thin vessels. Finally, using the recommended GCI-CBAM instead of U-net's short connections further improves segmentation performance. As a result of these developments, the proposed BFMD SN U-net with GCI-CBAM outperforms state-of-the-art results in U-net-based approaches for retinal vessel separation at important metrics such as sensitivity, AUC, and accuracy for both datasets.

A. PROPOSED NETWORK ARCHITECTURE

The proposed BFMD SN U-net with GCI-CBAM is shown in Fig.2. In the encoder part of the network, there are convolution-BFMD-SN-ReLU (C-BFMD-SNR) layers at each stage and max-pooling to carry out the down-sampling operation. The number of channels is doubled in each down-sampling part. In the decoder part, the up-sampling (Conv2D Transpose) function is employed and the channels are halved, this function is followed by a combination with the GCI-CBAM module. Therefore, we add GCI-CBAM between the short connection of the network. The output of the last stage produces a resulting separation map utilizing 1×1 convolution and sigmoid function.

B. SWITCHABLE NORMALIZATION

Normalization is a very effective element used in deep learning, it is commonly used for efficiency in natural language processing and computer vision [10]. The first reason for using normalization in deep networks is faster convergence. During this stabilization, not only the training time is reduced, but also increased performance is achieved because of the regularization effect of the normalization layer [12]. Models used so far used a single normalization type in all its layers for normalization, which did not produce the best results, and choosing different normalization types for different problems made designing the model more difficult [10]. As a solution to this, Switchable Normalization (SN) is proposed by [10] that can use different normalizations in different layers and combines the statistics of three normalization methods namely, Instance Normalization (IN) [13], Batch Normalization (BN) [12], and Layer Normalization (LN) [14]. SN chooses to use three normalizations in some problems, combining three types of normalizers by switching among them and learning their importance weights [10]. SN facilitates the use of normalizers by allowing each normalization part in the network to perform its own functioning and pushing the normalization limits in deep learning [15]. In many problems, it is more efficient to use combinations of these normalizations rather than a single normalization method. For example, the combination of three normalizers is preferred for image classification and object detection problems [10]. The ability to select different normalizers enables SN to be more robust to the minibatch dimension. For instance, when a small batch size is selected, the random noise coming from batch statistics of BN will be more dominant. To avoid this, SN reduces the weights of the batch normalization and instead increases the weights of the layer normalization, and the regularization from the BN is minimized and the learning capacity is improved by the LN. In other words, SN provides the balance of learning and generalization ability [10], [15].

SN is defined as [15]:

$$\hat{d}_{ncij} = \kappa \frac{d_{ncij} - \sum_{k\in\Omega} w_k \mu_k}{\sqrt{\sum_{k\in\Omega} w'_k \sigma_k^2 + \lambda}} + \beta \tag{1}$$

where d_{ncij} and $\hat{d}_{ncij}$ are the values of original and normalized pixels in 4D input tensor, where 4D tensor is the number of samples, the number of channels, height, and width (N,C,H,W). n, c, i, and j take values between 1 to N, 1 to C, 1 to H, and 1 to W, respectively. β represents a shift parameter and κ represents a scale parameter. λ is a small constant to maintain numerical stability. Ω is a group of moment statistics. w_k and w'_k are the scalar variables representing importance ratios and are employed to weight the means (μ) and variances (σ^2). These parameters take dynamic values between 0 and 1 and are learned by backpropagation. k is used to discriminate BN, IN, and LN methods.

C. DISOUT

Models trained with a small number of samples suffer from overfitting due to sampling noise and generate poor performance on the test time [8]. Dropout is one commonly used methodology to avoid overfitting [16]. However, the recently proposed Disout [11] regularization technique proposes to learn the distortion of the feature map by reducing the empirical Rademacher complexity (ERC) of the network, rather than fixing the value of perturbation. Secondly, the features in CNNs are spatially related and in order to make the Disout more useful in CNNs, it also proposes to distort certain semantic information, as DropBlock [17] suggests dropping certain semantic information is a more useful regularization technique for CNNs, and calls it Block Feature Map Distortion (BFMD). The algorithm for training the neural network is as follows; Firstly, it calculates the feature map of the corresponding layer $F^m(d_i)$, then it generates the distortion ε and corresponding sample mask M, then it obtains distorted feature map $\hat{F}^m(d_i)$, and feed-forward the network using this distorted feature map. It does this for all layers and then backpropagates to update the weights.

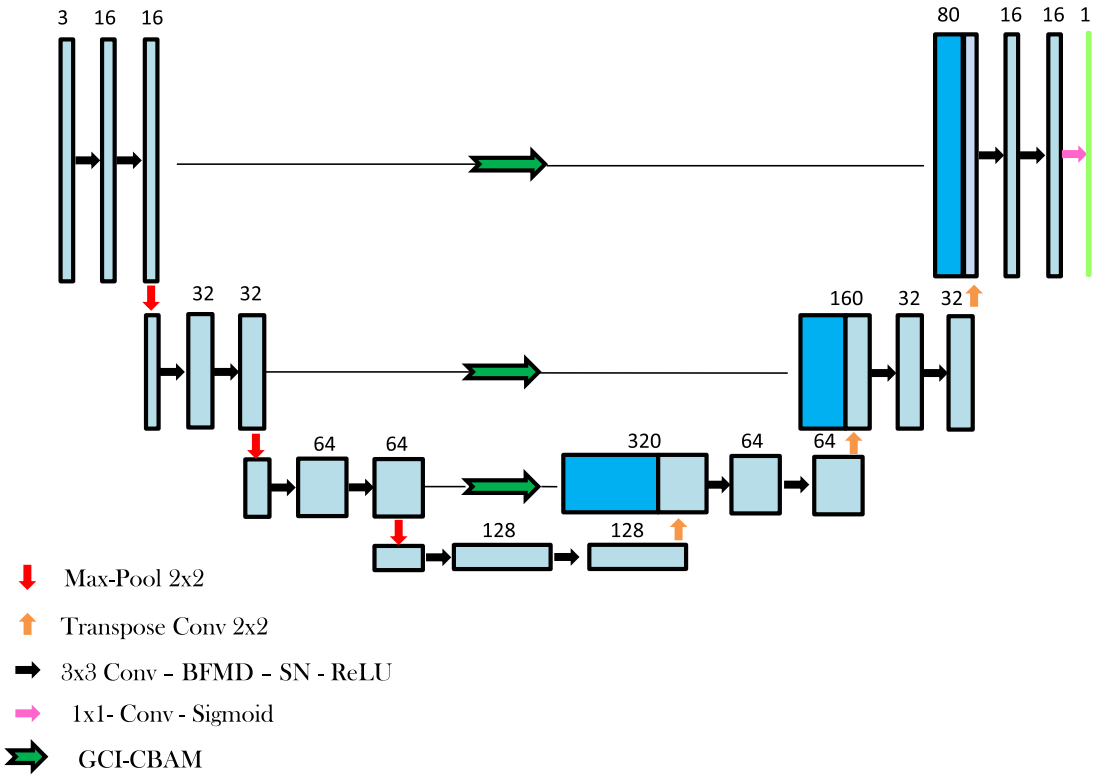

FIGURE 2. Proposed BFMD SN U-Net model with GCI-CBAM approach. Dark blue feature maps are the output of GCI-CBAM and light blue feature maps are the output of Transpose Conv.

Block Feature Map Distortion has the following formula [11];

$$\hat{F}^m(d_i) = F^m(d_i) - M_i^m \circ \varepsilon_i^m \tag{2}$$

where d_i is the input data, $F^m(d_i)$ is the feature map of the m-th layer, $\hat{F}^m(d_i)$ is the distorted feature map, M is the binary mask, and ε is the distortion utilized on the feature map, respectively. Circle denotes the element-wise product. As in the case of a fully connected layer, ERC in the $(m+1)$-th layer is employed to guide the optimization of epsilon (ε) in the m-th layer [11]. Instead of fixing the value, epsilon is automatically learned in the guide of ERC. i takes values from 1 to the total number of input samples.

D. SWITCHABLE NORMALIZATION AND BLOCK FEATURE MAP DISTORTION IN U-NET BFMD-SN BLOCK

Retinal vessel segmentation databases have a small number of samples, so deep-learning models trained for this segmentation suffer from overfitting. To make the optimization problem smoother and to increase generalization performance by making use of different normalization types, we integrate Switchable Normalization into the base U-net model, thus convolution-relu-convolution-relu block becomes convolution-SN-relu-convolution-SN-relu. We call this model SN U-net throughout the text. To solve the overfitting problem and make the model trainable even on small-size datasets, is integrated the Block Feature Map Distortion layer into the SN U-net. Thus, the new block becomes convolution-BFMD-SN-relu-convolution-BFMD-SN-relu instead of the convolution-SN-relu-convolution-SN-relu block, as shown in Fig.3. We call this model BFMD SN U-net throughout the text. Here, since BFMD and SN improve the regularization of the model in different ways, and SN also improves the convergence epochs, it prevents the prolongation of the convergence epoch that may occur as a result of the excessive distortion of the BFMD, and therefore it is very favorable to use these two layers together, in terms of both generalization performance and convergence speed.

E. GLOBAL CONTEXT INFORMATIVE CONVOLUTIONAL BLOCK ATTENTION MODULE

Channel Attention (CA) was initially used for classification networks to create channel attention maps as squeeze and excitation blocks. In [18], Convolutional Block Attention Module (CBAM) is proposed, which combines both Channel and Spatial Attention. Channel attention concentrates on what is relevant in the input picture, on the other hand, spatial attention concentrates on where the informative region is, which is supplementary to channel attention [18]. CBAM takes the channel attention and multiplies it with the input feature to get the channel attention map then takes the spatial attention of the channel attention map and multiplies it with the channel attention map to get the spatial attention map. As a result, the input feature is enhanced. Retinal blood

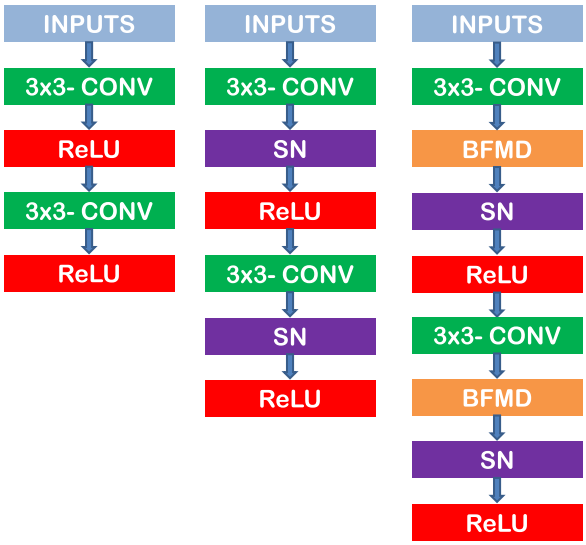

FIGURE 3. Blocks are named as U-net block (left column), SN U-net block (middle column), and BFMD SN U Net block (right column), respectively.

vessels have different thicknesses and for better segmentation of different thicknesses, we propose GCI-CBAM. Here, we use Dilated/Atrous convolution, which does not increase the number of parameters while increasing the receptive field of the model to capture retinal vessels of different sizes. Unlike channel attention used by CBAM, we use convolution instead of multilayer perceptron in channel attention. One advantage of using convolution here is computational efficiency since multilayer perceptron will calculate the new channel by considering all channels (giving weights to all of them), while 1D CNN with a kernel size of 3 will take 3 adjacent channels into account and will have fewer parameters. We select the dilation rates of the convolutions in SA and CA as 1, 2, 4, and 6. Thus, we obtain 4 different refined features and concatenate them. We call this module GCI-CBAM throughout the text. This adds global context information while filtering the image. GCI-CBAM first calculates the channel attention of the input feature at different rates, multiplies it by the input feature, and obtains 4 different channel refined features. The resulting channel-refined features are forwarded to the spatial attention module. In this Spatial attention section, channel-refined features from 4 different branches are calculated at different convolution rates and thus retinal vessels of different sizes are detected, and these spatial attention outputs are multiplied by the channel-refined features, and thus spatial and channel-refined features are obtained. The proposed GCI-CBAM is shown in Fig. 4. We add the proposed GCI-CBAM to the short links of BFMD SN U-net and call this model BFMD SN U-net with GCI-CBAM. We add GCI-CBAM in the skip connection (short links) of U-net because is a general principle for integrating attention blocks to vanilla U-net architectures. Guoheng et al. [27] propose Channel-attention U-Net and use the attention blocks in the skip connection, and Khanh et al. [28] enhance U-net with Spatial-Channel Attention Gate and use this gate in the skip connection. Performance is further enhanced as a result of the addition of the GCI-CBAM module.

IV. EXPERIMENTAL SETUP

We train BFMD SN U-net with GCI-CBAM on the DRIVE and CHASE DB1 databases, the two most common publicly accessible databases for retinal vascular separation. We prefer a combination of affine and pixel-level transforms augmentation approaches as detailed in [19]. After these data augmentations, we increase the number of images of both datasets to 240. Among affine transformations, we use 90, 180, 270-degree rotation, random rotation, shear, and crop and among pixel-level transformations, we use gamma correction, Gaussian noise, linear contrast stretching, and brightening.

A. DATASETS

We train and test the suggested approach on two publicly accessible databases, namely, DRIVE and CHASE DB1. DRIVE database consists of 40 pictures in total and usually, 20 pictures are reserved for testing and the remaining for training. In this database, 7 of the images belong to the diseased, and 33 belong to the healthy subjects. Some of these 7 images show diabetic retinopathy and some have pigment epithelium changes or both. All pictures in the dataset have a resolution of 584×565 pixels. A mask image for each retinal image is contained, showing the region of interest. We set the data to 592×592 by adding black pixels at the corners of the pictures so that the data is suitable for the proposed model, and after testing, we convert the predictions to their original size (584×565). Zero-padding is utilized for resizing step. A total of 20 images were given for training, and we increase it to 240 images using the above-mentioned data augmentation techniques.

The second publicly available dataset we use, CHASE DB1, consists of 28 images in total. In general, 20 of these pictures are employed for training and 8 are used for testing. Each sample image in the database is 999×960 pixels. We resize the images to 1008×1008 with the technique mentioned above to be compatible with the proposed model, and after testing, we resize the predictions to their original size (999×960). In this dataset, as in DRIVE, we increase 20 training images to 240 with the aforementioned augmentation techniques. We normalize the pictures in both datasets during training and testing.

B. EVALUATION METRICS

Aiming to measure the performance of the proposed system, we compare the ground truths of the data sets labeled by the experts with the prediction of the network. We calculate True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values by making a comparison on the basis of pixels and calculating Sensitivity (SE), Specificity (SP), Accuracy (ACC), and F1 Score (F1) as shown in the

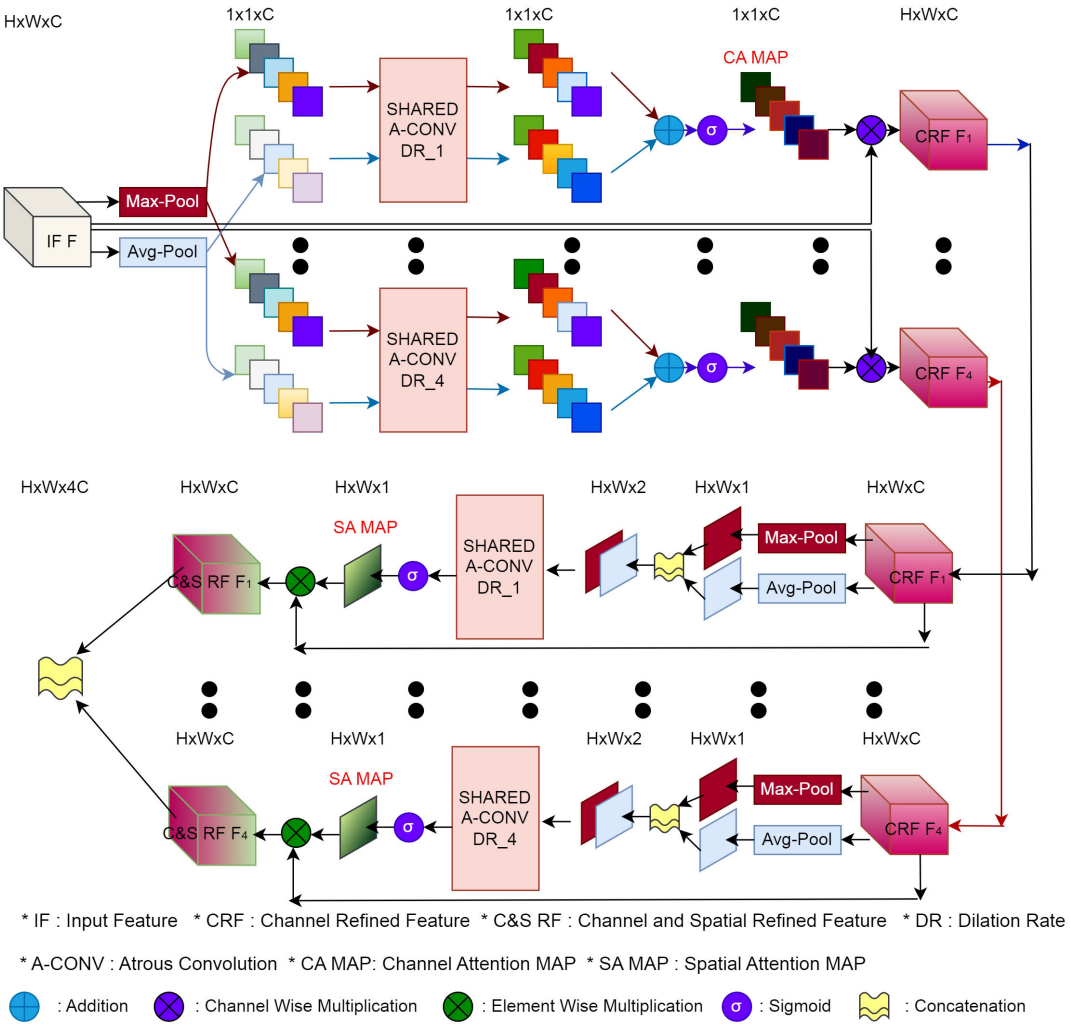

FIGURE 4. Proposed Global Context Informative Convolutional Block Attention Module (GCI-CBAM). On the left side of the module, after receiving channel attention at different dilation rates from top to bottom, it is multiplied by the input feature (IF) and a Channel Refined Feature (CRF F″) is obtained. This feature is then sent to spatial attention, and spatial attention is taken from top to bottom at different dilation rates, and after multiplying with the input (CRF F″), the Refined Feature (RF) is obtained and the information from 4 branches is combined. The first group of poolings is 2D Global Poolings applied to the spatial axis as can be seen from the top part of the figure giving an output of $1 \times 1 \times C$. The second group of the poolings is applied on the channel axis as can be seen from the bottom part of the figure giving an output of HxWx1.

following formulas, respectively.

$$SE = \frac{TP}{TP + FN} \tag{3}$$

$$SP = \frac{TN}{TN + FP} \tag{4}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \tag{6}$$

C. ABLATION EXPERIMENTS

We train and test the base U-net, SN U-net, BFMD SN U-net under the same conditions to show that Switchable Normalization (SN), the proposed GCI-CBAM, and BFMD increase segmentation performance for retinal vasculature. Moreover, we also compare the U-net with the proposed GCI-CBAM and with CBAM [18].

The performances of the models are given in Tables 1 and 2 in details. Our experimental studies on the DRIVE database showed that SN U-net is better than basic U-net by 0.6 % in accuracy, 6.48 % in sensitivity, 1.72 % in AUC, and 4.32 % in F1 score. Also, the convergence epoch decreased from 200 to 80. By adding BFMD to SN U-net, we observe an increase of 7.35 % in sensitivity, 0.27 % in AUC, and 0.42 % in F1 score as compared to SN U-net. Finally, we see a further increase in performance by adding the GCI-CBAM to the BFMD SN U-net. This increase is 0.98 % in sensitivity,

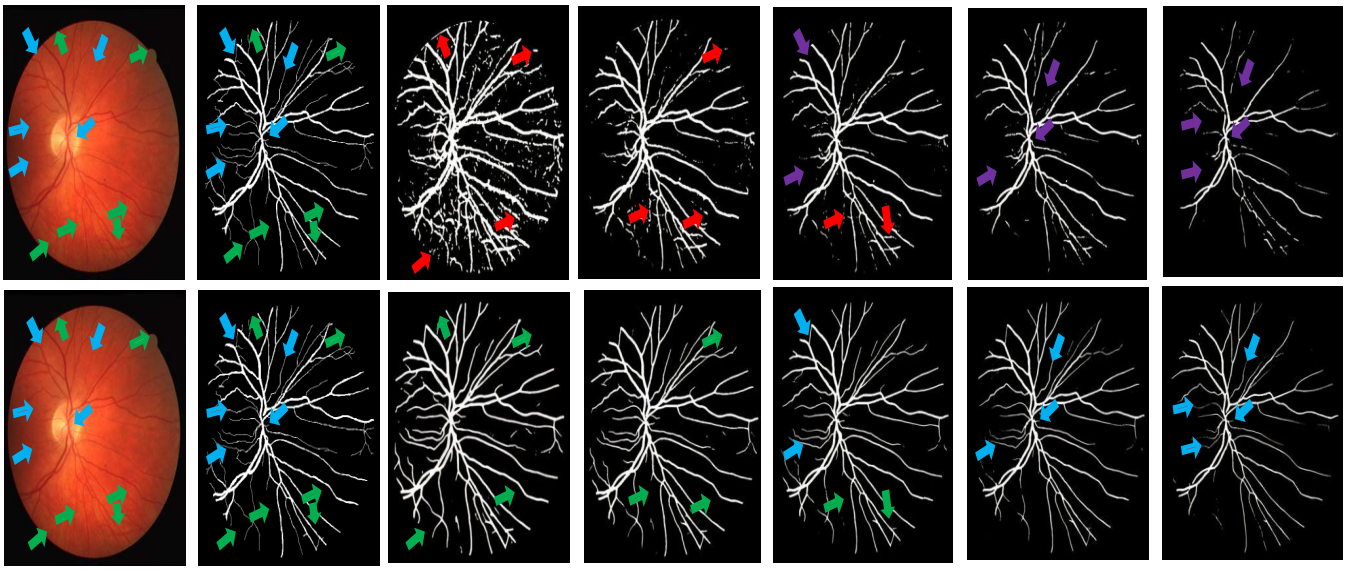

FIGURE 5. Segmentation results of U-net (top) and BFMD SN -U-net with GCI-CBAM (bottom) at different threshold values on an example image in the DRIVE test set. From left to right, the input color picture, the corresponding ground truth, and the predicted binary images for threshold values of 0.1, 0.3, 0.5, 0.7, and 0.9, respectively. Red, green, purple, and blue arrows indicate false positives, true negatives, false negatives, and true positives respectively.

0.35 % in AUC, and 0.9 % in F1 score. The segmentation result of BFMD SN U-net with GCI-CBAM on thin vessels is also shown in detail in Fig.6 and is shown with blue arrows in the segmentation outputs in Fig.7. Also, the experimental studies on the CHASE DB1 database show that SN U-net is better than basic U-net by 7.05 % in sensitivity, and by 1.91 % in F1 score. Then by adding BFMD to SN U-net, sensitivity, and F1 score are increased by 4.06 %, and 0.8 % respectively. Finally, the addition of GCI-CBAM to BFMD SN U-net positively affected the sensitivity by 2.44 % and the F1 score by 1.13 %. The segmentation outputs can be seen in Fig. 7.

D. IMPLEMENTATION DETAILS AND TRAINING PROCEDURE

For optimization, the Adam optimizer is used and we choose to employ binary cross-entropy for the loss function. The network is trained in 150 epochs. We choose the learning rate as 0.001 in the first 100 epochs and 0.0001 in the last 50 steps. The BFMD's block size is chosen as 7 for each layer, and the distortion rate as 0.10,0.15,0.20,0.25,0.20,0.15,0.10. As can be seen from Fig.8, when we select the momentum parameter of the Switchable Normalization in SN-U-net as 0.99, it fluctuates in the training step. To prevent this, we select the momentum as 0.97. We choose the epsilon of SN as 1e-3.

V. RESULTS

A. COMPARISON WITH U-NET

After adding SN to the U-net, the convergence time of the model is accelerated by 2 times. Moreover, with the addition of SN to the model, a significant improvement is observed not only in the convergence time of the system but also in the segmentation performance as can be seen in Table 1. SN U-net prefers to use 3 different normalizations, instance normalization, batch normalization, and layer normalization, and gain different advantages from each layer for this segmentation problem. The proposed SN U-net can also adapt to different segmentations by learning different normalization weights in different data. By adding BFMD to the SN U-net, we see a significant increase in sensitivity as can be seen in Table 1, which indicates that even thin retinal vessels are better segmented as well as thick vessels. Segmentation of these thin vessels in both datasets can be seen in Fig.7 and Fig.6. Performance is further improved by the addition of GCI-CBAM to the BFMD SN U-net. Finally, BFMD SN U-net with GCI-CBAM is more confident compared to U-net, for example, when we change the retinal vessel threshold, we see that the proposed model is less sensitive to this threshold value compared to U-net. So when we set the retinal vessel selection threshold to 0.1, the basic U-net will introduce too many false positives, while the BFMD SN U-net with GCI-CBAM behaves more consistently, and when we set the threshold to 0.9, we see that the basic U-net model predicts a lot of true positives as false negatives, whereas BFMD SN U-net with GCI-CBAM deal better, this way we can show that the proposed model is more robust. Segmentation results for different threshold values of U-net and BFMD SN U-net with GCI-CBAM are shown in Fig. 5.

In our network, instead of using 23 convolutional layers, we utilize 18 convolutional layers and a remarkable decrease in the number of parameters and making the network lighter and more effective. Thus, 23-layer U-net has 2,158,705 parameters, while 18-layer U-net has 535,793 parameters.

TABLE 1. DRIVE database ablation experimental results (in %). CE stands for Convergence Epoch.

Methods	ACC	SE	SP	AUC	F1	CE
U-net [5]	95.88	70.78	98.28	96.10	75.02	200
U-net+CBAM [18]	96.24	72.49	***98.52***	97.47	77.16	200
U-net+GCI-CBAM	96.26	77.06	98.11	97.56	78.33	200
SN U-net	96.48	77.26	98.32	97.82	79.45	***80***
BFMD SN U-net	96.27	84.61	97.38	98.09	79.83	150
BFMD SN U-Net+GCI-CBAM	96.42	***85.59***	97.46	98.44	80.73	150
BFMD SN U-net+GCI-CBAM+DA	***97.00***	83.00	98.30	***98.71***	***82.60***	150

TABLE 2. CHASE DB1 database ablation experimental results (in %). CE stands for Convergence Epoch.

Methods	ACC	SE	SP	AUC	F1	CE
U-net [5]	96.89	71.66	***98.59***	97.84	74.40	200
U-net+CBAM [18]	96.93	72.80	98.55	97.95	74.91	200
U-net+GCI-CBAM	97.02	76.60	98.31	98.03	75.33	200
SN U-net	97.09	78.71	98.26	97.25	76.31	***80***
BFMD SN U-net	97.09	82.77	98.00	97.79	77.19	150
BFMD SN U-Net+GCI-CBAM	97.19	85.21	97.95	98.42	78.32	150
BFMD SN U-net+GCI-CBAM+DA	***97.62***	***85.82***	98.37	***99.11***	***81.12***	150

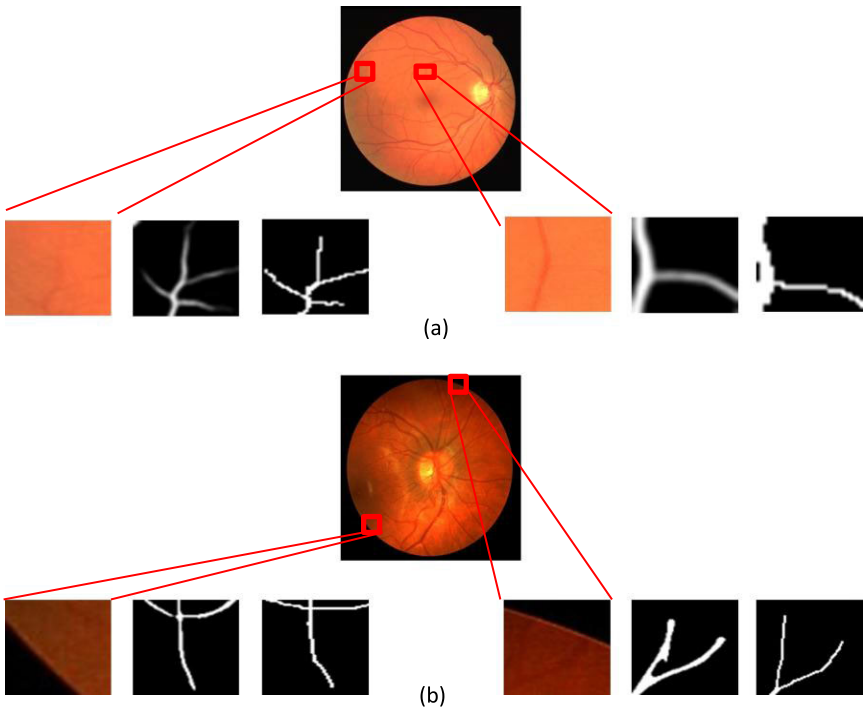

FIGURE 6. (a) Input fundus image from DRIVE test set (top), selected region (in color), predicted with BFMD SN U-net with GCI-CBAM (in binary), corresponding ground truth (in binary), from left to right (bottom), respectively. (b) Input fundus image from CHASE DB1 test set (top), selected region (in color), predicted with BFMD SN U-net with GCI-CBAM (in binary), corresponding ground truth (in binary), from left to right (bottom), respectively.

As a result of our additions to the 18-layer U-net, that is, BFMD SN U-net with GCI-CBAM has 685,069 parameters and has fewer parameters compared to the 23-layer U-net, while it has slightly more parameters than 18-layer U-net, but considering the convergence epoch and the performances of the network, it can be said that BFMD SN U-net with GCI-CBAM is much more effective. Such that, it is better by, 14.81 % in sensitivity, 2.34 % in AUC, 0.54 % in accuracy, and 5.71 % in F1 score as compared to basic U-net in Table 1.

B. COMPARISON WITH STATE OF THE ART STUDIES

As a result of these developments, as can be seen from Tables 3 and 4, the best performances are achieved in

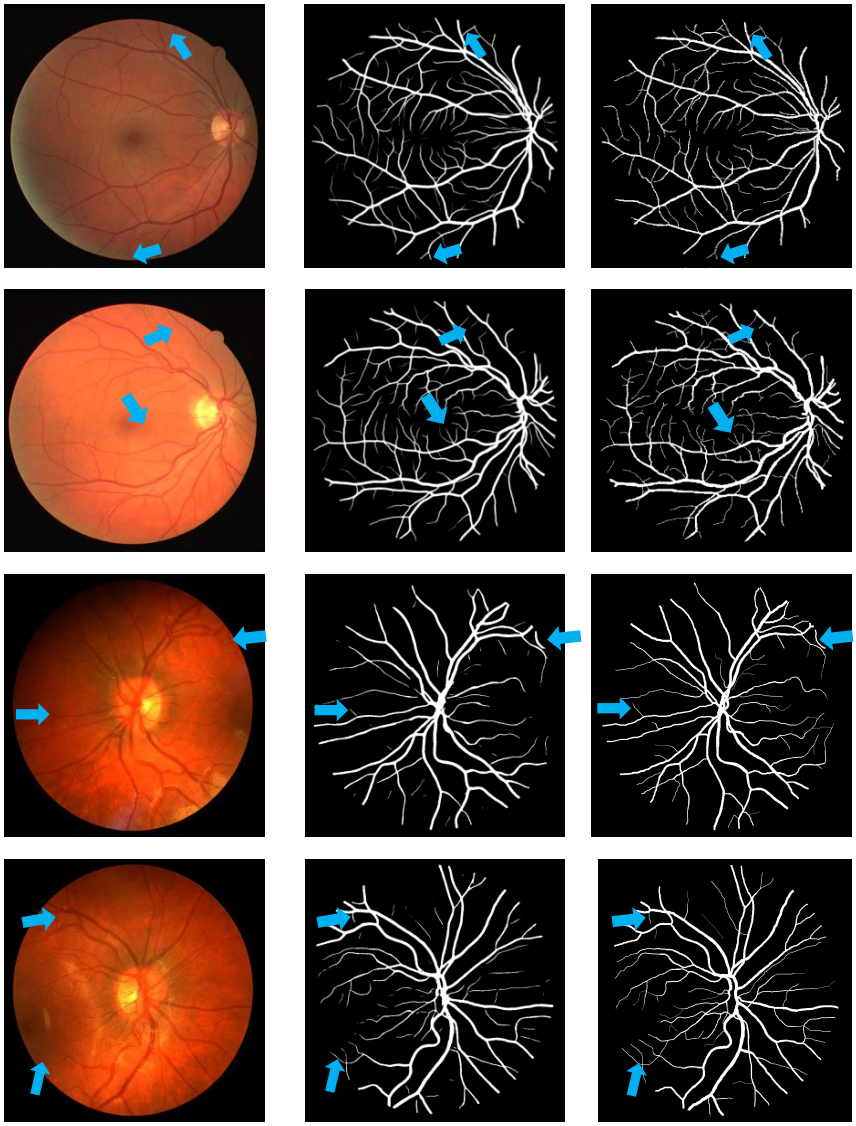

FIGURE 7. From left to right, input color fundus images, segmentation results (in binary), and corresponding ground truths (in binary), respectively. From top to bottom, the first two rows are from the DRIVE test database and the last two rows are from the CHASE DB1 database, respectively. Blue arrows indicate thin vessels.

TABLE 3. Comparison with state of the art studies on DRIVE database (in %).

Study	ACC	AUC	SE	SP	Year
Liskowski et al. [20]	95.35	97.90	78.11	98.07	2016
Orlando et al. [21]	94.54	95.07	78.97	96.84	2017
LADDERNET [22]	95.61	97.93	78.56	98.10	2018
SD U Net [8]	96.74	98.36	78.91	***98.48***	2019
AG-Net [7]	96.92	98.56	81.00	***98.48***	2019
SA-U Net [9]	96.98	98.64	82.12	98.40	2020
Zhang et al. [23]	96.15	98.15	82.13	98.07	2021
CRAU-Net [24]	95.86	98.30	79.54	N/A	2022
Su et al. [25]	95.81	98.27	77.91	98.42	2022
VG-DropDNet [26]	95.36	93.09	79.74	97.61	2022
Ours	***97.00***	***98.71***	***83.00***	98.30	2023

TABLE 4. Comparison with state of the art studies on CHASE DB1 database (in %).

Study	ACC	AUC	SE	SP	Year
Liskowski et al. [20]	96.28	98.23	78.16	98.36	2016
Orlando et al. [21]	94.58	95.24	72.77	97.12	2017
LADDERNET [22]	96.56	98.39	79.78	98.18	2018
SD U Net [8]	97.25	98.50	75.48	***98.99***	2019
AG-Net [7]	97.43	98.63	81.86	98.48	2019
SA-U Net [9]	97.55	99.05	85.73	98.36	2020
Zhang et al. [23]	96.39	98.32	80.35	97.87	2021
CRAU-Net [24]	96.59	98.64	82.59	N/A	2022
Su et al. [25]	96.59	98.65	77.38	98.75	2022
Ours	***97.62***	***99.11***	***85.82***	98.37	2023

the retinal vessel segmentation with the proposed model. Segmentation results are shown in Fig. 7 and segmentation results of thin vessels are shown in Fig. 6 in detail.

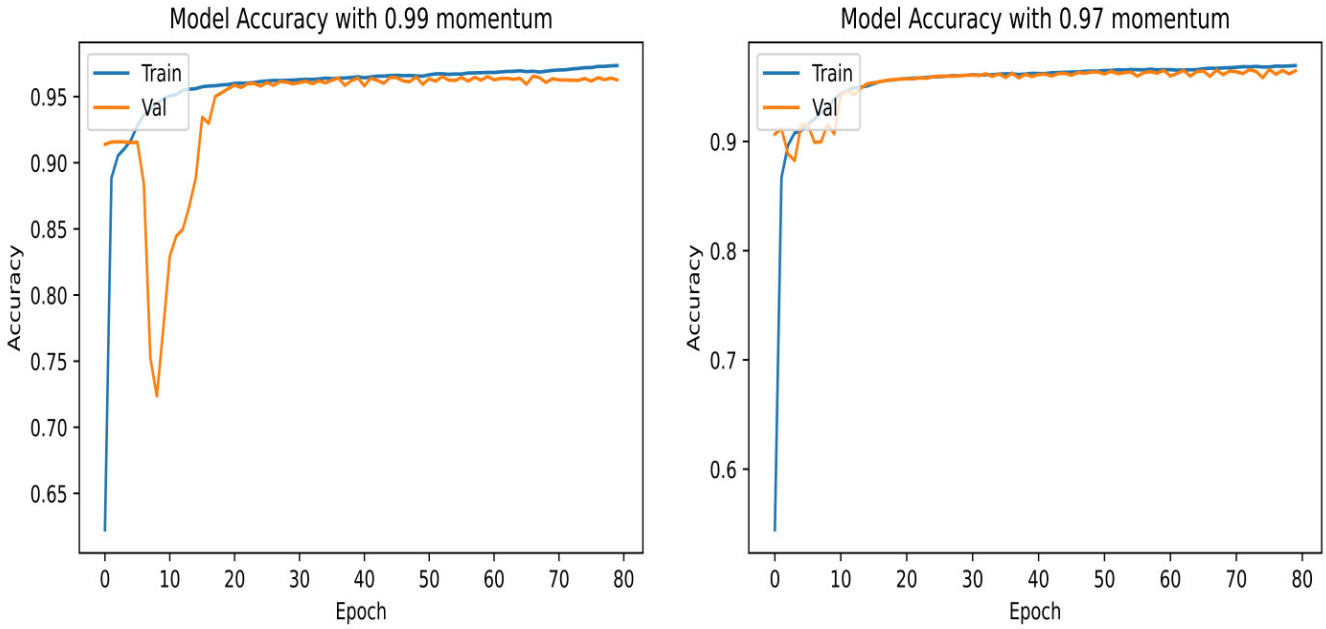

FIGURE 8. The performance of SN U-net with 0.99 momentum (left), and with 0.97 momentum (right). As seen on the left, when the Switchable Normalization moment is set to 0.99, the model fluctuates while training, and when the momentum is reduced to 0.97, as seen on the right, this fluctuation disappears.

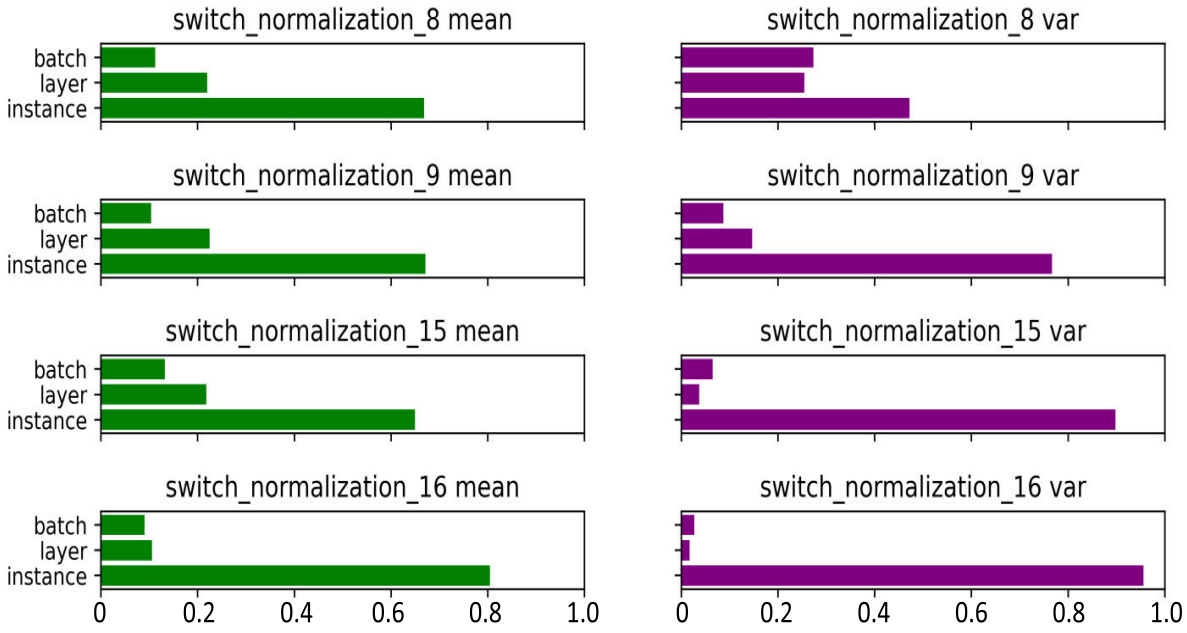

FIGURE 9. Selected mean (left) and variance (right) weights of SN in the BFMD SN U-net with GCI-CBAM at layers 8, 9, 15 and 16. Switchable Normalizations in these layers prefer to use three different types of normalization and assign mean and variance values greater than 0 to each normalization. The most weighted normalization type in these layers is instance normalization as it has the highest mean and variance.

C. SWITCHABLE NORMALIZATION WEIGHTS ON THE RETINAL VESSEL SEGMENTATION

BFMD SN U Net with GCI-CBAM, which we trained in DRIVE and CHASE DB1 datasets, chose to use 3 different types of normalization for most SNs. As illustrated in Fig. 9 switchable normalization prefers to use each of the three normalizations in the BFMD SN U-net with GCI-CBAM. Also, the mean and variance of the BN, IN, and LN, calculated by SN can be seen in Table 5. SN in the BFMD SN U-net with GCI-CBAM preferred to use instance

normalization in most of the cases. Fig.9 shows some of the SN weights selected in the proposed network.

TABLE 5. The average values of the mean and variance calculated by Switchable Normalizations for each normalization (in %). For the average mean calculation, the mean of each layer is summed and divided by the total number of layers, and for the average variance calculation, the variance of each layer is summed and divided by the total number of layers.

Normalization Type	Mean	Variance	Database
Batch	15.29	15.67	DRIVE
Layer	25.63	20.88	DRIVE
Instance	59.08	63.45	DRIVE
Batch	12.79	12.80	CHASE DB1
Layer	22.03	19.97	CHASE DB1
Instance	65.18	67.23	CHASE DB1

VI. CONCLUSION AND DISCUSSION

The precise separation of retinal vessels acts a vital role in classifying, the major causes of blindness, and disorders such as diabetic retinopathy, glaucoma, and hypertension. The retinal vessel segmentation studies conducted in recent years utilized variants of U-net architecture. We aim to address the shortcoming of the vanilla U-net network. Firstly, by adding switchable normalization (SN) to the U-net model, the convergence epoch of the model decreases, and the model becomes more flexible as SN can choose different normalization combinations for different data as it can learn different importance ratios automatically to weighted BN, LN, and IN and it tries to use the most effective normalization combinations from 3 normalizations accordingly so during the training phase it tries to find the best normalization combinations which minimize the loss function. Then, by adding BFMD to the network, the overfitting encountered in the baseline U-net is solved and generalization performance is improved. Thus, the proposed network can be trained effectively with few samples and converges quickly. Additionally, we add the Global Context Informative Convolutional Block Attention Module (GCI-CBAM) between the decoder and encoder parts of the model. With this addition, we get better feature refinement by taking into account spatial attention and channel attention. We obtain this feature refinement at different dilation rates thereby increasing the receptive field of the model to capture different sizes of retinal vessels.

In this manuscript, we propose a novel network called BFMD SN U-net with GCI-CBAM for the automatic segmentation of retinal vessels. Thanks to the block feature map distortion we added to this U-net-based network, the model can be well-trained even with a few samples, and we also improved the convergence speed of the model by adding SN to the model. We also proposed the GCI-CBAM module to suppress non-essential information and learn different sizes of information at the same time. We showed that these 3 different improvements can be used together in a network and the state-of-the-art performances are reached on DRIVE and CHASE DB1 databases among U-net based models.

One possible application area of our method can be coronary artery and lung vessel tree segmentation challenges in Computer Tomography images. The mis-classification of narrow vessels is a major problem in CT and our context information guided setup can address this issue. In the future, it can be efficient if the model can select automatically the distortion probability and block size to be based on the overfitting condition. Such an approach will make it possible to select better values for block size and distortion rate without much trial and error and loss of time. Dice loss can be used in combination with cross-entropy in the future to improve the performance further by developing a new loss function.

It is difficult to manually find the optimum value of block size and the distortion probability of the block feature map distortion in the proposed model. The block size and the distortion probability values will directly affect the performance. The small block size and distortion probability cause the model to overfit. On the other hand, selecting these values high will cause the model to underfit. Therefore, selecting these values carefully requires researchers to understand the behavior of the model in the data. Fortunately, these choices can be developed by trial and error method by looking at the performance of the model in the test dataset. However, it requires good initial values in order to prevent time lost due to trial and error.

DATA ACCESS STATEMENT

The retinal fundus data employed in the study is available at https://drive.grand-challenge.org/, and https://blogs.kingston.ac.uk/retinal/chasedb1/

CONFLICT OF INTEREST

No conflict of interest is declared.

DECLARATIONS

- The code will be available on the following GitHub repository: https://github.com/sabrid369/BFMD-SN-U-net
- Competing interests: None declared

ACKNOWLEDGMENT

This paper has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C353). However, the entire responsibility of the publication/paper belongs to the owner of the article. The financial support received from TUBITAK does not mean that the content of the publication is approved in a scientific sense by TUBITAK.

REFERENCES

[1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010, doi: 10.1109/RBME.2010.2084567.

[2] K. Narasimhan, V. C. Neha, and K. Vijayarekha, "Hypertensive retinopathy diagnosis from fundus images by estimation of AVR," *Proc. Eng.*, vol. 38, pp. 980–993, Jan. 2012, doi: 10.1016/j.proeng.2012.06.124.

[3] 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.

[4] Z. Feng, J. Yang, and L. Yao, "Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1742–1746, doi: 10.1109/ICIP.2017.8296580.
[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[6] B. Wang, S. Qiu, and H. He, "Dual encoding U-Net for retinal vessel segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 84–92.
[7] S. Zhang, H. Fu, Y. Yan, Y. Zhang, Q. Wu, M. Yang, M. Tan, and Y. Xu, "Attention guided network for retinal image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 797–805.
[8] C. Guo, M. Szemenyei, Y. Pei, Y. Yi, and W. Zhou, "SD-UNet: A structured dropout U-Net for retinal vessel segmentation," in *Proc. IEEE 19th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2019, pp. 439–444, doi: 10.1109/BIBE.2019.00085.
[9] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial attention U-Net for retinal vessel segmentation," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 1236–1242, doi: 10.1109/ICPR48806.2021.9413346.
[10] P. Luo, R. Zhang, J. Ren, Z. Peng, and J. Li, "Switchable normalization for learning-to-normalize deep representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 712–728, Feb. 2021, doi: 10.1109/TPAMI.2019.2932062.
[11] Y. Tang, "Beyond dropout: Feature map distortion to regularize deep neural networks," in *Proc. AAAI*, Apr. 2020, vol. 34, no. 4, pp. 5964–5971.
[12] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37. Lille, France, 2015, pp. 448–456.
[13] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.
[14] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.
[15] P. Luo, J. Ren, Z. Peng, R. Zhang, and J. Li, "Differentiable learning-to-normalize via switchable normalization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–18.
[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2004.
[17] G. Ghiasi, T. Y. Lin, and Q. V. Le, "DropBlock: A regularization method for convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 10727–10737.
[18] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[19] S. Deari, I. Oksuz, and S. Ulukaya, "Importance of data augmentation and transfer learning on retinal vessel segmentation," in *Proc. 29th Telecommun. Forum (TELFOR)*, Nov. 2021, pp. 1–4, doi: 10.1109/TELFOR52709.2021.9653400.
[20] 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.
[21] 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.
[22] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2018, *arXiv:1810.07810*.
[23] J. Zhang, Y. Zhang, H. Qiu, W. Xie, Z. Yao, H. Yuan, Q. Jia, T. Wang, Y. Shi, M. Huang, J. Zhuang, and X. Xu, "Pyramid-Net: Intra-layer pyramid-scale feature aggregation network for retinal vessel segmentation," *Frontiers Med.*, vol. 8, p. 2403, Dec. 2021, doi: 10.3389/fmed.2021.761050.
[24] F. Dong, D. Wu, C. Guo, S. Zhang, B. Yang, and X. Gong, "CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation," *Comput. Biol. Med.*, vol. 147, Aug. 2022, Art. no. 105651, doi: 10.1016/j.compbiomed.2022.105651.
[25] Y. Su, J. Cheng, G. Cao, and H. Liu, "How to design a deep neural network for retinal vessel segmentation: An empirical study," *Biomed. Signal Process. Control*, vol. 77, Aug. 2022, Art. no. 103761, doi: 10.1016/j.bspc.2022.103761.
[26] A. Desiani, B. Suprihatin, F. Efriliyanti, M. Arhami, and E. Setyaningsih, "VG-DropDNet a robust architecture for blood vessels segmentation on retinal image," *IEEE Access*, vol. 10, pp. 92067–92083, 2022, doi: 10.1109/ACCESS.2022.3202890.
[27] G. Huang, J. Zhu, J. Li, Z. Wang, L. Cheng, L. Liu, H. Li, and J. Zhou, "Channel-attention U-Net: Channel attention mechanism for semantic segmentation of esophagus and esophageal cancer," *IEEE Access*, vol. 8, pp. 122798–122810, 2020, doi: 10.1109/ACCESS.2020.3007719.
[28] T. L. B. Khanh, D.-P. Dao, N.-H. Ho, H.-J. Yang, E.-T. Baek, G. Lee, S.-H. Kim, and S. B. Yoo, "Enhancing U-Net with spatial-channel attention gate for abnormal tissue segmentation in medical imaging," *Appl. Sci.*, vol. 10, no. 17, p. 5729, Aug. 2020, doi: 10.3390/app10175729.

SABRI DEARI received the B.Sc. degree in electrical and electronics engineering from Trakya University. His research interests include biomedical image processing, machine learning, and deep reinforcement learning.

ILKAY OKSUZ received the B.Sc. degree in electronics engineering from Istanbul Technical University, the M.Sc. degree in electrical and electronics engineering from Bahcesehir University, and the Ph.D. degree in computer, decision, and systems science from the IMT School for Advanced Studies Lucca. He was a Research Associate with the Biomedical Engineering Department, King's College London, from 2017 to 2019. He has been an Assistant Professor with the Computer Engineering Department, Istanbul Technical University, since 2020. His research interests include medical image quality assessment, reconstruction, and analysis, with a focus on machine learning.

SEZER ULUKAYA received the B.Sc. degree in electronics engineering from Ankara University, in 2008, the M.Sc. degree in electrical and electronics engineering from Bahçeşehir University, in 2011, and the Ph.D. degree from Boğaziçi University, in 2017. He is currently an Assistant Professor with the Electrical and Electronics Engineering Department, Trakya University. He has won first place in the scientific challenge on respiratory signal processing which was sponsored by the IFMBE. His research interests include biomedical signal analysis, artificial intelligence, and image processing with applications to biometrics, lung sounds, and gait analysis.