2023 IEEE International Conference on Image Processing (ICIP) | 978-1-7281-9835-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIP49359.2023.10222431

A SHALLOW U-NET WITH SPLIT-FUSED ATTENTION MECHANISM FOR RETINAL VESSEL SEGMENTATION

Amit Bhati[⋆], *Samir Jain*[⋆], *Neha Gour*[†], *Pritee Khanna* [⋆], *Aparajita Ojha* [⋆], *and Naoufel Werghi* [†]

[⋆] Department of Computer Science and Engineering, PDPM IIITDM, Jabalpur
[†]C2PS, Departement of Electrical Engineering and Computer Science, Khalifa University, UAE

ABSTRACT

Extraction of retinal vascular parts is an important task in retinal disease diagnosis. Precise segmentation of the retinal vascular pattern is challenging due to its complex structure, overlapping with other anatomical structures, and crucial thin vascular structures. In recent years, complex and heavy deep learning networks have been proposed to segment retinal blood vessels accurately. However, these methods fail to detect the thin vascular structure among different patterns of thick vessels. An attention-based novel architecture is proposed to segment the thin vasculature to address this limitation. The proposed model comprises a shallow U-Net based encoder-decoder architecture with split-fuse attention (SFA) block. The proposed SFA block enables the network to identify the placement of pixels for the tree-shaped vessel patterns at their relative position during the reconstruction phase in the decoder. The attention block aggregates low-level and high-level semantic information, improving the vessel segmentation performance. Experimentation performed on publicly available fundus datasets, DRIVE, HRF, CHASE-DB1, and STARE show that the proposed method performs better than the current state-of-the-art methods. The results demonstrate the adaptability of the proposed model for clinical applications due to its low memory footprint and better performance.

Index Terms— Retinal Vessel Segmentation, Shallow U-Net, Split Fused Attention, Fully Convolution Network, Encoder-Decoder.

1. INTRODUCTION

Retinal vessels are the only human blood circulation system component that can be immediately and non-invasively observed. Retinal vascular characteristics like shape, tortuosity, and structural tree pattern play a vital role in early-stage ocular disease diagnosis [1]. Retinal vessel segmentation is crucial among other important tasks for ocular disease diagnosis, including optic disc and cup segmentation, ocular disease grading, and detection. Specifically, abnormalities like artery thickening, increase in retinal capillaries, and hemorrhages can be observed in the vessel structure of diabetic retinopathy patients. Additionally, retinal vasculature can predict patient age, hemoglobin level, blood pressure, and other heart disease parameters, which can aid in disease diagnosis and risk identification [2]. As manual retinal vessel extraction is tedious and time-consuming, an efficient and fast vascular segmentation algorithm is vital in computer-aided diagnosis [3].

Recently, deep learning-based methods using convolutional neural networks (CNN) yielded significant performance in medical image analysis. U-Net provided impressive performance in an end-to-end manner for retinal vessel segmentation [4, 5, 6, 7, 8, 9, 10, 11] . However, these approaches result in scattered vessel fragments achieving discontinuous blood vessels by misclassifying vessel breakpoints due to the distribution imbalance between the vessels and the background pixels in a fundus image. In this work, attention-based shallow encoder-decoder architecture is proposed, which learns the vessel tree branch structures and identifies vessel branching areas. This decreases vessel breakpoints and improves vessel connection in the segmentation results.

Yan et al. [5] proposed a U-Net-based architecture with a pixel-wise and segmentation-based joint loss function. Similar to this, Wang et al. [6] proposed a dual encoder-based segmentation model to combine spatial and contextual information for better network learning. A combination of local matting loss and global pixel-level loss was suggested by Zhao et al. [7]. Libacher et al. [8] proposed a MobileNet-based M2U-Net architecture consisting of fewer trainable parameters but a marginal loss of performance. Similarly, Zhuo et al. [9] proposed a fusion of dense connectivity and size-invariant feature maps. But, the model is unable to beat the performance of the model proposed by Wang et al. Similarly, Mou et al. [12] trained a dilated attention-based U-Net model for getting better performance. They utilized multi-scale DICE loss function and probabilistic random walk for vessel fracture corrections. Li et al. [10] proposed a multi-scale residual similarity gathering (MRSG) and response cue erasing (RCE) based method for allowing the model to focus more on retinal vessel branching. This architecture outperformed other state-of-the-art methods. Although all these methods are good in performance, these models have a large number of trainable parameters which makes them unsuitable to deploy in resource constraint environments. Recently Galdran et al. [11] proposed a shallow W-Net using a pair of U-Net

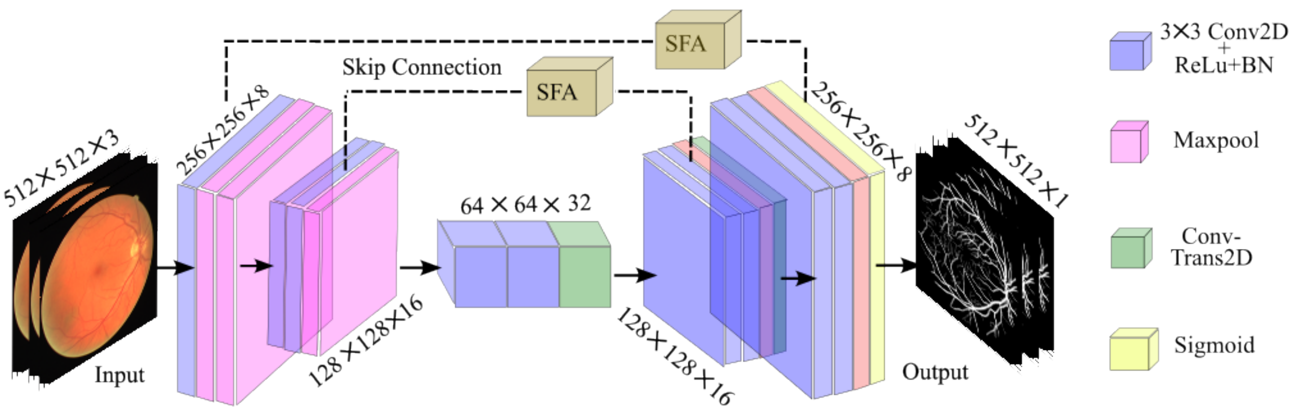

Fig. 1. Block diagram of the proposed model.

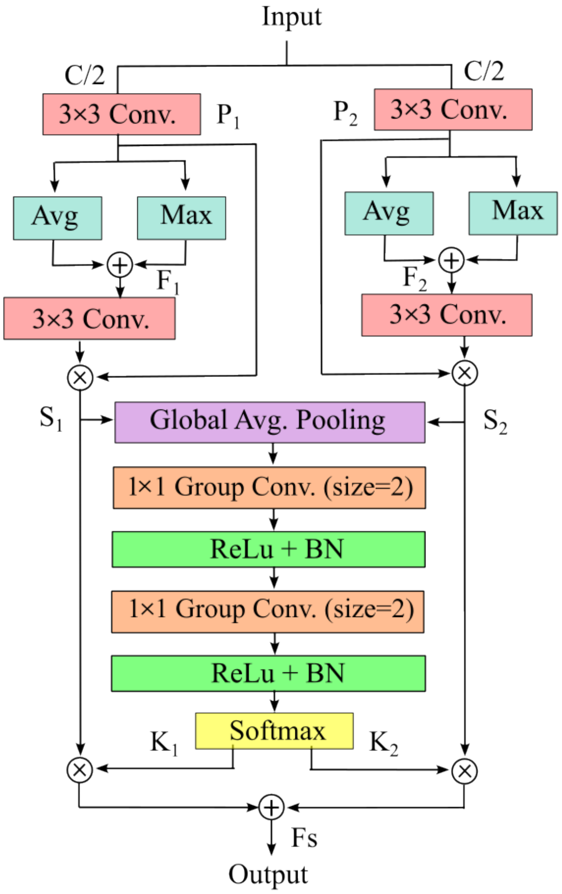

Fig. 2. Block diagram of Split Fused Attention (SFA) block.

to give a comparable performance with a very less number of trainable parameters.

A self-attention-based Split Fused Attention (SFA) block proposed in this work strengthens the vessel branching features extraction and enhances the network's capacity by using re-calibrated weighted attention maps to decode vessel structure. Encoder features in the skip connections are enhanced with the SFA block that utilizes spatial and channel-wise group convolution operations. Experiments performed on benchmark datasets including DRIVE [13], HRF [14], CHASE-DB1 [15], and STARE [16] show that the proposed method outperforms state-of-the-art methods.

The paper is organized into four sections. The proposed methodology is given in Section 2. The datasets, experimental setup, and results are compiled in Section 3. Lastly, the work is concluded in Section 4.

2. METHODOLOGY

2.1. Proposed Network

Recent methods in the literature showed that an optimized shallow U-Net perform better for retinal vessel segmentation compared to complex U-Net architecture [11]. The proposed model shown in Fig. 1 employs a 5-layer shallow U-Net. The encoder has three convolution layers, while the decoder has two convolution layers. Convolution layers with filters of size 3×3 followed by a batch-normalization layer are used in the encoder-decoder modules. Max-pooling with a factor of 2 is exploited in the encoder module for the downsampling of input. The decoder upsamples the input by 2 using transposed convolution operation to learn the pixel mapping corresponding to features extracted from the encoder. This shallow network is configured with 8, 16, and 32 filters in the encoder convolution layers while the decoder layers are having 16 and 8 filters. As depicted in Fig. 1, the first two encoder layers share the feature maps with the corresponding decoder layers, such that these feature maps are processed by SFA blocks. The architecture of SFA block is detailed in the next subsection. Finally, the decoder's output is passed through the sigmoid activation function for the generation of segmentation masks.

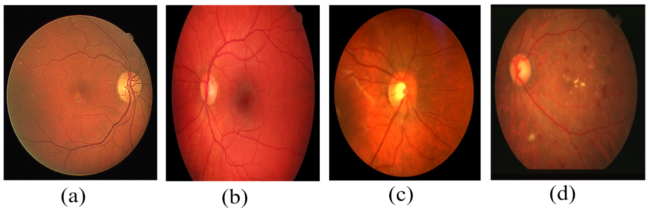

Fig. 3. Retinal Fundus Image from four datasets (a) DRIVE [13] (b) HRF [14] (c) CHASEDB [15] (d) STARE [16].

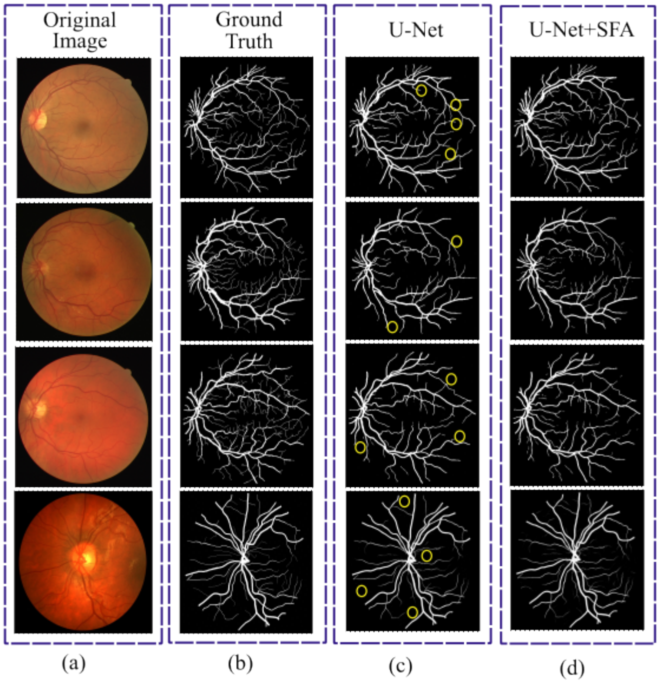

Fig. 4. Visual comparison of retinal vessel segmentation (a) original fundus images (b) ground truths (c) shallow U-Net segmentation results with vessel breakpoints/missing areas highlighted (d) vessels extracted by shallow U-Net with SFA.

2.2. Split Fused Attention Block

In the process of retinal vessel segmentation factors like uneven brightness in a thin vessel region and background color similarity with negligible contrast between tiny vessels are common causes of retinal vessel discontinuity in the prediction. This makes the segmentation task challenging for the model to distinguish between the background and the vessels. To enable the network for veins' actual branch breakpoint identification, it is important to learn vessels' branch breakpoint patterns. This can be achieved by making the network pay attention to this breakpoint area and extract the relevant information. SFA blocks are introduced in the proposed segmentation model. SFA is rooted in the convolution block attention module (CBAM) [17], which exploits both spatial and channel attention. In the SFA block shown in Fig. 2, the process starts with the splitting of the input feature map into two halves C_1 and C_2. Each of these halves is separately passed through an initial 3×3 convolution layer whose output is referred to as P_1 and P_2. Both outputs are passed to spatial average and max pooling in parallel followed by the pixel-wise addition operation. So obtained feature maps, F_1 and F_2, are processed by 3×3 convolution blocks followed by an element-wise multiplication operation performed with P_1 and P_2 respectively to create spatial attention maps S_1 and S_2. Here the spatial average helps in the rough estimation of boundaries in broad regions while the max operation gives more emphasis on the vascular region. Thus, the spatial attention map enables the network to pay more attention to vessel branching and breakpoint areas. Global average pooling is performed on S_1 and S_2 to achieve channel-wise attention that reduces channel-wise redundant information and allows the network to focus more on informative features. This is followed by 2 consecutive layers of 1×1 grouped convolution with ReLu activation and batch normalization. Batch normalization is used here to avoid over-fitting during the model training. The output obtained from channel attention is split into two halves, K_1 and K_2, and are multiplied with S_1 and S_2 followed by channel-wise concatenation operation to obtain the final fused attention map F_s as shown in Fig. 2.

3. RESULTS

Datasets: The model is validated on four publicly available fundus image datasets (Fig. 3). DRIVE [13] dataset has 40 fundus images including seven pathological cases. HRF [14] dataset contains 45 fundus images of normal, glaucoma, and diabetic retinopathy classes. CHASE-DB1 [15] dataset comprises 14 pairs of patients' left and right fundus images. STARE [16] dataset contains 40 fundus images with hand-labeled ground truth for vessel segmentation.

Experimental Setup: The proposed model is implemented on NVidia DGX-A100 having AMD ROM 7742, 2.25 GHz 128 cores CPU with 512 GB memory size, and NVidia Tesla P100 40 GB GPU. The PyTorch model utilizes an Adam optimizer with an initial learning rate of $1e-4$ with a decay rate of $1e-5$ for the vanishing gradient issue. The models are trained for 100 epochs with a batch size of 8 using the DICE loss function expressed in equation 1:

$$Loss(y, \hat{y}) = 1 - \frac{2y\hat{y}}{y+\hat{y}} \quad (1)$$

where y refers to the ground truth label and $\hat{y}$ represents pseudo label. The performance of the proposed model is evaluated on AUC and Dice Score. AUC evaluates the discriminating capability of the model ranging between 0 and 1. In contrast, the dice score measures the similarity between the

Table 1. Performance comparison of SFA block with shallow U-Net architecture.

Method	Params	DRIVE		HRF		CHASE-DB 1		STARE	
		AUC	DICE	AUC	DICE	AUC	DICE	AUC	DICE
U-Net	34.20K	97.98	82.41	98.11	80.59	98.22	80.29	98.04	79.65
U-Net+SFA	**74.24K**	**98.05**	**83.89**	**98.67**	**83.11**	**99.01**	**83.02**	**98.67**	**81.43**

Table 2. Performance comparison of the proposed model with state-of-the-art methods.

Method	Params	DRIVE		HRF		CHASE-DB 1		STARE	
		AUC	DICE	AUC	DICE	AUC	DICE	AUC	DICE
Yan et al., 2018 [5]	-	97.52	81.83	-	78.14	97.81	-	98.01	-
Wang et al., 2019 [6]	-	97.72	82.7	-	-	98.12	80.37	98.46	
Zhao et al., 2018 [7]	-	-	78.82	-	76.59	-	-	-	74.84
Laibacher et al., 2019 [8]	549K	97.14	80.91	-	78.14	97.03	80.06	-	-
Zhuo et al., 2020 [9]	-	97.54	81.63	-	-	-	-	98.24	-
Mou et al., 2019 [12]	56M	97.96	-	-	-	98.12	-	98.58	-
Li et. al., 2022 [10]	2.01M	**98.43**	-	-	-	98.35	-	98.43	-
Galdran et. al., 2022 [11]	68.48K	98.1	82.79	98.25	81.03	98.47	81.69	98.28	79.76
Proposed Method	**74.24K**	98.05	**83.89**	**98.67**	**83.11**	**99.01**	**83.02**	**98.67**	**81.43**

predicted label to the ground truth, where a higher score indicates a better performance.

Ablation Study: Experiments were carried out to study the performance of the proposed shallow U-net and the effect of the SFA block on segmentation results. The outcomes are summarized in Table 1. The proposed shallow U-Net with SFA block performs better as compared to the baseline shallow U-Net. An average improvement of 1.79%, 3.12%, 3.40%, and 2.23% in Dice scores is obtained for DRIVE, HRF, CHASE-DB1, and STARE datasets, respectively. The characteristic response and visual analysis of vessel breaking areas are conducted through the gradient maps shown in Fig 4. The qualitative results depict the effectiveness of the SFA block. It can be observed from Fig. 4 that shallow U-Net with SFA block can identify the complex structures of vessels more accurately and also be able to reconstruct them with more precision. The yellow circles in Fig. 4 (c) are showing vessel breakpoints and missing areas that could not be captured by shallow U-Net. It can be minutely observed that these vessel breakpoints and missing areas are covered with the application of SFA block as shown in Fig. 4 (d).

Comparative Study: The performance of the proposed model is compared with recently published state-of-the-art methods. Table 2 summarized the comparative study using AUC and dice parameters. The proposed model performs better as compared to the methods proposed by Yan et al. [5], Wang et al. [6], Zhao et al. [7], Laibacher et al. [8], Zhuo et al. [9], and Mou et al. [12] for all four datasets in terms of AUC and Dice scores. However, the method proposed by Li et al. [10] and Galdran et al. [11] performed slightly better (0.38% and 0.05% respectively) in terms of AUC only for the DRIVE dataset. The total training parameters for models are also reported in Table 2. It can be noted that the slight improvement of 0.38% is obtained by Li et al. [10] at the cost of 2.01M parameters used in their model as compared to 0.74M parameters used in the proposed model.

The quantitative and quantitative results demonstrate the effectiveness of the proposed method for enhancing vascular segmentation performance. The use of SFA blocks enables the network to focus on regions prone to vessel break points and learns the branching patterns and structures of such cases to avoid pixel placement error in the decoder part. Due to its low memory footprint and better performance, the proposed model is preferably suited for clinical applications also.

4. CONCLUSION & FUTURE WORK

Retinal vessel segmentation is the primary step for the early-stage detection of retinal diseases. It is found to be challenging due to the presence of thin vessels and complex breakpoints. This work presents an attention-based shallow U-Net architecture for efficient retinal vessel segmentation. The key idea is to utilize the U-Net like model with an attention mechanism. A self-attention based SFA block that exploits spatial and channel information is introduced in the proposed network. As a result, the proposed network effectively learns vessel breakpoints and branching patterns. Experimentation shows that the proposed model can detect both thicker and micro-level vessels efficiently. The method outperforms state-of-the-art methods for three benchmark datasets except for the DRIVE dataset in terms of AUC and that too by a narrow margin. Future work will focus on further refinement in the model architecture and the reduction of false positives of micro-vessels.

5. REFERENCES

[1] C. L Srinidhi, P. Aparna, and J. Rajan, "Recent advancements in retinal vessel segmentation," *Journal of medical systems*, vol. 41, pp. 1–22, 2017.

[2] R. Poplin, A. Varadarajan, K. Blumer, Y. Liu, M. Mcconnell, G. Corrado, L. Peng, and D. Webster, "Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. arxiv 2017," *arXiv preprint arXiv:1708.09843*, 2017.

[3] K. Sun, Y. Chen, Y. Chao, J. Geng, and Y. Chen, "A retinal vessel segmentation method based improved u-net model," *Biomedical Signal Processing and Control*, vol. 82, p. 104574, 2023.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[5] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1912–1923, 2018.

[6] X. Wang, X. Jiang, and J. Ren, "Blood vessel segmentation from fundus image by a cascade classification framework," *Pattern Recognition*, vol. 88, pp. 331–341, 2019.

[7] H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, and L. Cheng, "Supervised segmentation of un-annotated retinal fundus images by synthesis," *IEEE transactions on medical imaging*, vol. 38, no. 1, pp. 46–56, 2018.

[8] T. Laibacher, T. Weyde, and S. Jalali, "M2u-net: Effective and efficient retinal vessel segmentation for real-world applications," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[9] Z. Zhuo, J. Huang, K. Lu, D. Pan, and S. Feng, "A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index," *Computer methods and programs in biomedicine*, vol. 196, p. 105508, 2020.

[10] M. Li, S. Zhou, C. Chen, Y. Zhang, D. Liu, and Z. Xiong, "Retinal vessel segmentation with pixel-wise adaptive filters," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2022, pp. 1–5.

[11] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, and I. B. Ayed, "State-of-the-art retinal vessel segmentation with minimalistic models," *Scientific Reports*, vol. 12, no. 1, p. 6174, 2022.

[12] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "Cs-net: channel and spatial attention network for curvilinear structure segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, 2019, pp. 721–730.

[13] A. Asad, A. T. Azar, N. El-Bendary, A. E. Hassaanien *et al.*, "Ant colony based feature selection heuristics for retinal vessel segmentation," *arXiv preprint arXiv:1403.1735*, 2014.

[14] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International journal of biomedical imaging*, vol. 2013, 2013.

[15] H. Y. Henry, X. Feng, Z. Wang, and H. Sun, "Mix-module: Mixed cnn kernel module for medical image segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1508–1512.

[16] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[17] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.