22 th International Conference on
Information Technology,
Computer and Telecommunication
Event Place: Tbilisi,Georgia
www.itctconf.ir
بیست و دومین کنفرانس بین المللی
فناوری اطلاعات، کامپیوتر و مخابرات | گرجستان
22th International Conference on Information Technology,Computer and Telecommunication
مجلات معتبر بین المللی ۲۶ اسفندماه ۱۴۰۲

Accurate segmentation of retinal vessels using U-Net and spatial convolution neural network

First Author Ali Ghanbarzhade
Amirkabir University of Technology

Second Author[1] Saman Amini Serajgah
Kharazmi University Tehran Iran

Third Author Zeinab Razmi Hamzeh Khan Lou
Kharazmi University Tehran Iran

Fourth Author Maral Mirzamohammadi
Iran University of Science and Technology

Abstract

Early diagnosis of diseases such as diabetes and high blood pressure, which have a direct effect on retinal blood vessels, is very important. In this research, we propose a complex convolutional network with a spatial U-Net, which is used without the need for a large number of training data in a data augmentation manner for optimal use of samples. U-Net is a spatial module that expands the attention map along the spatial dimension and multiplies it to the input feature map for adaptive feature refinement. The input of the complex convolutional network is the structured output blocks from the previous step and does not use the initial spatial U-Net convolution, which avoids additional processing and increases accuracy. To evaluate the proposed method, you use two retina data sets. Two sets of retinal vessel extraction data (DRIVE) and data (CHASE_DB1) show that the proposed method performs well in both data sets.

Keywords: U-Net, retinal vessel, image segmentation, convolution network, denoising

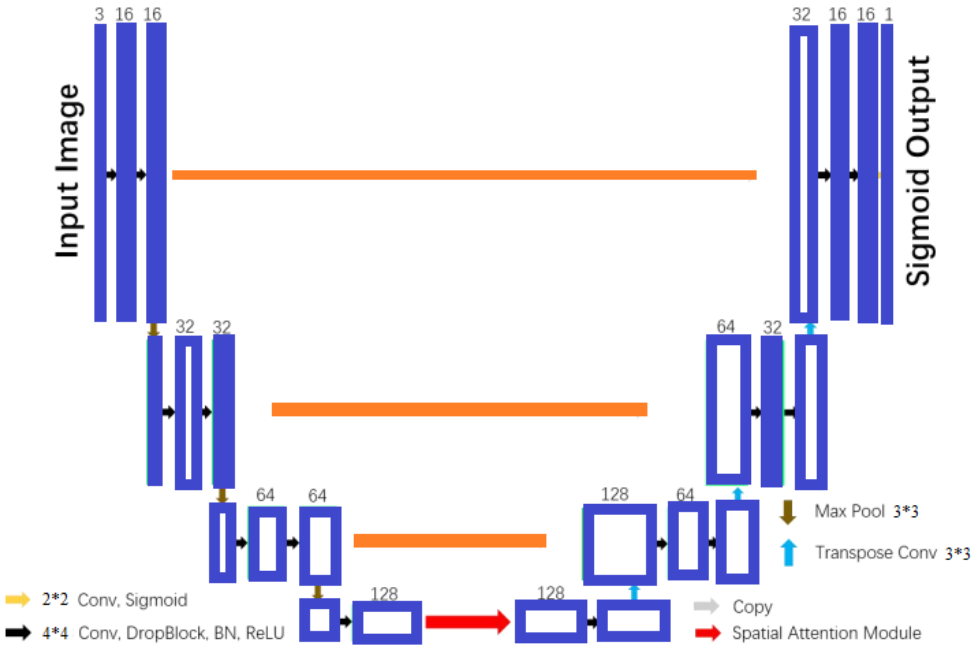
1-Corresponding Author

Introduction

Observing the fundus vascular system can help diagnose and track many diseases, including diabetes and hypertension, which can cause changes in the morphology of the blood vessels of the retina. Diabetes, in particular, can cause systemic microvascular and small vessel diseases, with the retinal vascular disease being the most vulnerable[1]. This can result in Diabetic Retinopathy (DR), which requires special attention if swelling is observed in the blood vessels of the retina. Patients with long-term hypertension may notice blood vessel curvature or vascular stenosis due to increased arterial blood pressure, known as hypertensive retinopathy (HR)[2] [3].

Segmenting the retinal blood vessels is essential for a quantitative analysis of fundus images. This process helps to obtain relevant morphological information about the retinal blood vessel tree, such as the curvature, length, and width of the blood vessels[5]. The unique characteristics of the vascular tree of Retinal vessels can also be applied to biometric recognition. Therefore, accurate segmentation of retinal blood vessels is of great significance. However, retinal blood vessels are fragile and numerous, and the blood vessels are closely connected, making the retinal blood vessel tree structure complex. The segmentation of retinal blood vessels is a challenging task due to various factors such as the lack of clear distinction between the blood vessel area and the background, as well as the susceptibility of fund images to uneven lighting and noise[6]. In the past few decades, many retinal blood vessel segmentation methods have been proposed, mainly divided into manual and automatic segmentation methods. Manual segmentation is time-consuming and labor-intensive, requiring high professional skills.

On the other hand, automatic segmentation algorithms can reduce the burden of manual segmentation. With the recent advancements in deep learning, U-Net has become a common network architecture for retinal segmentation. However, these variants of U-Net can make the network more complex and less interpretable[7]. To address this issue, the Spatial Attention U-Net (SA-UNet) was proposed, which employs a structured dropout convolutional block integrating Drop Block and batch normalization (BN) to replace the original U-Net convolutional block[8]. Additionally, spatial attention was introduced to enhance important features, such as vascular features, and suppress unimportant features. SA-UNet was evaluated on two public retinal fundus image datasets and achieved state-of-the-art performance compared to other existing methods for retinal vascular segmentation.



Figureure(1). Diagram of the proposed SA-UNet

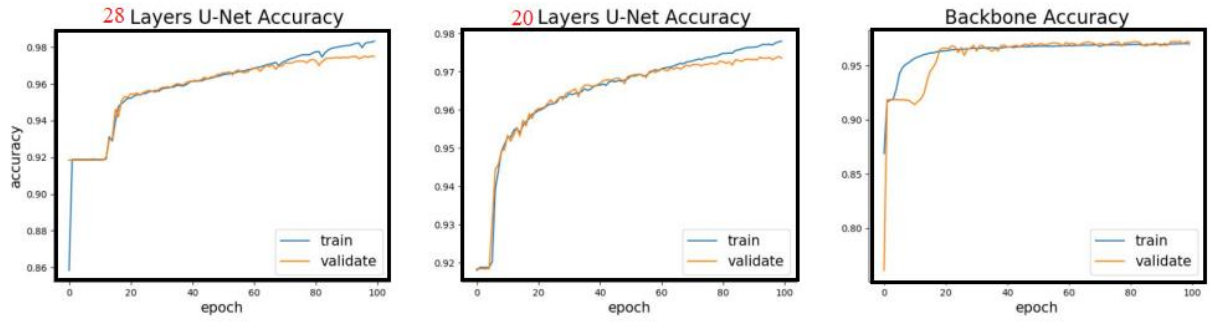
II. METHODOLOGY

A. Network Architecture

The proposed SA-UNet architecture is shown in Figure. 1, and it consists of a U-shaped encoder-decoder structure. In the encoder, each step contains a structured dropout convolutional block and a 3*3 max pooling operation. The convolutional block of each step is made up of a DropBlock, batch normalization (BN) layer, and rectified linear unit (ReLU), followed by the max pooling operation for down-sampling, with a stride size of 3. At each down-sampling step, the number of feature channels is doubled.
In the decoder, each step involves a 3*3 transposed convolution operation for up-sampling, which halves the number of feature channels. The output is then concatenated with the corresponding feature map from the encoder. This is followed by a structured dropout convolutional block. The spatial attention module is added between the encoder and decoder. For the output segmentation map, a 2*2 convolution and Sigmoid activation function are used at the final layer.

B. Structured dropout complex convolutional block

Data augmentation is a common technique used to prevent overfitting in machine learning models. However, even after performing data augmentation on the original datasets, we observed serious overfitting during the training of the original U-Net, as shown in Figureure 2 (left). To address this issue, we employed a lightweight U-Net with 28 convolutional layers as our basic architecture. However, this architecture still suffered from an overfitting problem, as demonstrated in Figureure 2 (middle). To solve this problem, we adopted DropBlock, a structured form of dropout, which is known to effectively prevent overfitting problems in convolutional networks. Unlike traditional dropout, DropBlock discards contiguous areas from a feature map of a layer instead of dropping independent random units. We constructed a structured dropout convolutional block, where each convolutional layer is followed by a DropBlock, a layer of batch normalization (BN), and a ReLU activation unit. This block was employed instead of the original convolutional block of U-Net to build a U-shaped network as our "Backbone". Our Backbone has only 20 convolutional layers, compared to the 29 convolutional layers of the original U-Net. As shown in Figureure 2 (left), the overfitting problem was perfectly solved by using this structured dropout convolutional block and it accelerated the convergence of the network. Additionally, we introduced batch normalization (BN) to the structured dropout convolutional block to improve the network's convergence.



Figureure (2).Comparison of different models training 100 epochs on DRIVE.

C. Spatial Attention Module (SAM)

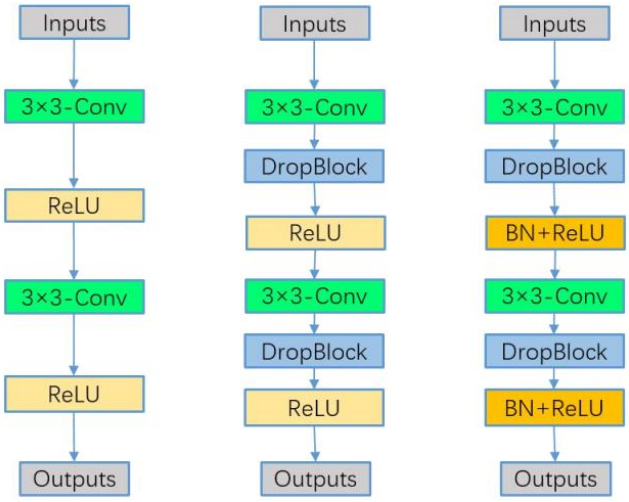
The Spatial Attention Module (SAM) is a component of the convolutional block attention module which is used for classification and detection [14]. SAM leverages the spatial relationship between features to generate a spatial attention map. To calculate the spatial attention, SAM applies max-pooling and average-pooling operations along the channel axis and concatenates them to create an efficient feature descriptor, as depicted in Figureure 4.
Formally, input feature $F \in R^{H\times W\times C}$ is fowarded through the channel-wise max poling and average-pooling to generate outputs$F^S_{mp} \in R^{H\times W\times 1}$ and$F^S_{ap} \in R^{H\times W\times 1}$, respectively. Then a convolutional layer followed by the Sigmoid activation function on the concatenated feature descriptor is used to generate a spatial attention map$M^S(F) \in R^{H\times W\times 1}$. In short, the output feature $F^S(F) \in R^{H\times W\times C}$ of spatial attention module is calculated as:

$$F^S = F.M^S(F) =$$
$$F.\sigma\left(f^{7*7}([MaxPool(F); AvgPool(F)])\right) =$$
$$F.\sigma\left(f^{7*7}\left([F^S_{mp}; F^S_{ap}]\right)\right) \quad (1)$$

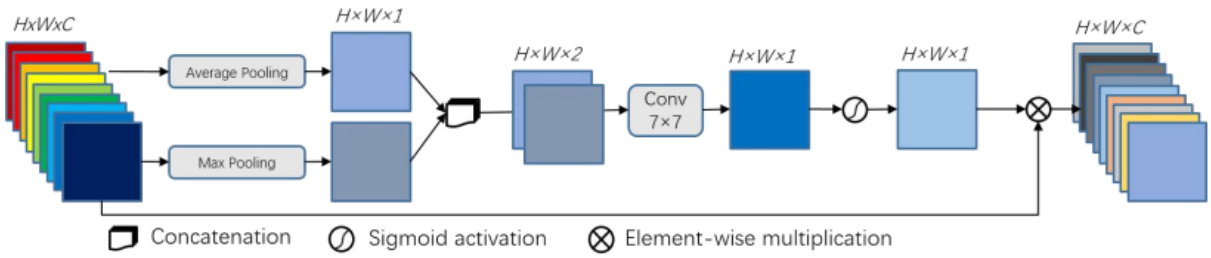
III. EXPERIMENTS

A. Datasets

Our proposed SA-UNet was evaluated on two publicly available retinal fundus image datasets: DRIVE and CHASE DB1. Table 1 provides specific information on the two datasets. It is important to note that the original size of the datasets was not appropriate for our network. Therefore, we adjusted their size by adding zero padding around them. However, during evaluation, the size was cropped back to the initial size. To enhance the data, we used four data augmentation methods, as shown in the last column of Table 1, for both datasets. Each method generated three new images from the original image. This augmentation process increased the number of training images from the original 20 to 256, for both datasets.



Figureure (3) . Original U-Net block (left), SD-Unet block (middle), Structureddropout convolutional



Figureure (4). Diagram of the Spatial Attention Module

B. Evaluation Metrics

To evaluate the performance of our model, we compare the segmentation results with the corresponding ground truth. We divide the results of each pixel comparison into four categories: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Then, we use sensitivity (SE), specificity (SP), F1-score (F1), and accuracy (ACC) to assess the model's performance. In retinal vessel segmentation, only a small percentage (9%-14%) of the pixels belong to blood vessels, while the rest are considered background pixels. For performance measurement of binary classifications with two categories of different sizes, the Matthews Correlation Coefficient (MCC) is suitable. The MCC value helps to determine the optimal setting for the vessel segmentation algorithm.

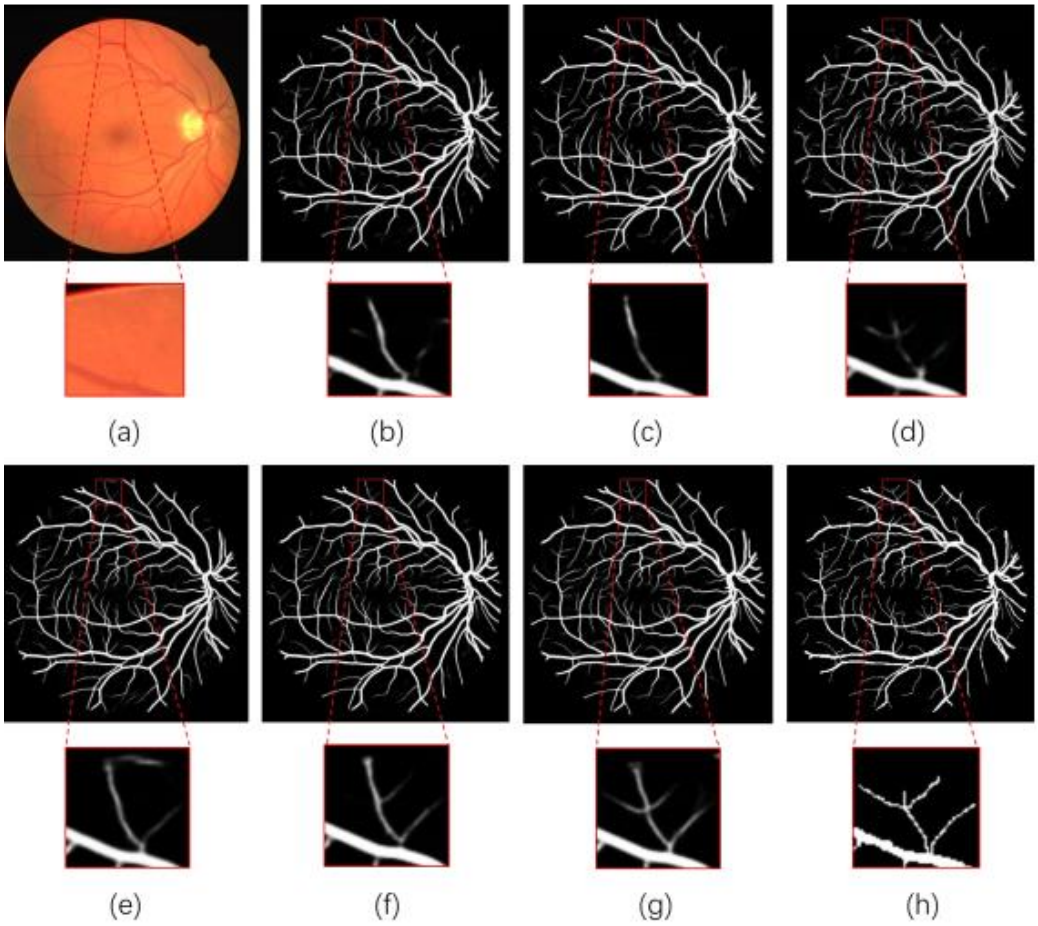
MCC is defined as follows:

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \qquad (2)$$

Table I the specific information of DRIVE and CHASE-DB1 Datasets

Datasets	DRIVE	CHASE-DB1

Obtained form	Dutch Diabetic Retinopathy Screening Program	Child Heart and Health Study
Total number	40	28
Train/ Test number	20/20	20/8
Resolution (pixel)	584*565	999*960
Resize(pixel)	592*592	1008*1008



Figureure. (5). (a) A test image from DRIVE **dataset; (b) Segmentation result by** U-Net**; (c) Segmentation result by** UNet+SA**;(d)Segmentation result by** AG-Net**; (e) Segmentation result by** SD-Unet**; (f) Segmentation result by** Backbone**; (g) Segmentation result by** SA-UNet**; (h) Corresponding ground truth segmentation**

C. Implementation Details

To ensure that our network is not overfitting, we randomly selected 26 and 13 images from the DRIVE and CHASE DB1 augmented datasets as the validation set. We trained the SA-UNet from scratch using the augmented training set, and employed the Adam optimizer and the binary cross entropy loss function for both datasets. To reduce the number of parameters, we set the number of channels after the first convolutional layer to 16. We trained the model for 150 epochs, with a learning rate of 0.001 for the first 100 epochs and 0.0001 for the last 50 epochs. The size of the discard blocks of DropBlock is set to 7. For the DRIVE dataset, we set the batch size to 8 and the dropout rate of DropBlock to 0.18. For CHASE DB1, the batch size is set to 4 and the dropout rate is 0.13. We implemented the model using the public Keras with Tensorflow as the backend, and all experiments were run on an NVIDIA TITAN XP GPU with 12 Gigabyte memory. Figure. 2 shows the case of training 100 epochs on the DRIVE dataset.

IV. Results Discussion

A. Ablation Experiments

A series of experiments were conducted on DRIVE and CHASE_DB1 to demonstrate that each component of the proposed SA-UNet can improve the performance of retinal vascular segmentation. In Tables II and III, the segmentation performance of U-Net, U-Net + SA, SD-Unet, Backbone, and SA-UNet are shown in decreasing order of performance, respectively. The parameter quantities of different models are shown in Table IV. From the results, several useful observations were made: (1) U-Net + SA outperforms U-Net with only 98 additional parameters, demonstrating the effectiveness of introducing spatial attention. (2) When using a structured dropout convolutional block based on U-Net, the ACC, AUC, F1, and MCC of the Backbone are higher than U-Net by 0.28%/0.22%, 0.73%/0.59%, 2.42%/2.48%, and 2.48%/2.64% on DRIVE and CHASE_DB1, respectively. This shows that the newly constructed structured dropout convolutional block is effective in building the Backbone. (3) The Backbone has a better performance compared to the original SD-Unet, even with a slightly increased number of parameters. This shows that adding batch normalization (BN) can improve network performance to a certain extent. (4) The proposed SA-UNet achieves the best performance on most metrics and has a much smaller number of parameters compared to AG-Net and the original U-Net with 23 convolutional layers. Therefore, for the task of retinal blood vessel segmentation, SA-UNet is a lightweight and effective network. In Figureure 5, a test example on the DRIVE dataset is shown, which includes the segmentation results obtained by U-Net, U-Net + SA, AG-Net, SD-Unet, Backbone, and the proposed SA-UNet, and the corresponding ground truth. Compared to U-Net and U-Net + SA, AG-Net has certain advantages in the segmentation of the edge structure, but is not strong enough at the intersection of small blood vessels. SD-Unet ignores some edge and small vascular structures and even produces incorrect segmentation. The Backbone produces more accurate small vessel segmentation than the U-Net and SD-Unet, demonstrating the effectiveness of the Backbone constructed using structured dropout convolutional blocks. Compared to the Backbone, the SA-UNet proposed can produce more accurate segmentation results for border blood vessels and retain more retinal blood vessel spatial structure, demonstrating that spatial attention mechanism can highlight blood vessels and reduce the influence of background. To better observe the results, more segmentation examples of U-Net, Backbone, and SA-UNet on DRIVE and CHASE_DB1 are shown in Figureures 6 and 7, respectively.

TABLE II. ABLATION STUDIES ON DRIVE **DATASET**

Methods	*SE*	*SP*	*ACC*	*AUC*	*F1*	*MCC*
U-Net	0.7677	0.9857	0.9666	0.9789	0.8012	0.7839
U-Net + SA	0.7883	0.9845	0.9673	0.9809	0.8085	0.7909
SD-Unet	0.7978	**0.9860**	0.9695	0.9858	0.8208	0.8045
Backbone	**0.8246**	0.9832	0.9694	0.9862	0.8254	0.8087
SA-UNet	0.8212	0.9840	**0.9698**	**0.9864**	**0.8263**	**0.8093**

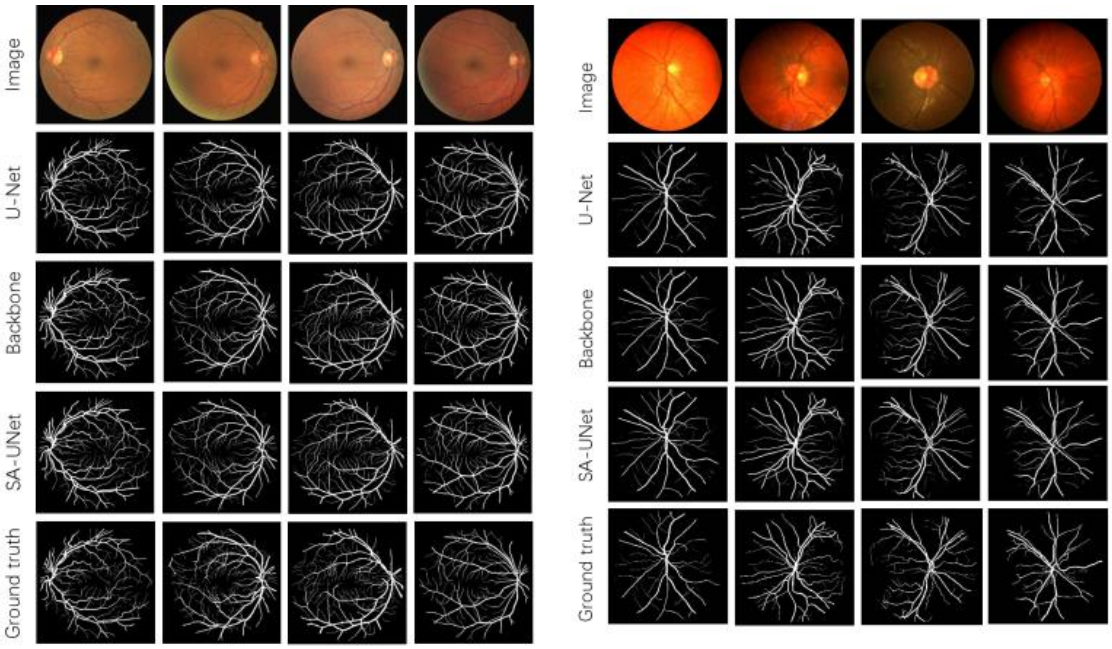
TABLE III. ABLATION STUDIES ON CHASE_DB1 **DATASET.**

Methods	***SE***	***SP***	***ACC***	***AUC***	***F1***	***MCC***
U-Net	0.7842	0.9861	0.9733	0.9838	0.7875	0.7733
U-Net + SA	0.7840	**0.9865**	0.9738	0.9852	0.7902	0.7763
SD-Unet	0.8297	0.9854	**0.9756**	0.9897	0.8109	0.7981
Backbone	0.8422	0.9844	0.9755	0.9897	0.8123	0.7997
SA-UNet	**0.8573**	0.9835	0.9755	**0.9905**	**0.8153**	**0.8033**

Table IV. AMOUNT OF PARAMETERS ON DIFFERENT MODELS

Models	Total	Tranable	Non-trainable
AG-Net	9335340	9335340	0
28 Layers-U Net	2158705	2158705	0
20 Layers-U Net	535793	535793	0
U Net+ SA	535891	535891	0

SD-Unet	535793	535793	0
Backbone	538609	537201	1408
SA-Unet	538707	537299	1408



B. Comparisons with state-of-the-art methods

We have compared the performance of SA-UNet with other state-of-the-art methods currently being used in retinal vessel segmentation tasks. In Tables V and VI, we have provided a summary of the release year of different methods, their performance, and the CHASE_DB1 datasets. The results show that SA-UNet has achieved the best performance in both DRIVE and CHASE_DB1, with the highest sensitivity of 0.8512 / 0.8973, the highest accuracy of 0.9998/0.9855, and the highest AUC of 0.9964 / 0.9985. Although the specificity is comparable with other methods, SA-UNet has better segmentation performance than the best-performing AG-Net in the previous methods, particularly at the intersection of small blood vessels, as shown in Figure. 5. Remarkably, the parameter amount of SA-UNet is much smaller than that of AG-Net. The above results demonstrate that SA-UNet has achieved state-of-the-art performance in the retinal vessel segmentation challenge.

TABLE VI. RESULTS OF SA-UNET AND OTHER METHODS ON DRIVE DATASETS.

Datasets	**CHASE-DB1**			
Metrics	SE	SP	ACC	AUC
[15]	0.7811	0.9807	0.9535	0.9790
[16]	0.7897	0.9684	0.9454	0.9507
[17]	0.7653	0.9818	0.9542	0.9752
[18]	0.7844	0.9819	0.9567	0.9807
[7]	0.7940	0.9816	0.9567	0.9772
[8]	0.8038	0.9802	0.9578	0.9821
[9]	0.8100	0.9848	0.9692	0.9856
This Work	0.8503	0.9845	0.9898	0.9805

TABLE VI. RESULTS OF SA-UNET AND OTHER METHODS ON CHASE_DB1 DATASETS.

Datasets	**CHASE-DB1**

Metrics	SE	SP	ACC	AUC
[15]	0.7816	0.9836	0.9628	0.9823
[16]	0.7277	0.9712	0.9458	0.9524
[17]	0.7633	0.9809	0.9610	0.9781
[18]	0.7538	0.9847	0.9637	0.9825
[7]	0.8074	0.9821	0.9661	0.9812
[8]	0.8132	0.9814	0.9661	0.9812
[9]	0.8186	0.9848	0.9743	0.9863
This Work	0.8573	0.8573	0.9755	0.9905

Conclusions

Most datasets of retinal fundus images are relatively small, which can make it difficult to train deep neural networks. To help with learning, data augmentation is often employed, but even with this approach, it's common to observe overfitting. To address this issue, we developed a new approach inspired by the successful use of DropBlock and batch normalization in convolutional neural networks. Our approach replaces the convolutional block of U-Net with a structured dropout convolutional block that uses DropBlock and batch normalization as its backbone. Additionally, in retinal fundus images, it can be challenging to distinguish between the blood vessel area and the background, especially when it comes to small blood vessels and edges. To help the network learn these features, we've added a spatial attention module between the encoder and decoder of the backbone, resulting in a Spatial Attention U-Net (SA-UNet). By using spatial attention, we're able to help the network focus on important features and suppress unnecessary ones, improving its representation capability. We've tested SA-UNet on two publicly available datasets of retinal fundus images, DRIVE and CHASE_DB1, and have found that our approach is effective. In fact, when compared with other state-of-the-art methods for retinal vessel segmentation, our lightweight SA-UNet achieves state-of-the-art performance. We believe that SA-UNet is a general network and can be applied to other retinal vessel segmentation tasks due to the similar vascular structure characteristics of the retinal image.

References

[1] Chen Y, Wang J, Chen X, Zhu M, Yang K, Wang Z, Xia R (2019) Single-image super-resolution algorithm based on structural self-similarity and deformation block features. IEEE Access 7:58791–58801

[2] K. Kipli, M. E. Hoque, L. T. Lim, M. H. Mahmood, S. K. Sahari, R. Sapawi, N. Rajaee, and A. Joseph, ‘‘A review on the extraction of quantitative retinal microvascular image feature,’’ Comput. Math. Methods Med., vol. 2018, pp. 1–21, Jul. 2018.

[3] Jin, Qiangguo, et al. "DUNet: A deformable network for retinal vessel segmentation." Knowledge-Based Systems 178 (2019): 149-162.

[4] Marcos Ortega, M.G. Penedo, J. Rouco, N. Barreira, M.J. Carreira, "Personal verification based on extraction and characterization of retinal feature points."Journal of Visual Languages & Computing 20.2 :80-90, 2009.

[5] Simon and I. Goldstein. A new scientific method of identification. New York State Journal of Medicine, 35(18):901–906, Sept. 1935.

[6] Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham, 2015

[7] Wang B., Qiu S., He H. (2019) Dual Encoding U-Net for Retinal Vessel Segmentation. In: Shen D. et al. (eds) MICCAI 2019. Lecture Notes in Computer Science, vol 11764. Springer, Cham, 2019..

[8] Y Wu, Y Xia, Y Song, D Zhang, D Liu, C Zhang, W Cai. (2019) Vessel-Net: Retinal Vessel Segmentation Under Multi-path Supervision. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science, vol 11764. Springer, Cham, 2019.

[9] S, Zhang., H, Fu., Y, Yan., Y, Zhang., Q, Wu., M, Yang., M, Tan. Attention Guided Network for Retinal Image Segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science, vol 11764. Springer, Cham, 2019

[10] C. Zhu, B. Zou, R. Zhao et al., “Retinal vessel segmentation in colour fundus images using Extreme Learning Machine,” *Computerized Medical Imaging and Graphics*, vol. 55, no. 5, pp. 68–77, 2017.

[11] 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 Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 16–27, 2017.

[12] M. Tarom and M. M. Mohammadi, "Segmentation of retinal blood vessel based on Radon Fourier Transform and Gaussian Process," 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP), Isfahan, Iran, 2017, pp. 105-109, doi: 10.1109/IranianMVIP.2017.8342377.

[13]

[14] Fan D, Cheng M, Liu Y, Li T, Borji A (2017) Structure-measure: a new way to evaluate foreground maps, IEEE international conference on computer vision (ICCV). Venice 2017:4558–4567

[15] Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels withdeepneural networks. TMI 35, 2369–2380, 2016.

[16] Orlando, J.I., Prokofyeva, E., Blaschko, M.B.: Adiscriminativelytrained fully connected conditional random field model for bloodvesselsegmentation in fundus images. IEEE Trans. Biomed. Eng. 64(1), 16–27, 2017.

[17] Yan, Z., Yang, X., Cheng, K.T.: Joint segment-Level andpixel-Wiselosses for deep learning based retinal vessel segmentation. IEEETrans.Biomed. Eng. 65(9), 1912–1923, 2018.

[18] Y, Wu., Y, Xia., Y, Song., Y, Zhang., W, Cai.: Multiscalenetworkfollowed network model for retinal vessel segmentation. In: Frangi,A.,Schnabel, J., Davatzikos, C., Alberola-Lopez, C., Fichtinger,G. (eds.)MICCAI 2018. LNCS, vol. 11071, pp. 119–126. Springer, Heidelberg,2018.