Biomedical Signal Processing and Control 69 (2021) 102953

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Robust segmentation of vascular network using deeply cascaded AReN-UNet

Aamer Abdul Rahman [a], Birendra Biswal [b,*], Geetha Pavani P [b], Shazia Hasan [a], M.V.S. Sairam [b]

[a] Department of Electrical and Electronics Engineering, BITS-Pilani, Dubai, United Arab Emirates
[b] Center for Medical Imaging Studies, Gayatri Vidya Parishad College of Engineering (A), Visakhapatnam, India

ARTICLE INFO

Keywords:
Retinal fundus images
Convolutional neural networks
Segmentation
Cascaded AReN-UNet
Attention module
Residual module

ABSTRACT

Retinal vessel segmentation is an essential step for non-invasive diagnosis and analysis of ocular pathologies such as diabetic retinopathy, glaucoma, etc. Although several deep learning networks have been implemented for segmenting vascular maps, still further modification can be carried out on the existing deep learning networks for precise segmentation of vascular maps. This paper presents a novel cascaded AReN-UNet (Attention Residual U Network), driven by the integration of attention and residual modules. The proposed network is implemented by cascading two deep learning networks of depth 4. In the second network, each encoder receives the feature maps from the previous convolutional block. In addition to this, the feature maps of a respective convolutional block of the preceding network are also fed as input to the convolutional block of the second network. Furthermore, aggregated residual and attention modules in the cascaded AReN-UNet are used to improve convergence and stability of the network which eventually reduces the vessel breakdowns in the vascular map. The proposed model is trained and evaluated on different datasets such as DRIVE, CHASE_DB1, and one locally collected dataset. The proposed network illustrates the state-of-the-art performance by achieving an accuracy, F1 score, sensitivity, specificity, and Area Under the Curve (AUC) of 96.96%, 82.63%, 83.68%, 98.35%, and 98.67% respectively on the DRIVE dataset and 97.70%, 82.01%, 85.60%, 98.35%, and 99.01% respectively on the CHASE_DB1 dataset.

1. Introduction

A cardinal organ, the eye perceives the visual information from our surroundings and transfers it to the brain through the nervous system in impulse signals. A report by the WHO in 2019 stated that at least 2.2 billion people worldwide had been affected by vision impairment [1]. Furthermore, it was also reported that approximately 11.9 million people had moderate to severe risk of vision loss due to pathologies such as diabetic retinopathy, glaucoma, and trachoma that could have been avoided if these manifestations are detected earlier. Two studies in the US estimated that financial expenditures are amounted to $139 billion due to the prevalence of ocular diseases such as cataracts, diabetic retinopathy, glaucoma, and refractive errors [2,3]. If these are not identified in the early stages, they progress to a severe stage called neovascularization, i.e., the formation of new fragile blood vessels on the retinal surface. Traditionally, fluorescein angiography was carried out to diagnose neovascularization [4]. However, fluorescein angiography is an invasive technique that involves the injection of a fluorescent dye into the bloodstream. As a result, it is associated with several side effects such as nausea, vomiting, breathing difficulties, and in some rare cases, cardiac arrests also. The segmentation of vascular maps from retinal images by employing deep learning techniques with image processing resolves the demerits of fluorescein angiography.

Manual annotation of blood vessels from retinal fundus images for a growing number of patients is a difficult and time-consuming process. Hence, there is an essential need to implement a non-invasive automated screening system that addresses the problems faced by ophthalmologists in routine check-ups of patients. These automated screening systems can be useful for the diagnosis, screening, and treatment planning of ocular diseases such as glaucoma, age-related macular degeneration, and diabetic retinopathy [5–8]. Apart from this, they also reduce the cost burden on patients and prevent permanent blindness. Furthermore, the analysis of retinal images using deep learning networks helps to diagnose other pathologies such as cardiovascular diseases, strokes, and arteriosclerosis [9].

Retinal vessel segmentation is a challenging problem due to the

* Corresponding author at: Center for Medical Imaging Studies (CMIS), Gayatri Vidya Parishad College of Engineering, Visakhapatnam, India.
E-mail addresses: birendrabiswal@gvpce.ac.in, birendra_biswal1@yahoo.co.in (B. Biswal).

https://doi.org/10.1016/j.bspc.2021.102953
Received 12 March 2021; Received in revised form 22 May 2021; Accepted 29 June 2021
Available online 8 July 2021

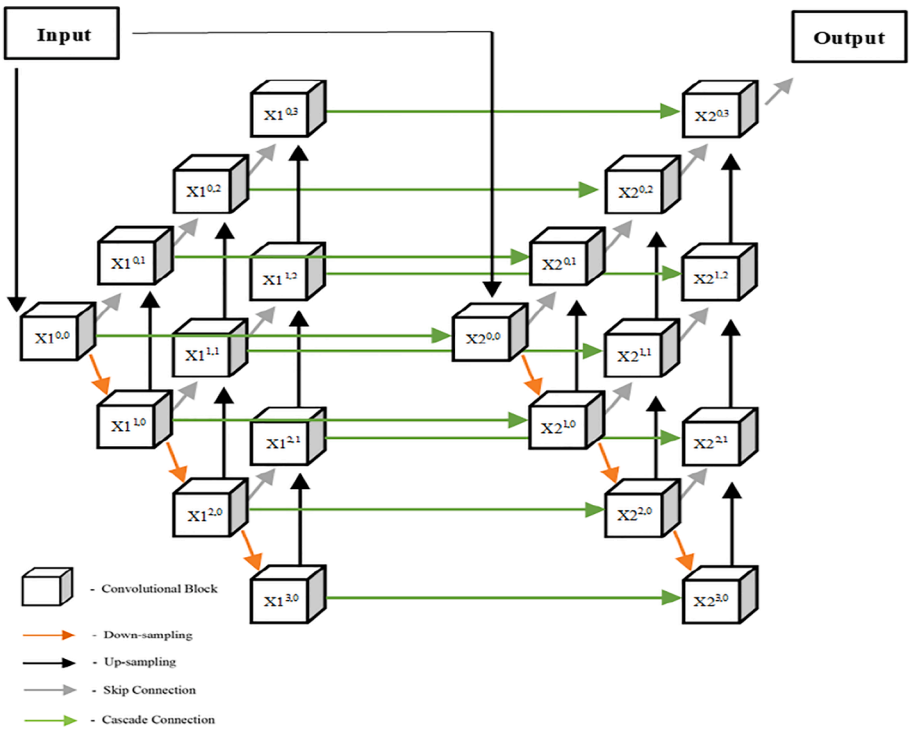

Fig. 1. Cascaded AReN-UNet architecture.

complexity and dissimilarity in vessel structures. In addition to this, the quality of images is affected by different irregularities such as variational lighting, blur, and noise. The presence of numerous anatomical structures such as the optic disk, macula, fovea, and various other abnormalities such as haemorrhages, cotton wool spots, and exudates on retinal surfaces deteriorates the algorithm's performance in segmenting the vascular network [10]. Over the past many years, multiple algorithms have been implemented for the segmentation of retinal blood vessels. These algorithms are broadly divided into two types, supervised and unsupervised methods. Unsupervised methods are rule-based segmentation algorithms such as matched filtering [11,12] multi-scale methods [13,14], vessel tracking [15,16] and morphological methods [17]. However, these algorithms lack generalization ability, which results in the formation of false edges.

With the progressive growth of computer vision, in present days, deep learning models play a vital role in medical image analysis. Among supervised models, Convolutional Neural Network's (CNN) are the most popular methods that have been developed and are used effectively to diagnose different diseases. The widespread of these models enabled several researchers to develop a network for vessel segmentation [18]. A U-Net is mainly composed of the downsampling encoder and upsampling decoder connected with inter skip connections to fill the information gap between the encoder and decoder. There are several U-Net variants [19–26] in which each has its own merits and demerits. Wu et al. [27] presented VesselNet by combining inception and residual modules. Guo et al. [28] proposed the Residual Spatial Attention U-Net that used the spatial attention block introduced in [29] and integrated it into every level of the encoder and decoder structure. Zhou et al. [30] developed a U-Net variant and is termed U-Net++, which consists of nested U-Net with dense connections between the convolutional blocks of the encoder and decoder at every depth. Consequently, the model is refined to capture each pixel of information from the target image.

Due to the imbalance of annotated data, existing models are suspected to over-fitting problems. Apart from this, retinal vessel differences also deteriorate the performance of the network. To address these issues, algorithms [31,32] used the concept of cascaded U-Nets. In these methods, the probabilistic map of the front network is fed to subsequent networks as input. The subsequent network acts as a refinement module that enhances the performance of the overall network. However, these methods suffer from large memory consumption and serious computational problems.

To circumvent the above problems in the literature survey, this paper developed a novel architecture by cascading the two identical deep learning networks of depth 4. The encoders and decoders of each network are connected with the respective encoders and decoders of the other network for transferring the feature maps accordingly. The convolutional layers in the upper blocks of the model consist of only 16 channels. As a result, the increase in memory consumption is relatively lower than the memory consumption occupied by the convolutional blocks in the lower levels. During the downsampling of the first network $(X1)$, the respective feature maps of each convolutional block are retained and stored. These retained feature maps are fed as input to the convolutional blocks of the second network $(X2)$. Consequently, $X2$uses the feature maps of $(X1)$ and enhances the performance of the overall network by extracting the feature maps of each encoder. Apart from this, integration of attention and residual modules with the cascaded network by inter skip connections for semantic segmentation of retinal vessels improved the model's spatial and structural representation and generalization abilities. The implementation of the Convolutional Block Attention Module (CBAM) in cascaded AReN-UNet with pre-activation aggregated residual layers utilizing the dropblocks with a dropout rate of 20%. This block prevents the degradation of the model and simultaneously improves the network regularization. As a result, minor and thin blood vessels of the vascular map are also segmented more accurately. Finally, it is demonstrated through experimental evaluation and comparison that the proposed modifications made to the U-Net architecture significantly improve the baseline performance. As a result, it achieves better state-of-the-art results. The contributions of this paper are as follows:

1. Implementing a novel cascading AReN-UNet that integrates with the attention and residual modules to reduce vessel misclassification. The second cascaded network improves pixel connective ability and reduces the bifurcation breakdowns in the vascular map.

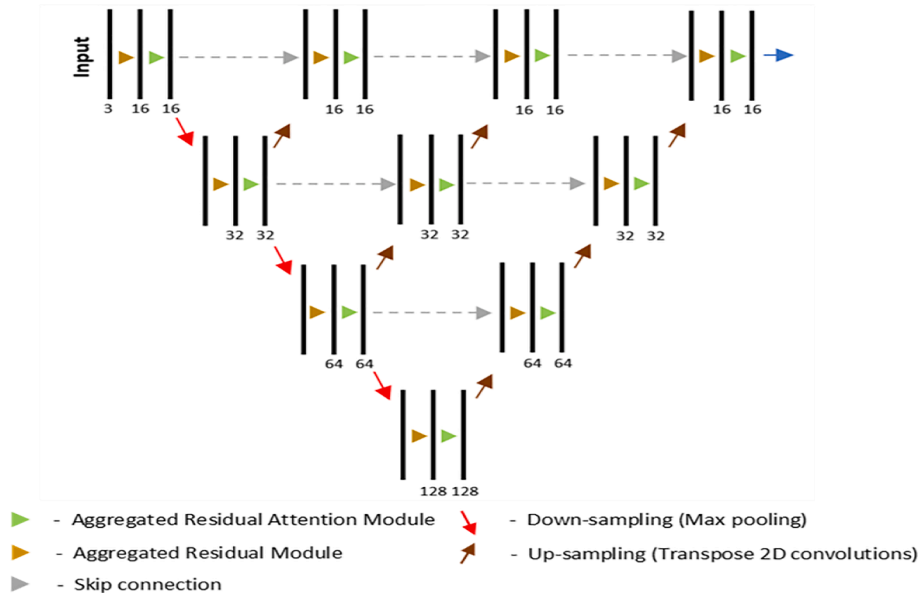

Fig. 2. Proposed Architecture of AReN-UNet.

2. The inclusion of intra and inter skip connections in integrating the attention module and cascading the two networks enables the AReN-UNet and improves its performance in capturing thin vessels.
3. The efficacy of the proposed model is estimated by using different publicly available datasets like CHASE-DB1, DRIVE datasets, and one local dataset.

The remaining part of this paper is organized as follows: Section 2 describes the architecture of cascaded AReN-UNet and its application in blood vessel segmentation. The datasets and experimental results are discussed in Section 3. Finally, the discussion and conclusions of the proposed work are depicted in Sections 4 & 5, respectively.

2. Methodology

The proposed network, cascaded AReN-UNet represented in Fig. 1, follows the encoder-decoder architecture of the U-Net. The cascaded AReN-UNet is a nested U-Net variant model with each network comprising of one encoder and three decoders as depicted in Fig. 2. As a result, two similar deep learning backbone networks are implemented and are cascaded with inter and intra skip connections.

The loss of spatial information in deep learning models is addressed by considering three decoders at each up-sampling block. This results in the formation of 3 fully segmented vascular maps at the end of each network. The convolutional layers utilize 3×3 kernels and use zero paddings with a stride of 1. The Leaky ReLU activation replaces the conventional ReLU function. This preserves the relative feature maps for semantic segmentation of blood vessels. The Leaky ReLU activation function ($\alpha = 0.3$) produced better results in comparison to the ReLU activation function. Mathematically, Leaky ReLU of a function with input x_i and output y_i is defined as:

$$y_i = \begin{cases} x_i if x_i > 0 \\ \alpha^* x_i if x_i < 0 \end{cases} \quad (1)$$

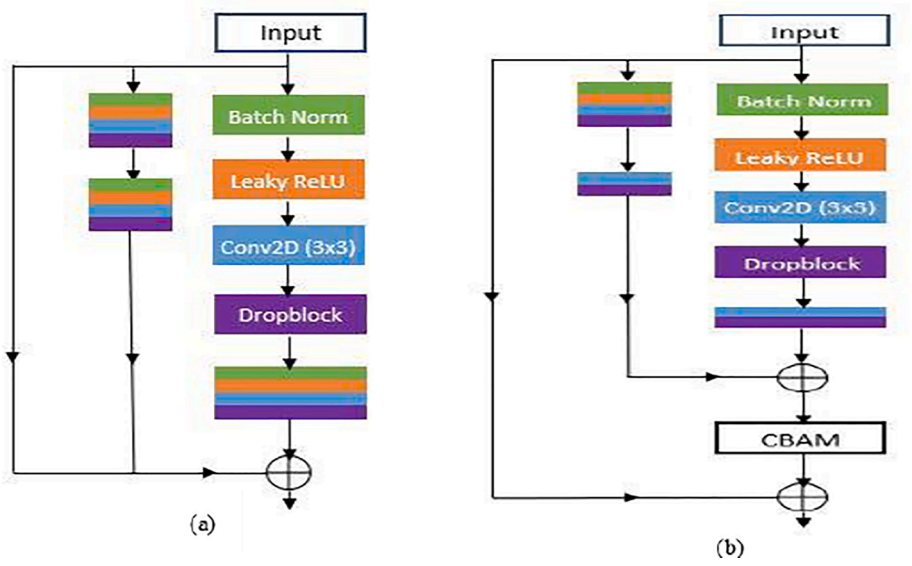

Fig. 3. (a) Pre-activation aggregated residual block (PRB) (cardinality = 2) (b) Pre-activation aggregated residual attention block (PRAB) (cardinality = 2).

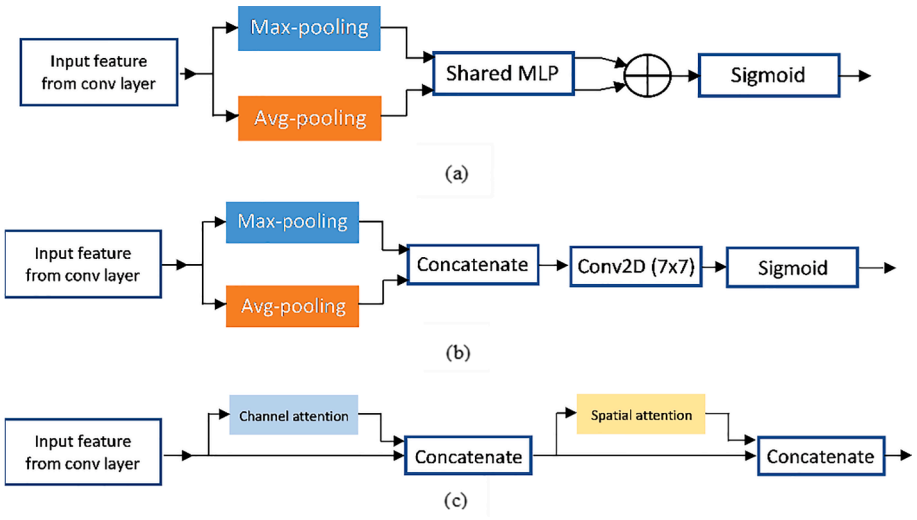

Fig. 4. (a) Channel attention block (b) Spatial attention block (c) Convolutional block attention module (CBAM).

2.1. Cascaded AReN-UNet

To circumvent the problems faced by U-Net variants, we adapted the mechanism of cascaded U-Nets [33], i.e., the output of the first U-Net is fed into a subsequent U-Net, which serves as a refinement model and aims to enhance the structural representation of the vascular map. In addition to this, the cascaded model is combined with attention modules and driven by preactivated residual layers, which will reduce the misclassification rate of retinal vessels.

During the forward pass of the cascaded AReN-UNet, the image is initially fed into $X1$, and the respective feature maps of each convolutional block of $X1$ are saved and fed to $X2$. The feature maps produced by $X1^{0,0}$ are then concatenated with the input image and fed into$X2^{0,0}$. Further, the output feature maps produced by $X2^{0,0}$ are concatenated with the feature maps of $X1^{0,1}$ and fed into $X2^{0,1}$. As a result, each convolution block in the second network concatenates feature maps of the front network and preceding block for segmenting the vessels. This yields better results by fine refinement of vessel structures.

2.2. Pre-activation aggregated residual block

The integration of residual blocks [34] circumvents the degradation problem using a skip connection for residual learning. These pre-activation residual structures improve the regularization ability of the model and ease the optimization process compared to traditional residual structures. The aggregated residual network (Res-NeXt) introduced by Xie et al. showed that increasing the cardinality of the residual connections was more effective than increasing the width or depth of the network [24]. Cardinality is the number of residual transformations to be aggregated and resultant sum of individual residual paths, $F(i)$defined as:

$$F(i) = \sum_{j=1}^{C} T_j(i) \tag{2}$$

Where C is the cardinality and $T_j(i)$ is the individual residual path for input i. Furthermore, applying the residual function yields the output y as:

$$y = \sum_{j=1}^{C} T_j(i) + i \tag{3}$$

In addition to this, dropblocks [35,36] are employed to avoid redundancy errors in AReN-UNet architecture. The dropblock layer effectively improves the regularizing capabilities of convolutional layers, further improving the overall performance network. Fig. 3(a). illustrates the pre-activated aggregated residual blocks with cardinality 2 and Fig. 3 (b). represents the pre-activated aggregated residual attention blocks with cardinality 2.

2.3. Pre-activation aggregated residual attention module

The Convolutional Block Attention Module (CBAM) [29] is a type of attention module that can be integrated with the feed-forward convolutional neural networks to improve their representation ability. Different types of attentional modules along with conventional attentional modules are represented in Fig. 4. Fig. 4(c) illustrates the CBAM that consists of a sequence of channels and spatial modules that have been designed to effectively emphasize the relative features. This eventually suppresses irrelevant features along the channel and spatial axis respectively. CBAM processes an intermediate feature map ($F \in R^{C*H*W}$). This sequentially computes the 1-dimensional channel attention map ($M_c \in R^{C*1*1}$) and the 2-dimensional spatial attention map ($M_s \in R^{1*H*W}$). These attention feature maps are obtained by Eq. (4).

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F' \tag{4}$$

where $\otimes$ mbolizes element-wise multiplication. During multiplication, the values of channel attention are broadcasted along the spatial dimension, vice versa.

The channel attention module, as illustrated in Fig. 4. (a), is designed to improve the model's representation power along the channel dimension and emphasizes focus on relevant information to realize the image by employing both max-pooling and average-pooling to compute F^c_{max} and F^c_{avg} respectively. These are then fed into a shared Multi-Layer Perceptron (MLP), and the respective concatenated outputs are passed through an activation function, sigmoid layer, to form the channel attention map. Channel attention feature maps are mathematically represented as:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{5}$$

3. Experimental results

3.1. Datasets

The proposed network is trained and tested on different publicly available datasets such as DRIVE [37] and CHASE_DB1 [38,39]. Coloured fundus image databases were used to evaluate the proposed models and test their clinical applicability.

Table 1
Dataset information.

Dataset	DRIVE	CHASE_DB1
Number of Images	40	28
Original Size	584 × 565	999 × 960
Resized	592 × 592	1008 × 1008
Train/Val./Test Split	18/2/20	19/2/7
K-fold Cross-Validation	None (Official Split)	K = 4

The DRIVE dataset includes 20 training and 20 testing images of size 584 × 565 pixels. They were collected as part of a diabetic retinopathy screening program in the Netherlands. Only 7 of the images in this dataset consist of pathological manifestations. The CHASE_DB1 dataset, are collected from both the eyes of 14 school children in England, consists of 28 retinal fundus images of size 999 × 960 pixels. In addition to this, a local dataset collected by conducting medical camps in rural areas of Andhra Pradesh, India, is used for testing the performance of the proposed algorithm. The segmented results of the local dataset are evaluated by a team of ophthalmologists from our collaborating hospital "Gayatri Vidya Parishad Institute of Healthcare and Medical Technology." Initially, the network is trained with the training dataset of the DRIVE dataset. To maintain homogeneity and fair comparison, CHASE_DB1 is used for training with a 4-fold cross-validation technique. Initially from all the input images, the green channel is extracted and subjected to CLAHE and gamma correction as a prior pre-processing step. Thus, the edges of the blood vessels in the fundus image can differentiate more precisely. Table 1 represents the information of images collected from different datasets to train the network.

3.2. Implementation details

Initially, the retinal images are resized to 592 × 592/ 1008 × 1008 and subjected to data augmentation that enlarges the training dataset. The Adam optimizer is used with the learning rate initialized from 0.001, and the models were trained over 200 epochs with the early stopping regularization technique. The size of the dropblocks was set to 7 with a dropout rate of 0.2. The implementation of the proposed algorithm is performed using publicly available cloud GPU services, Google Collaboratory, and Kaggle notebooks with Tensorflow 1.14 framework and Keras library. For the DRIVE dataset, the models were trained on images with a batch size of 2. While the training on the CHASE_DB1 dataset, the batch size was set to 1 due to computational

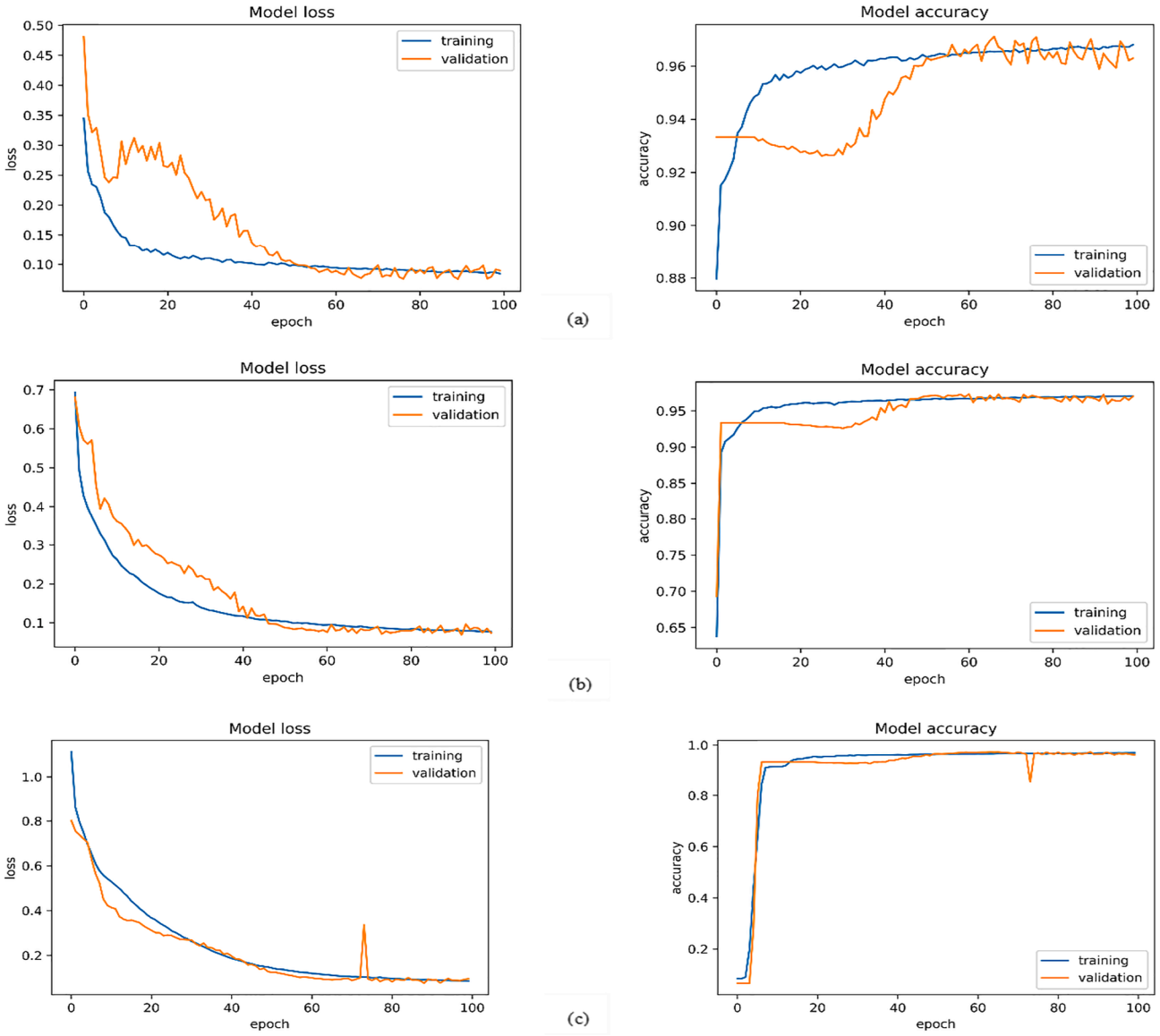

Fig. 5. Performance of different models during the training process of (a) U-Net++ (backbone) (b) AReN-UNet (c) Cascaded AReN-UNet.

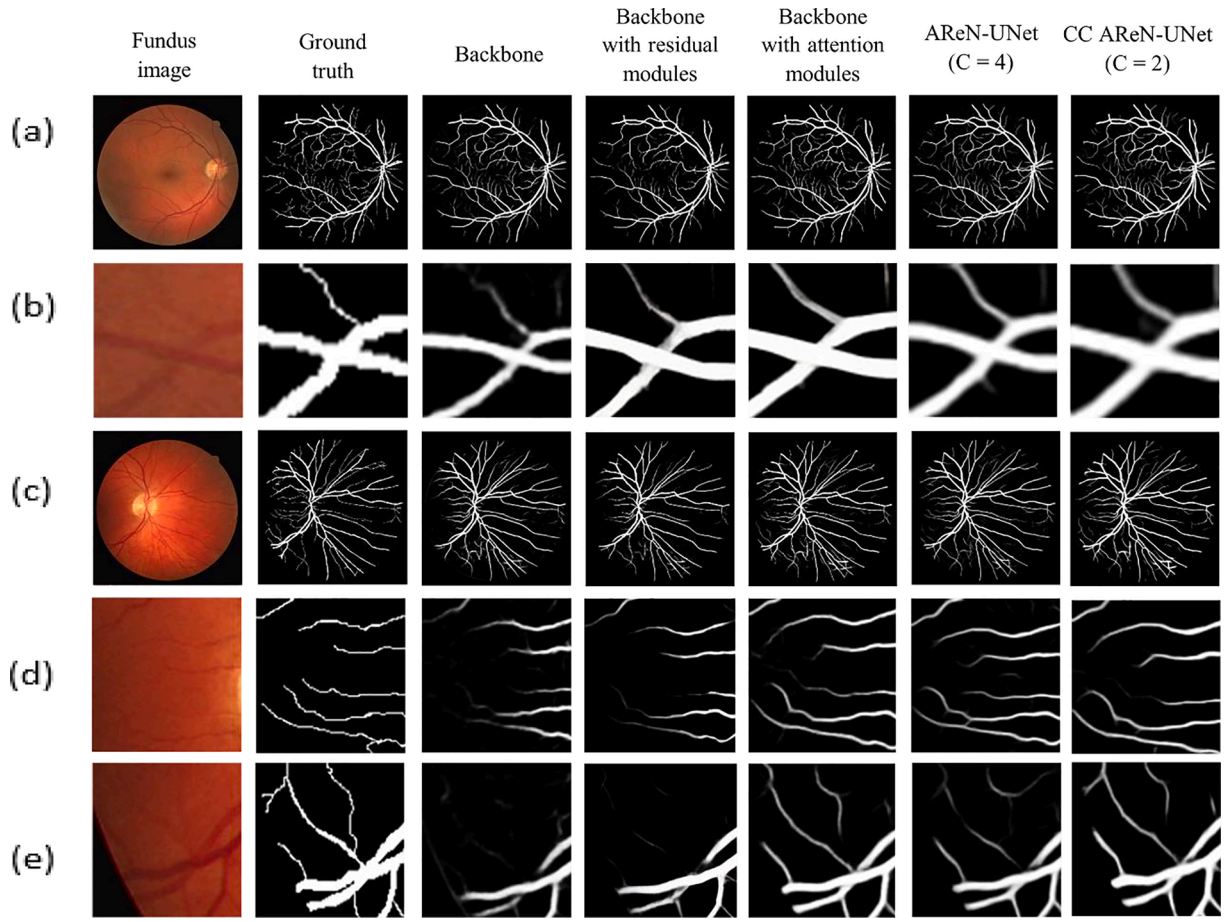

Fig. 6. Visualization of vascular maps of retinal images from different datasets (a) & (c) are the rows with sample input images from DRIVE [37] and CHASE_DB1 [38,39] datasets along with ground truth and segmentation results obtained from different module of proposed model (b) A small rectangular region of (a) to show the significance of each module in segmenting the vessels. (d) & (e) are the rectangular regions of (c) to show the significance of each module in segmenting the vessels.

limitations. Similarly, the cardinality hyperparameter was limited to 2 while training the cascaded AReN-UNet on the DRIVE dataset. However, on the CHASE_DB1 the cardinality of the model was set to 1. The training performance of the proposed algorithm is depicted in Fig. 5. The fully trained cascaded AReN-UNet segments the vascular maps with an average accuracy of 96.9%. The segmented vascular maps of retinal images from different datasets like DRIVE, CHASE_DB1 are illustrated in Fig. 6. Fig. 5 describes the performance metrics such as accuracy and loss values of different U-Net variants along with the proposed model. These values are monitored and plotted during the training of the model. From Fig. 5, it is evident that the training curves of AReN-UNet and cascaded AReN-UNet depict relatively lower deviation than the U-Net model in evaluating training and validating data. This indicates the high stability of the proposed cascaded AReN-UNet over the baseline U-Net model. At the end of the training, an average dice coefficient of 0.87 is observed, and the loss value is 0.2 with a specificity of 0.98 and sensitivity of 0.8245.

3.3. Evaluation metrics

The Qualitative and Quantitative analysis of the proposed model is performed to depict the efficacy of the proposed network. Different quantitative evaluation metrics like sensitivity, specificity, F1 score, accuracy, and area under the ROC curve (AUC) were evaluated and compared with baseline models.

$$Sensitivity = TP/TP + FN$$

Table 2
Ablation analysis of proposed model.

Method	F1 Score	Sensitivity	Specificity	Accuracy	**AUC**
Base U-Net	0.7675	0.7030	0.9876	0.9627	0.9454
Backbone (Nested – UNet)	0.8021	0.7513	0.9838	0.9642	0.9712
Backbone with residual module	0.8079	0.7576	**0.9886**	0.9676	0.9833
Backbone with attention module	0.8169	0.8163	0.9783	0.9667	0.9851
Cascaded backbone with residual module	0.8132	0.8037	0.9834	0.9684	0.9846
Cascaded backbone with attention module	0.8214	0.8227	0.9816	0.9682	0.9855
AReN-UNet (C = 2)	0.8217	0.8307	0.9816	0.9684	0.9859
Proposed Cascaded AReN-UNet (C = 2)	0.8234	**0.8368**	0.9812	0.9686	0.9861
Proposed AReN-UNet (C = 4)	**0.8263**	0.8245	0.9835	**0.9696**	**0.9867**

The bold values represent the highest values of each performance metric, when compared with previously existing algorithms.

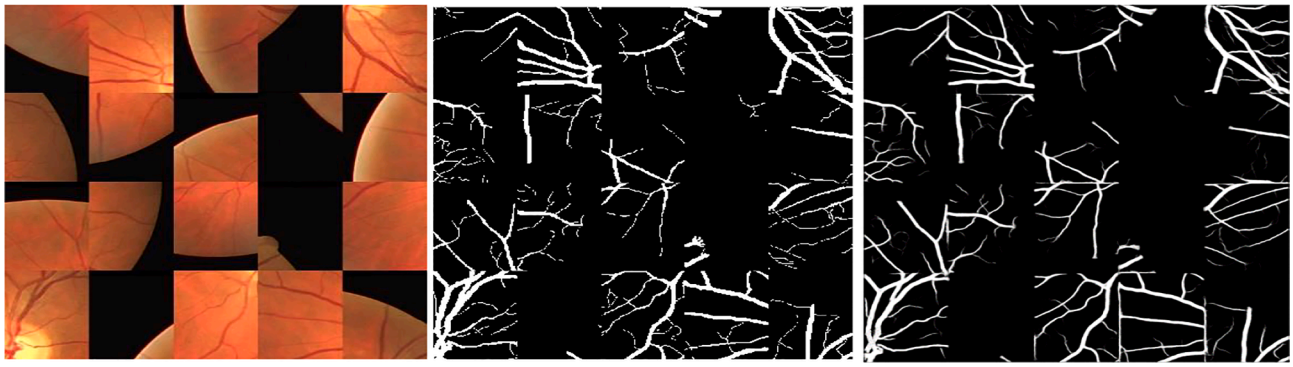

Fig. 7. The performance of proposed network on shuffled patches input image [37] (Input image, labelled image, Cascaded AReN-UNet output).

$$Specificity = TN/TN + FP$$

$$Accuracy = (TP + TN)/(TP + TN + FN + FP)$$

Where TP, FP, FN, and TN indicate True Positives, False Positives, False Negatives, and True Negative pixels, respectively.

3.4. Segmentation results

The supremacy of the proposed network is illustrated in Table. 2. Table 2 represents the ablation studies of the proposed networks over the baseline models. The model is trained and monitored by evaluating its performance metrics like Dice Coefficient, Loss value, Sensitivity, and Specificity. Fig. 6 depicts the segmentation results obtained by each model and the ground truth. A small rectangular region of the input image is considered and enlarged as shown in the second, fourth, and fifth rows of Fig. 6 which contain major vessels and minor vessels respectively. The subfigures in Fig. 6 (b), (d) & (e) illustrate the significant impact of cascaded AReN-UNet and inter-network skip connections. As a result, the proposed cascaded AReN-UNet model can not only extract accurate thin vessels but also eliminate misclassified areas than other models. Moreover, it can connect the cracked branches and enhance the vascular structure of the retinal image. Furthermore, to evaluate the efficacy, the networks are tested using input images by shuffling the patches randomly. The results are shown in Fig. 7 and achieved an accuracy of 94%. This evaluation justifies the supremacy of the proposed network over the earlier existing networks. In addition to this, the optimization of hyperparameters in the training process yields better results by achieving an F1 score of 82.3%. Furthermore, a local dataset is used to evaluate the performance of the proposed algorithm and attained an accuracy of 94.5% with specificity and sensitivity values of 95% and 83% respectively. The segmentation results of the sample image from the local dataset are depicted in Fig. 9.

3.5. Comparison of the proposed model with baseline models and earlier existing models

Initially, ablation experiments were carried out on the DRIVE dataset to evaluate the significance of individual features obtained from the attention module, residual module, and the cascaded networks. The

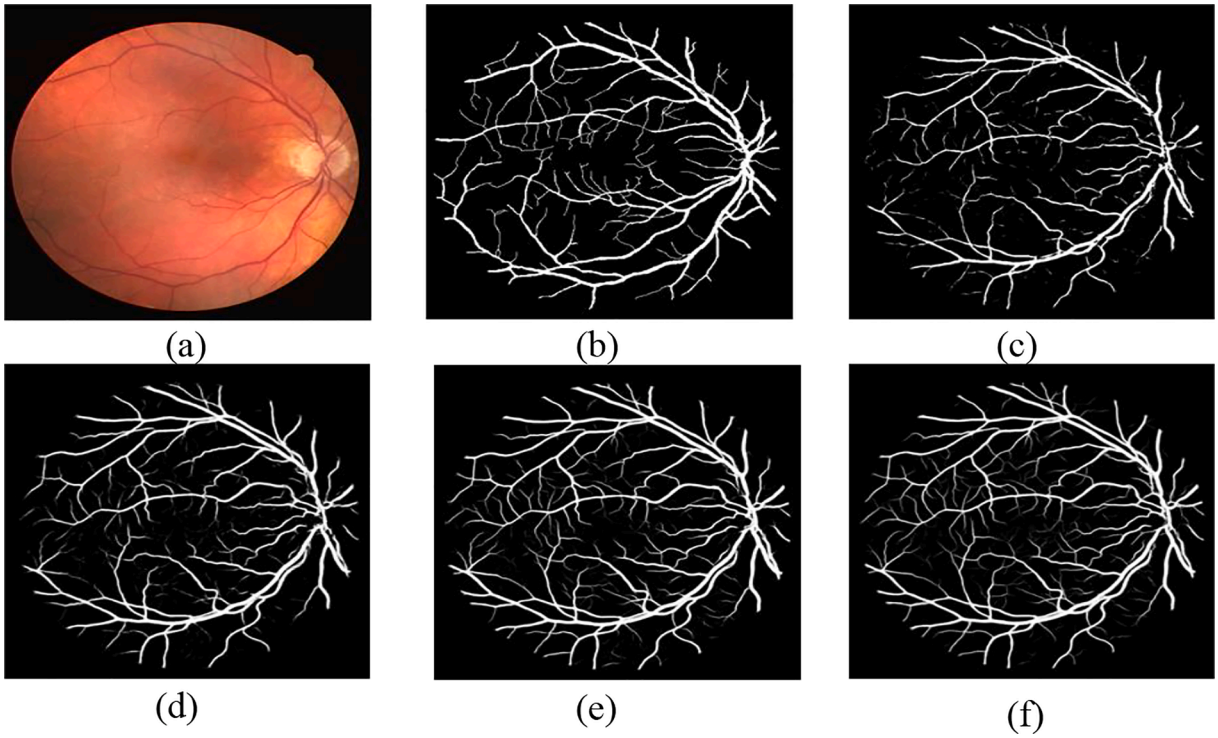

Fig. 8. (a) Fundus image (b) Ground Truth of vascular map, Vessel segmentation outputs of different models (c) U-Net (d) U-Net++ (e) VesselNet (f) Proposed cascaded AReN-UNet.

Table 3
Performance comparison with the state-of-the-art on the DRIVE dataset [37]

Method	F1 Score	Sensitivity	Specificity	Accuracy	AUC
Unsupervised methods					
Al-Diri et al. [40]	–	0.7282	0.9551	0.9258	–
You et al. [11]	–	0.7410	0.9751	0.9434	–
Fraz et al. [38]	–	0.7152	0.9768	0.9430	–
Roychowdhury et al. [41]	–	0.7395	0.9782	0.9494	–
Fan et al. [42]	–	0.7360	0.9810	0.9600	–
Supervised methods					
U-Net [18]	0.8012	0.7677	0.9857	0.9666	0.9789
R2UNet [43]	0.8146	0.7792	0.9813	0.9556	0.9784
LadderNet [44]	0.8202	0.7856	0.9810	0.9561	0.9793
BTS-DSN [36]	0.8237	0.7963	0.9800	0.9566	0.9806
Vessel-Net [29]	–	0.8038	0.9802	0.9578	0.9821
IterNet [31]	0.8218	0.7791	0.9831	0.9574	0.9813
HAnet [37]	0.8293	0.7991	0.9813	**0.9581**	0.9852
Tang et al. [38]	0.8155	–	–	0.9551	0.9769
RSAN [28]	0.8222	0.8149	0.9839	0.9691	0.9855
NFN+ [33]	**0.8295**	0.7996	0.9813	0.9582	0.9830
EEA Unet [45]	0.9770	0.7918	0.9708	0.9577	–
SAT-Net [46]	0.8174	0.8117	**0.9870**	0.9684	0.9822
Wang et al. [41]	0.7863	0.8060	0.9869	0.9512	0.9748
Proposed AReN-UNet(Cardinality = 4)	0.8263	0.8245	0.9835	**0.9696**	**0.9867**
Proposed Cascaded AReN-UNet (Cardinality = 2)	0.8234	**0.8368**	0.9812	0.9686	0.9863

The bold values represent the highest values of each performance metric, when compared with previously existing algorithms.

Table 4
Performance comparison with the state-of-the-art on the CHASE_DB1 dataset [38,39]

Method	**F1 Score**	**Sensitivity**	**Specificity**	**Accuracy**	**AUC**
Unsupervised methods					
Fraz et al. [30]	–	0.7224	0.9711	0.9469	0.9712
Roychowdhury et al. [31]	–	0.7201	0.9824	0.9530	0.9532
Fan et al. [42]	–	0.6570	0.9730	0.9510	–
Supervised methods					
U-Net [18]	0.7783	0.8288	0.9701	0.9578	0.9772
R2UNet [34]	0.7982	0.7756	0.9820	0.9634	0.9815
LadderNet [44]	0.7895	0.7856	0.9799	0.9620	0.9772
BTS-DSN [36]	0.7983	0.7888	0.9801	0.9627	0.9840
Vessel-Net [29]	0.7911	0.7819	0.9807	0.9624	0.9758
IterNet [31]	0.8146	0.7715	**0.9919**	0.9782	**0.9915**
HAnet [37]	0.8191	0.8239	0.9813	0.9670	0.9871
RSAN [28]	0.8111	0.8486	0.9836	0.9736	0.9894
NFN+ [33]	–	0.8003	0.9880	0.9688	0.9894
EEA Unet [45]	0.6453	0.6457	0.9653	0.9340	–
SAT-Net [46]	0.7911	0.8340	0.9868	0.9739	0.9855
Proposed AReN-UNet (Cardinality = 4)	**0.8201**	0.8420	0.9838	**0.9770**	0.9901
Proposed Cascaded AReN-UNet (Cardinality = 2)	0.8154	**0.8560**	0.9835	0.9763	0.9896

Table 5
Comparison of different models in terms of trainable parameters.

Method	**Number of parameters (in millions)**
U-Net [18]	7.76 M
R2UNet [43]	1.04 M
UNet++ [30]	9.04 M
Proposed AReN-UNet (Cardinality = 4)	4.17 M
Proposed Cascaded AReN-UNet (Cardinality = 2)	4.88 M

proposed model performance metrics were evaluated by combining the residual module and attention module in the first step. In the second step, the same experiment was conducted by combining the residual module with the cascading network. Finally, the attention module is integrated with the cascading network to compare the result of the performance metrics such as F1 score, Accuracy, Specificity, Sensitivity, and AUC. The result of the performance metrics is depicted in Table 2.

From Table 2, It is observed that the individual integration of both residual and attention modules with the nested U-Net (backbone) improved the performance of the model to a greater extent. As a result, the residual and attention modules improved the AUC by 1.21% and 1.45% respectively. In addition to this, the individual integration of these modules increased the F1 score by 0.53% and 0.45% respectively. Furthermore, the integration of these modules to cascading network, i.e., the proposed network increased the AUC by 1.59%. These improvements in the performance metrics depicts the robustness of the proposed model over nested U-Net (backbone). The segmentation results shown in Fig. 6 depict the impact of attention and residual modules in the proposed network. The combined integration of attention and residual modules with the nested U-Net (backbone) can extract more relevant information from low-resolution areas of the fundus image yielding an accurate segmentation of the vascular map. Finally, from Fig. 6, it is evident that the precise segmentation result is obtained by the proposed cascaded AReN-UNet enabling the improvement of different performance metrics by integrating with the attention module, residual module, and cascading networks.

For a better comparison with previously existing models, the convolutional networks [18] are considered, and all the models are trained with identical parameters like Adam optimizer, maximum epochs, learning rate, etc. Fig. 8 describes the segmentation results of different models such as U-Net, U-Net++, VesselNet along with proposed cascaded AReN-UNet. These results illustrate the significance of the proposed model in reducing the breakdowns of vessels and increasing the true positive rate. The proposed model can efficiently segment the minor vessels that are being neglected in U-Net, U-Net++, etc. The developed cascaded AReN-UNet integrated with aggregated residual and attention modules reduces the generalization problems and further increases the representation ability of the network. The single AReN-UNet and the cascaded AReN-UNet outperformed the existing models by achieving an Area Under the Curve (AUC) of 98.7%. At the same time, the sensitivity of the proposed models exceeded the U-Net and its variants by over 5.5%. The AReN-UNet with a cardinality of 4 achieved a better F1 score, accuracy, and AUC while the cascaded AReN-UNet with a cardinality of 2 achieved superior sensitivity than previously existing models. Table 3 and 4 represents the comparison of the proposed model with earlier existing algorithms. In addition to this, due to the presence of minimum channels in the upper convolutional layers of the proposed cascaded AReN-UNet, the memory consumption is relatively lower than the memory consumption of other models. Table 5 represents the comparison of memory consumption of different models with the proposed network in terms of trainable parameters.

The vessel maps produced by AReN-UNet and cascaded AReN-UNet are more effective in differentiating the smaller vascular structures and the background compared to the nested U-Net (backbone). The segmentation maps produced by the nested U-Net struggle to locate and identify vessels in low contrast regions of the input retinal image. As a result, it leads to high false-negative rate. The attention modules and the aggregated residual layers integrated with AReN-UNet improves the

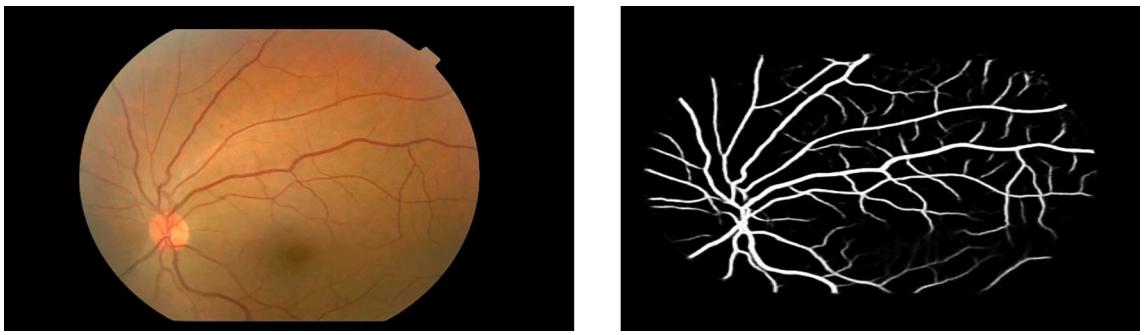

Fig. 9. A sample image from local dataset along segmented vascular map.

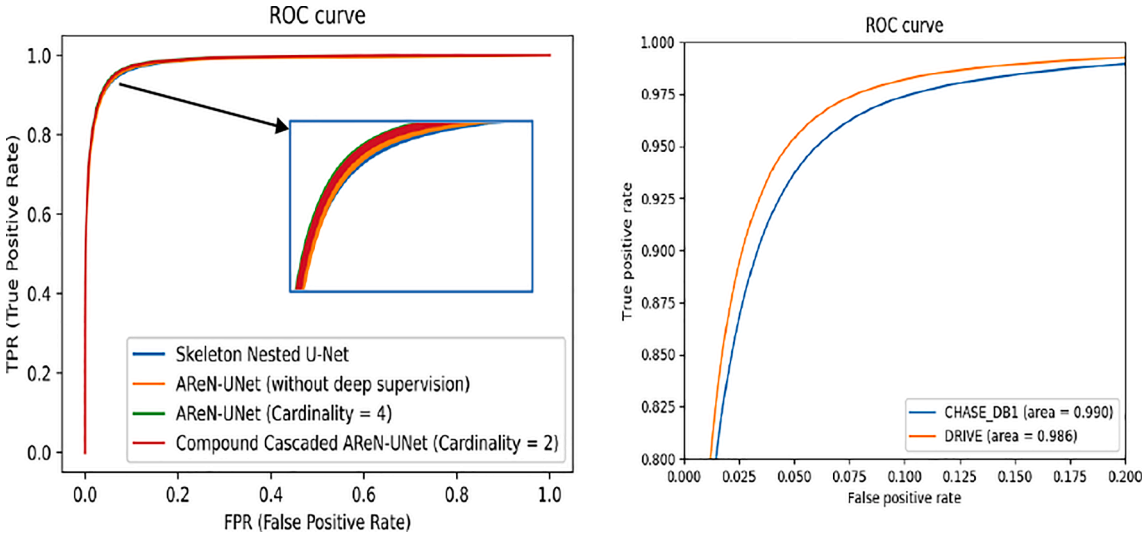

Fig. 10. Receiver operative characteristics of proposed work in comparison with baseline networks.

information extraction capabilities of the model. Consequently, the network addresses the problems of low contrast images and efficiently differentiates the foreground and background pixels. The single AReN-UNet may suffer due to minute false-positive vascular structures. The cascaded AReN-UNet refines the vascular networks using probabilistic feature maps of the front network. Fig. 10 represents the ROC proposed with comparisons to baseline models and different datasets. Fig. 10 has shown that larger AUC is produced by the AReN-UNet and Cascaded AReN-UNet which indicates the superior performance of the proposed models in differentiating between the foreground and background. Table 4 presents a comparison between the proposed model and the existing algorithms using the CHASE_DB1 dataset. The proposed models trained on this dataset are scaled-down due to computational limitations. However, they still managed to produce competitive results in comparison to existing models in the literature. AReN-UNet (cardinality = 2) achieved the highest F1-Score of 81.91%, the second-highest accuracy of 97.50%, and the third-highest AUC of 99.01% respectively. The proposed cascaded AReN-UNet (cardinality = 1) produced excellent results in terms of the second-highest F1-Score and sensitivity of 81.54% and 85.60% respectively.

4. Discussion

Vessel segmentation is an essential non-invasive tool for developing an automated screening system for diagnosing ocular patients. Improper vessel segmentation fails to detect the earlier stage of eye-related complications like Diabetic Retinopathy, Hypertensive Retinopathy, etc, and progresses to a severe stage. So, it is a challenging task to segment retinal vascular maps with high precision to detect the retinal disorders for preventing vision impairment.

In the proposed network cascaded AReN-UNet, two deep learning networks are cascaded such that the output of the front network is fed to the second network along with the input retinal image. In the cascaded encoder-decoder architecture, the feature maps are estimated by integrating the model with residual and attention modules. This inclusion preactivated the network to segment vascular maps more effectively with high representation ability. At every stage of the encoder, two pooling operations, MaxPooling and Average Pooling are performed. The resultant feature maps obtained from each pooling technique are concatenated. The CBAM of the network improves the generalization ability of layers in identifying the vessel pixel differences. The proposed network, cascaded AReN-UNet of depth 4 having 4.8 M trainable parameters is trained for 200 epochs on cloud GPU services, Google Colaboratory for 4 h. Categorical Cross-Entropy is used to train the network, at the same time the network gets optimized. The Adam optimizer with an initial learning rate of 0.0001 is used for training the proposed cascaded AReN-UNet. The learning rate will be reduced by a factor of 0.1 if no reduction in loss value is found for seven consecutive epochs. The less deviation in training curves of the proposed network in Fig. 5 (c) describes the stability of the network over other networks. Furthermore, to illustrate the significance of each module in the proposed network, we performed ablation studies and respective results are depicted in Fig. 6 and Table 2 also. Table 2 reveals that the integration of attention and residual modules with the nested U-Net improved the AUC by 1.41%. Further, cascading the similar networks integrated with attention and residual modules by inter-network skip connections to

implement the proposed cascaded AReN-UNet raised the AUC by 1.59%. In addition to this, different evaluation techniques like testing the input images by randomly shuffling their patches and testing on images collected locally are also performed. The segmentation results in Figs. 7 and 9 describe the better performance of cascaded AReN-UNet in detecting the thin blood vessels. Table 5 displayed that the memory consumption of the proposed algorithm is relatively lower than other models. Apart from this, the time complexity of the proposed network is estimated to be less than 8 s for segmenting an image when tested on the different datasets.

5. Conclusion

This work introduces a deep cascading mechanism, called a cascaded AReN-UNet, that feeds the feature maps delivered by each convolutional block of the primary U-Net to a subsequent U-Net for refinement. The integration of the network with aggregated residual structures and attention mechanisms efficiently increases the segmentation results by improving the generalization and representation abilities of the proposed network. The cracked vessels detected in the vascular map of the front network are refined by the second network of cascaded AReN-UNet by considering the probabilistic feature maps of the front network and feature maps of the second network. Furthermore, the comparison of the proposed network outperformed the baseline models when tested on the DRIVE and CHASE_DB1 datasets. To validate the network's ability over heterogeneity nature, it is tested over a local dataset and achieved an accuracy of 94.5%.

The authors express their gratitude and thanks to Dr. U.S.N. Murthy, Head of Ophthalmology Department, Gayatri Vidya Parishad Institute of Health Care and Medical Technology, for helping to label the dataset.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

CRediT authorship contribution statement

Aamer Abdul Rahman: Resources, Methodology, Investigation, Data curation, Visualization. **Birendra Biswal:** Resources, Investigation, Project administration, Supervision. **P. Geetha Pavani:** Conceptualization, Software, Formal analysis, Writing - original draft. **Shazia Hasan:** Software, Writing - review & editing. **M.V.S. Sairam:** Software, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors express their gratitude and thanks to Dr. U.S.N. Murthy, Head of Ophthalmology Department, Gayatri Vidya Parishad Institute of Health Care and Medical Technology, for helping to label the dataset.

References

[1] WHO, World report on vision, 2019.
[2] D.B. Rein, P. Zhang, K.E. Wirth, P.P. Lee, T.J. Hoerger, N. McCall, R. Klein, J. M. Tielsch, S. Vijan, J. Saaddine, The economic burden of major adult visual disorders in the United States, Arch. Ophthalmol. 124 (2006) 1754–1760, https://doi.org/10.1001/archopht.124.12.1754.
[3] International Diabetes Federation, International Federation on Ageing, International Agency for the Prevention of Blindness, The Diabetic Retinopathy Barometer Report: Global findings, 2019. https://www.iapb.org/wp-content/uploads/DR-Global-Report-1.pdf.
[4] Domalpally, Fluorescein Angiography in Neovascular AMD, Jobson Med. Inf. LLC. (2015). http://www.reviewofophthalmology.com/content/d/retinal_insider/i/1225/c/23061/.
[5] C.L. Srinidhi, P. Aparna, J. Rajan, Recent Advancements in Retinal Vessel Segmentation, J. Med. Syst. 41 (2017), https://doi.org/10.1007/s10916-017-0719-2.
[6] G.P. Pappu, B. Biswal, T.K. Gandhi, M.V.S. Sai Ram, Classification of neovascularization on retinal images using extreme learning machine, Int. J. Imaging Syst. Technol. (2020) 1–15, https://doi.org/10.1002/ima.22529.
[7] B. Biswal, Geetha Pavani P, P.K. Biswal, Controlled differential evolution based detection of neovascularization on optic disc using support vector machine, (2020) 1–10. 10.1515/bmt-2020-0110.
[8] G.P. Pavani, B. Biswal, M.V.S. Sairam, P.K. Biswal, An exclusive-disjunction-based detection of neovascularisation using multi-scale CNN, IET Image Process. (2021) 1–12, https://doi.org/10.1049/ipr2.12122.
[9] R. Poplin, A.V. Varadarajan, K. Blumer, Y. Liu, M.V. McConnell, G.S. Corrado, L. Peng, D.R. Webster, Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning, Nat. Biomed. Eng. 2 (3) (2018) 158–164, https://doi.org/10.1038/s41551-018-0195-0.
[10] M.D. Abràmoff, M.K. Garvin, M. Sonka, Retinal Imaging and Image Analysis, IEEE Rev. Biomed. Eng. 3 (2010) 169–208, https://doi.org/10.1109/RBME.2010.2084567.
[11] X. You, Q. Peng, Y. Yuan, Y.-ming. Cheung, J. Lei, Segmentation of retinal blood vessels using the radial projection and semi-supervised approach, Pattern Recognit. 44 (10-11) (2011) 2314–2324, https://doi.org/10.1016/j.patcog.2011.01.007.
[12] J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, E. Angelopoulou, Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database, IET Image Process. 7 (2013) 373–383, https://doi.org/10.1049/iet-ipr.2012.0455.
[13] U.T.V. Nguyen, A. Bhuiyan, L.A.F. Park, K. Ramamohanarao, An effective retinal blood vessel segmentation method using multi-scale line detection, Pattern Recognit. 46 (3) (2013) 703–715, https://doi.org/10.1016/j.patcog.2012.08.009.
[14] B. Biswal, T. Pooja, N. Bala Subrahmanyam, Robust retinal blood vessel segmentation using line detectors with multiple masks, IET Image Process. 12 (3) (2018) 389–399, https://doi.org/10.1049/ipr2.v12.310.1049/iet-ipr.2017.0329.
[15] P.K. Karn, B. Biswal, S.R. Samantaray, Robust retinal blood vessel segmentation using hybrid active contour model, IET Image Process. 13 (3) (2019) 440–450, https://doi.org/10.1049/ipr2.v13.310.1049/iet-ipr.2018.5413.
[16] Y. Yin, M. Adel, S. Bourennane, Retinal vessel segmentation using a probabilistic tracking method, Pattern Recognit. 45 (4) (2012) 1235–1244, https://doi.org/10.1016/j.patcog.2011.09.019.
[17] C. Heneghan, J. Flynn, M. O'Keefe, M. Cahill, Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis, Med. Image Anal. 6 (2002) 407–429, https://doi.org/10.1016/S1361-8415(02)00058-0.
[18] O. Ronneberger, P. Fischer, T. Brox, O. Ronneberger, P. Fischer, Thomas Brox, U-net: Convolutional networks for biomedical image segmentation, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9351 (2015) 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
[19] V. Iglovikov, A. Shvets, TernausNet, U-Net with VGG11 encoder pre-trained on imagenet for image segmentation, ArXiv. (2018).
[20] P. Chudzik, S. Majumdar, F. Calivá, B. Al-Diri, A. Hunter, Exudate segmentation using fully convolutional neural networks and inception modules, (2018) 105. 10.1117/12.2293549.
[21] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR (2017. 2017-Janua (2017)) 6450–6458, https://doi.org/10.1109/CVPR.2017.683.
[22] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, Proc. IEEE Int. Conf. Comput. Vis. (2015 Inter (2015)) 1520–1528, https://doi.org/10.1109/ICCV.2015.178.
[23] W. Liu, J. Chen, C. Li, C. Qian, X. Chu, X. Hu, A cascaded inception of inception network with attention modulated feature fusion for human pose estimation, 32nd AAAI Conf, Artif. Intell. AAAI 2018 (2018) 7170–7177.
[24] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, Proc. - 30th IEEE Conf, Comput. Vis. Pattern Recognition, CVPR (2017. 2017-Janua (2017)) 6230–6239, https://doi.org/10.1109/CVPR.2017.660.
[25] D.C. Cireşan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, Adv. Neural Inf. Process. Syst. 4 (2012) 2843–2851.
[26] J.H. Tan, H. Fujita, S. Sivaprasad, S.V. Bhandary, A.K. Rao, K.C. Chua, U. R. Acharya, Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network, Inf. Sci. (Ny) 420 (2017) 66–76, https://doi.org/10.1016/j.ins.2017.08.050.
[27] Y. Wu, Y. Xia, Y. Song, D. Zhang, D. Liu, C. Zhang, W. Cai, Vessel-Net: Retinal vessel segmentation under multi-path supervision, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11764 (2019) 264–272, https://doi.org/10.1007/978-3-030-32239-7_30.
[28] C. Guo, M. Szemenyei, Y. Yi, W. Zhou, H. Bian, Residual spatial attention network for retinal vessel segmentation, ArXiv. 2 (2020) 509–519.
[29] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, ArXiv. 1 (2018) 3–19.
[30] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++, A nested u-net architecture for medical image segmentation, Springer International Publishing (2018), https://doi.org/10.1007/978-3-030-00889-5_1.

[31] L. Li, M. Verma, Y. Nakashima, H. Nagahara, R. Kawasaki, IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks, ArXiv. (2019) 3645–3654.
[32] J. Hu, H. Wang, S. Gao, M. Bao, T. Liu, Y. Wang, J. Zhang, S-UNet: A Bridge-Style U-Net Framework with a Saliency Mechanism for Retinal Vessel Segmentation, IEEE Access. 7 (2019) 174167–174177, https://doi.org/10.1109/ACCESS.2019.2940476.
[33] Y. Wu, Y. Xia, Y. Song, Y. Zhang, W. Cai, NFN + : A novel network followed network for retinal vessel segmentation, Neural Networks. 126 (2020) 153–162, https://doi.org/10.1016/j.neunet.2020.02.018.
[34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis, Pattern Recognit. (2016-Decem (2016)) 770–778, https://doi.org/10.1109/CVPR.2016.90.
[35] C. Guo, M. Szemenyei, Y. Pei, Y. Yi, W. Zhou, SD-Unet: A Structured Dropout U-Net for Retinal Vessel Segmentation, Proc. - 2019 IEEE 19th Int. Conf. Bioinforma. Bioeng. BIBE 2019. (2019) 439–444. 10.1109/BIBE.2019.00085.
[36] G. Ghiasi, T.Y. Lin, Q.V. Le, Dropblock: A regularization method for convolutional networks, Adv. Neural Inf. Process. Syst. (2018-Decem (2018)) 10727–10737.
[37] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging. 23 (4) (2004) 501–509, https://doi.org/10.1109/TMI.2004.825627.
[38] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, S.A. Barman, An ensemble classification-based approach applied to retinal blood vessel segmentation, IEEE Trans. Biomed. Eng. 59 (9) (2012) 2538–2548, https://doi.org/10.1109/TBME.2012.2205687.
[39] C.G. Owen, A.R. Rudnicka, R. Mullen, S.A. Barman, D. Monekosso, P.H. Whincup, J. Ng, C. Paterson, Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (caiar) program, Investig. Ophthalmol. Vis. Sci. 50 (2009) 2004–2010, https://doi.org/10.1167/iovs.08-3018.
[40] B. Al-Diri, A. Hunter, D. Steel, An Active Contour Model for Segmenting and Measuring Retinal Vessels, IEEE Trans. Med. Imaging. 28 (9) (2009) 1488–1497, https://doi.org/10.1109/TMI.2009.2017941.
[41] S. Roychowdhury, D.D. Koozekanani, K.K. Parhi, Iterative Vessel Segmentation of Fundus Images, IEEE Trans. Biomed. Eng. 62 (7) (2015) 1738–1749, https://doi.org/10.1109/TBME.1010.1109/TBME.2015.2403295.
[42] Z. Fan, J. Lu, C. Wei, H. Huang, X. Cai, X. Chen, A Hierarchical Image Matting Model for Blood Vessel Segmentation in Fundus Images, IEEE Trans. Image Process. 28 (5) (2019) 2367–2377, https://doi.org/10.1109/TIP.2018.2885495.
[43] M.Z. Alom, C. Yakopcic, M. Hasan, T.M. Taha, V.K. Asari, Recurrent residual U-Net for medical image segmentation, J. Med. Imaging. 6 (2019) 1, https://doi.org/10.1117/1.jmi.6.1.014006.
[44] J. Zhuang, LadderNet,, Multi-path networks based on U-Net for medical image segmentation, ArXiv. (2018) 2–5.
[45] S. V., I. G., Encoder Enhanced Atrous (EEA) Unet architecture for Retinal Blood vessel segmentation, Cogn. Syst. Res. 67 (2021) 84–95, https://doi.org/10.1016/j.cogsys.2021.01.003.
[46] H. Tong, Z. Fang, Z. Wei, Q. Cai, Y. Gao, SAT-Net: a side attention network for retinal image segmentation, Appl. Intell. (2021), https://doi.org/10.1007/s10489-020-01966-z.