25th International Conference on Pattern Recognition (ICPR 2020)
January 10-15, 2021 Milan, Italy

SA-UNet: Spatial Attention Network for Retinal Vessel Segmentation

Changlu Guo[1],*, Márton Szemenyei[1], Yugen Yi[2],*, Wenle Wang[2], Buer Chen[1], Changqi Fan[1]

[1]Budapest University of Technology and Economics, Budapest, Hungary

[2]Jiangxi Normal University, Nanchang, China

1. Introduction

The precise segmentation of retinal blood vessels is of great significance for early diagnosis of eye-related diseases such as diabetes and hypertension. Due to the excellent performance of U-Net, many recent methods for retinal blood vessel segmentation are based on U-Net. Although previous U-Net variants perform well, they inevitably make the network more complex and less interpretable. In order to address these problems, we introduce spatial attention in U-Net and propose a lightweight network model, which we named Spatial Attention U-Net (SA-UNet). We evaluate SA-UNet on two public retinal fundus image datasets: DRIVE and CHASE_DB1. Experiments prove that SA-UNet is an effective method for retinal vessel segmentation.

2. Proposed Method

- Network Architecture

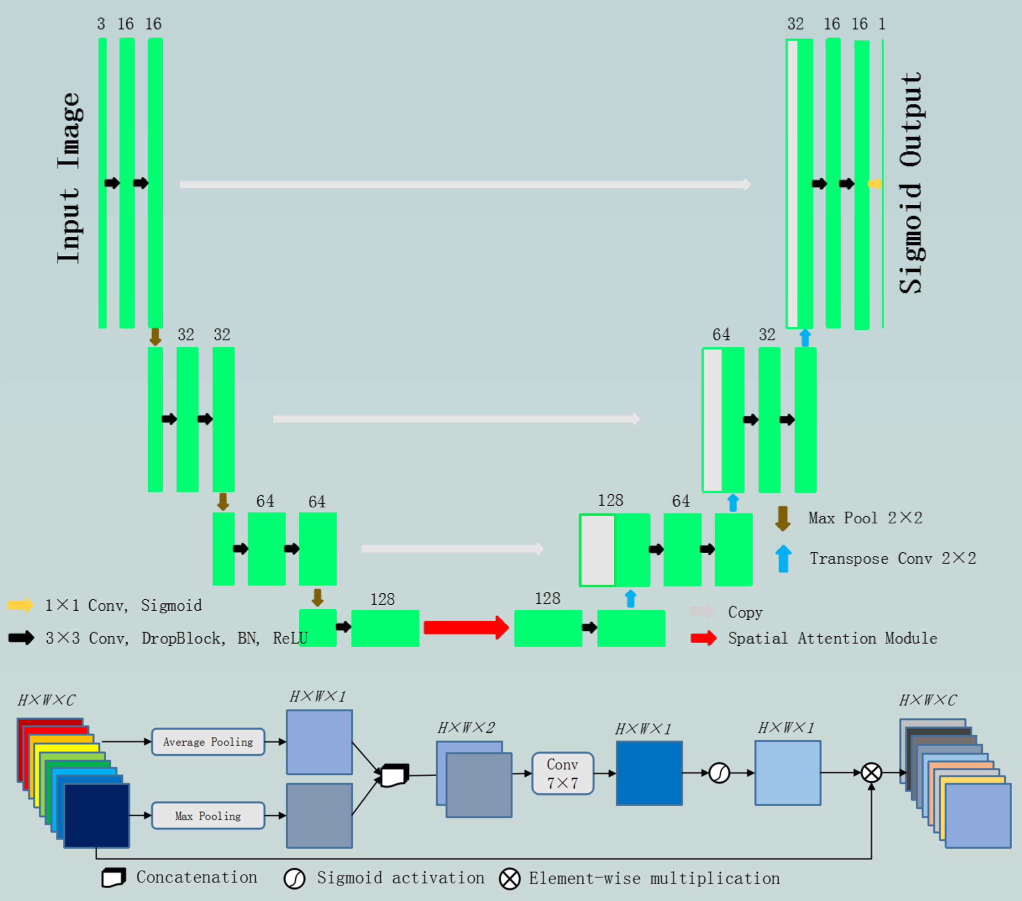

Figure. 1

Figure 1 shows the network architecture of SA-UNet and the spatial attention module. SA-UNet with a U-shaped encoder (left side)-decoder (right side) structure. The spatial attention module is added between the encoder and the decoder. At the final layer, a 1×1 convolution and Sigmoid activation function is used to get the output segmentation map. The spatial attention can help the network focus on important features and suppress unnecessary ones to improve the network's representation capability

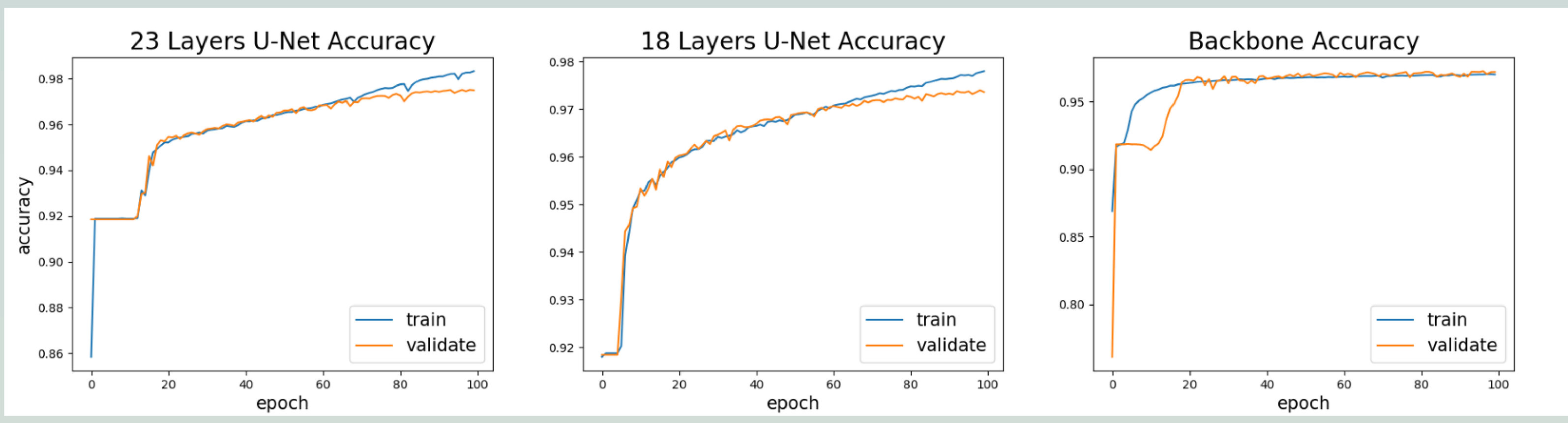

Figure. 2

- Structured Dropout Convolutional Block

Although data augmentation is performed for the original datasets, serious overfitting is still observed during original U-Net training, as shown in Fig. 2 (left). Therefore a lightweight U-Net with 18 convolutional layers is employed as our basic architecture, but it still has over-fitting problem, as shown in Fig. 2 (middle). Motivated by the successful application of DropBlock in recent computer vision works, we adopt DropBlock to regularize the network.

DropBlock, a structured form of dropout, can effectively prevent over-fitting problems in convolutional networks. We construct a structured dropout convolutional block, that is, each convolutional layer is followed by a DropBlock, a layer of batch normalization (BN) and a ReLU activation unit, as shown in the right side of Fig. 3. We employ this structured dropout convolutional block to build a U-shaped network as our "Backbone". As shown in Fig. 2. (left), the over-fitting problem is perfectly solved and accelerates the convergence of the network.

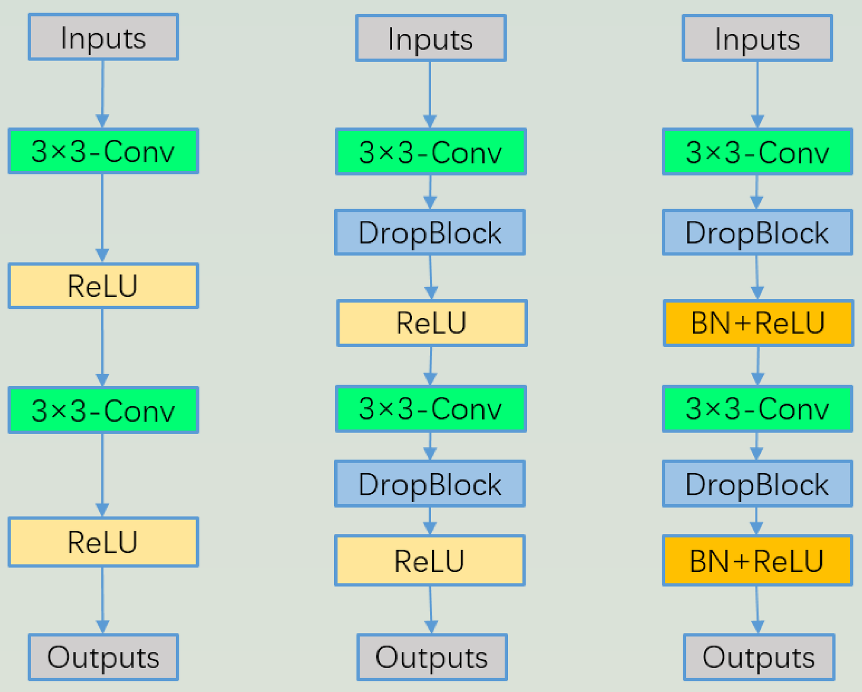

Figure. 3

3. Experiments

- Databases

We evaluate our proposed SA-UNet on two public retinal fundus image datasets: DRIVE and CHASE DB1. We use the following data augmentation methods: (1) Random rotation; (2) adding Gaussian noise; (3) color jittering; (4) horizontal, vertical and diagonal flips.

- Evaluation Metrics

For the purpose that we can estimate the performance of our proposed SA-UNet, the following metrics are employed: Sensitivity (*SE*), Specificity (*SP*), F1-score (*F1*), Accuracy (*ACC*), Area Under the ROC Curve (*AUC*) and Matthews Correlation Coefficient (*MCC*) .

- Results

TABLE I. 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**	**0.8097**

TABLE II. 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 III. AMOUNT OF PARAMETERS ON DIFFERENT MODELS.

Models	Total	Trainable	Non-trainable
AG-Net	9,335,340	9,335,340	0
23 Layers U-Net	2,158,705	2,158,705	0
18 Layers U-Net	535,793	535,793	0
U-Net + SA	535,891	535,891	0
SD-Unet	535,793	535,793	0
Backbone	538,609	537,201	1,408
SA-UNet	538,707	537,299	1,408

From the results in tables I, II and III, we could obtain several useful observations: (1) The strategy of introducing spatial attention is effective. (2) In the case of using structured dropout convolutional block based on U-Net, it demonstrates the effectiveness of adopting the newly constructed structured dropout convolutional block to build the Backbone. (3) Adding the batch normalization (BN) can improve the network performance to a certain extent. (4) Finally, the proposed SA-UNet achieves the best performance on most metrics, and compared with AG-Net and the original U-Net, our SA-UNet has a much smaller amount of parameters, so for the task of retinal vessel segmentation, SA-UNet is a lightweight and effective network. Figure 4 shows the sample segmentation results of different models.

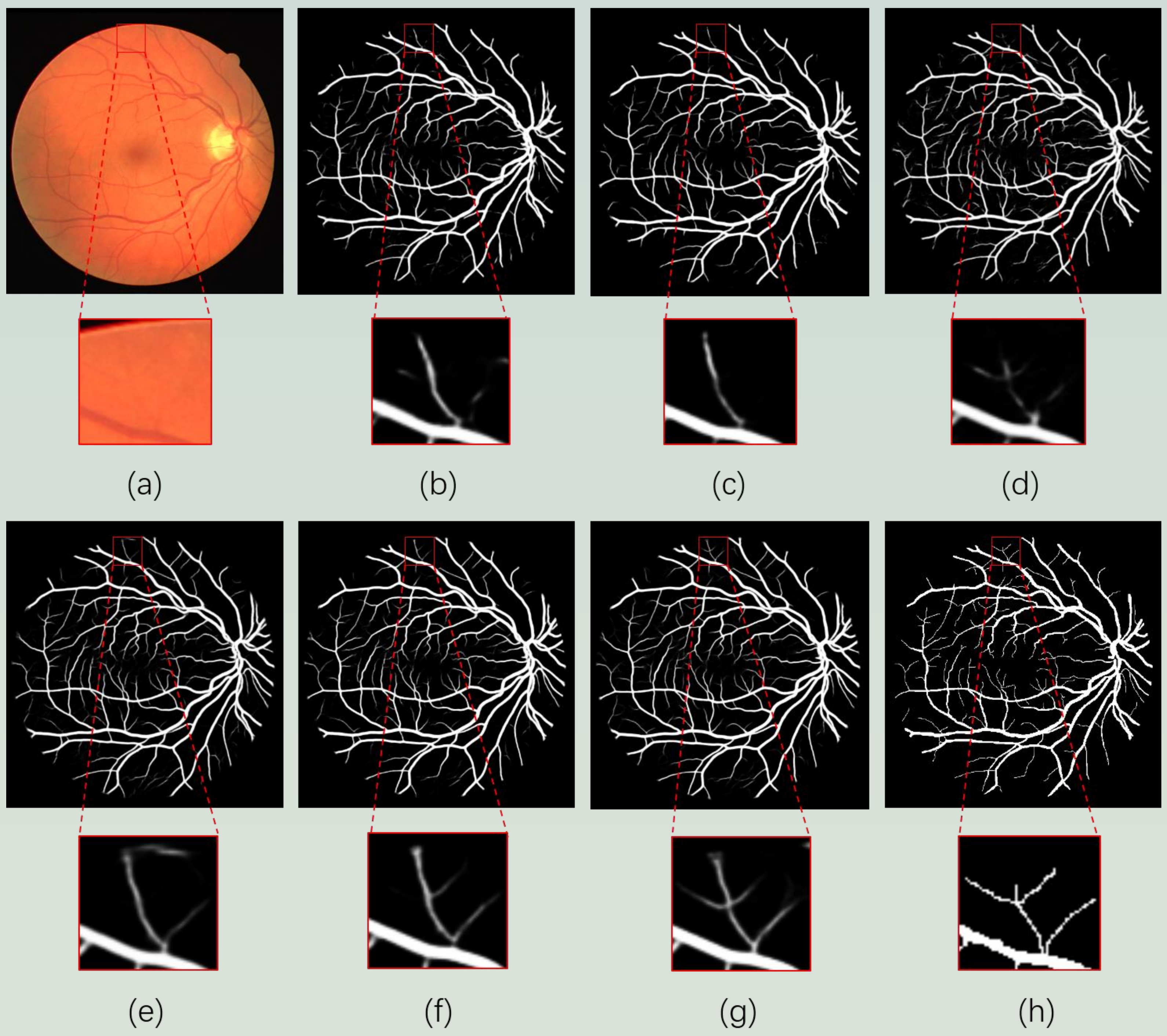

Figure. 4. A test image from DRIVE dataset; (b) Segmentation result by U-Net; (c) Segmentation result by U-Net+SA; (d) Segmentation result by AG-Net; (e) Segmentation result by SD-Unet; (f) Segmentation result by Backbone; (g) Segmentation result by SA-UNet; (h) Corresponding ground truth segmentation.

4. Conclusion

Inspired by the successful application of DropBlock and batch normalization in convolutional neural networks, we replace the convolutional block of U-Net with a structured dropout convolutional block that integrates DropBlock and batch normalization as our Backbone. In addition, in the retinal fundus images, the difference between the blood vessel area and the background is not obvious, especially the edges and small blood vessels. To help the network learn these, we add a spatial attention module between the encoder and decoder of the Backbone and propose Spatial Attention U-Net (SA-UNet). The experimental results demonstrate that using structured dropout convolutional blocks and the introducing spatial attention are effective, and SA-UNet is a very promising method for retinal vessel segmentation.

BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
Faculty of Electrical Engineering and Informatics