MŰEGYETEM 1782

SA-UNet: Spatial Attention Network for Retinal Vessel Segmentation

Changlu Guo, Márton Szemenyei, Yugen Yi*, Wenle Wang, Buer Chen, Changqi Fan*

https://github.com/clguo/SA-UNet

clguo.ai@gmail.com

MŰEGYETEM 1782

Abstract

The precise segmentation of retinal blood vessels is of great significance for early diagnosis of eye-related diseases such as diabetes and hypertension. In this work, we propose a lightweight network named Spatial Attention U-Net (SA-UNet) that does not require thousands of annotated training samples and can be utilized in a data augmentation manner to use the available annotated samples more efficiently.

MŰEGYETEM 1782

Introduction

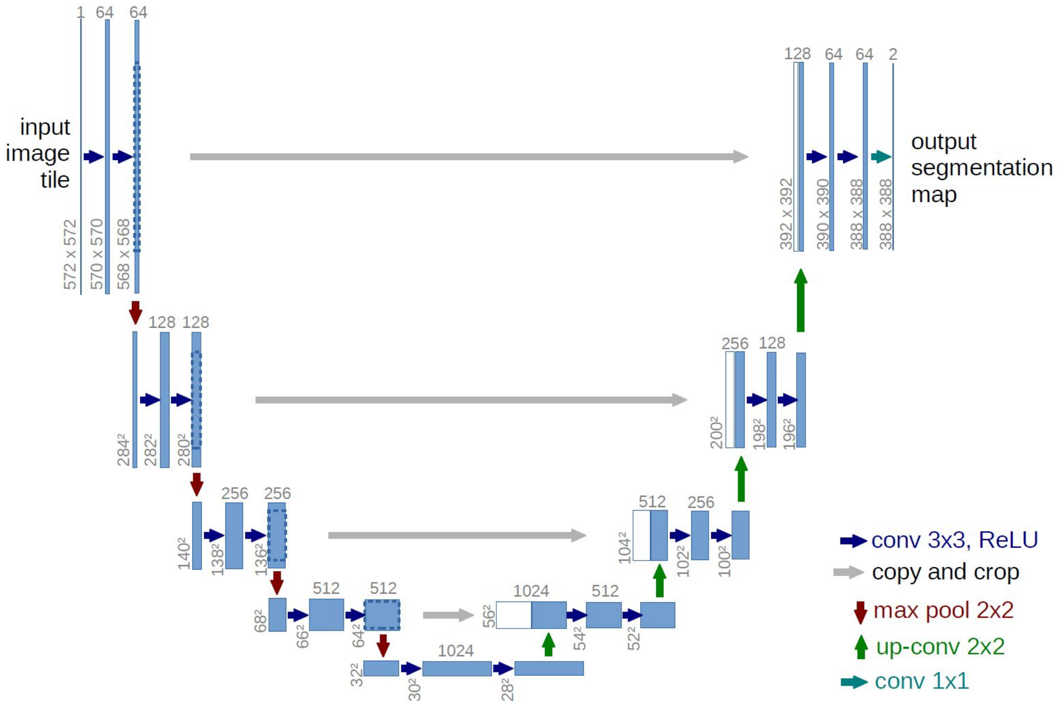
Due to the excellent performance of U-Net, many recent methods for retinal blood vessel segmentation are based on U-Net. Although previous U-Net variants perform well, they inevitably make the network more complex and less interpretable. In order to address these problems, we introduce spatial attention 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. Compared with other existing state-of-the-art methods for retinal vascular segmentation, our proposed SA-UNet achieves state-of-the-art performance.

I.

Methodology

U-Net

The famous U-Net that combines coarse features with fine features through skip connections to have superior performance in the field of medical image processing.



DropBlock

Its main difference from dropout is that it drops contiguous regions from a feature map of a layer instead of dropping out independent random units

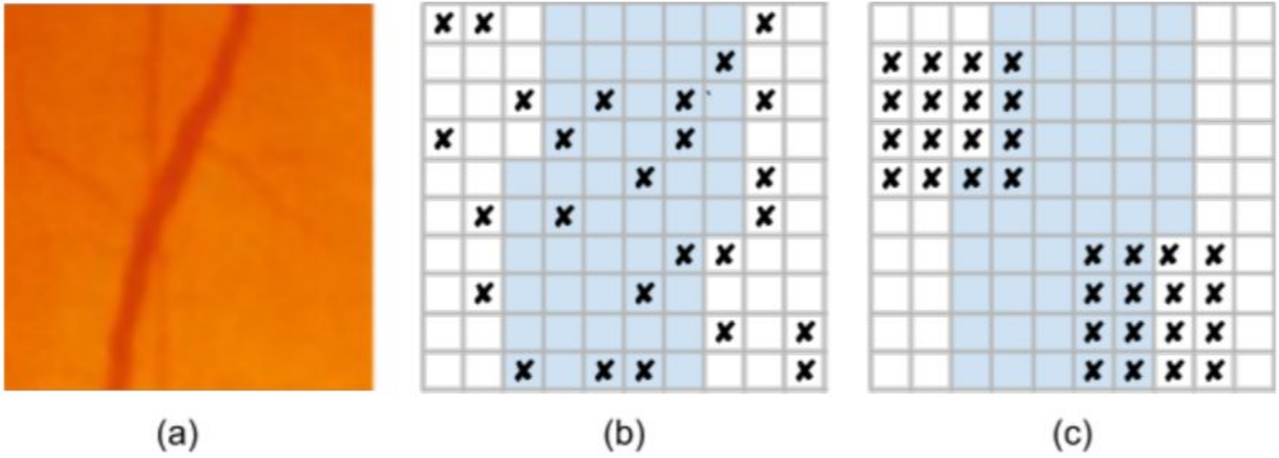
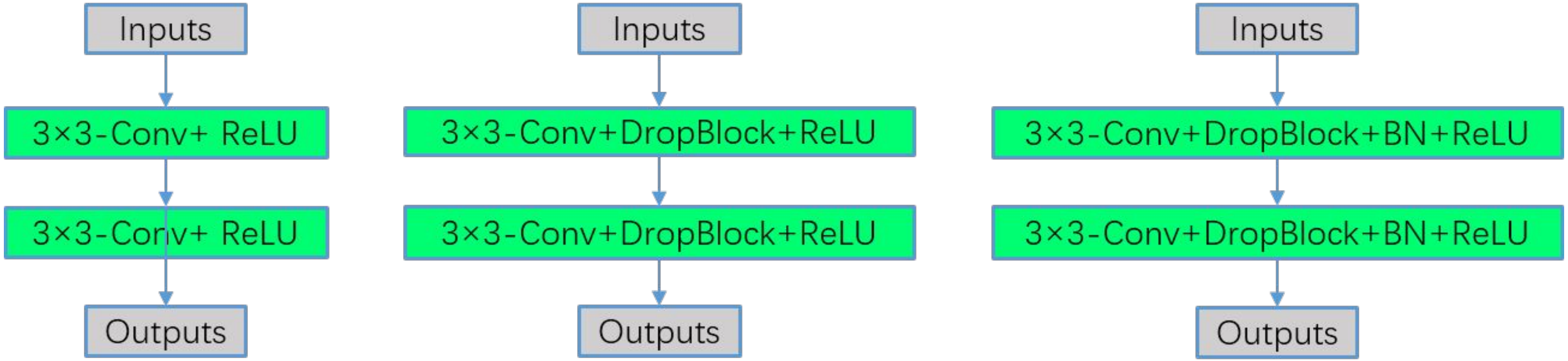


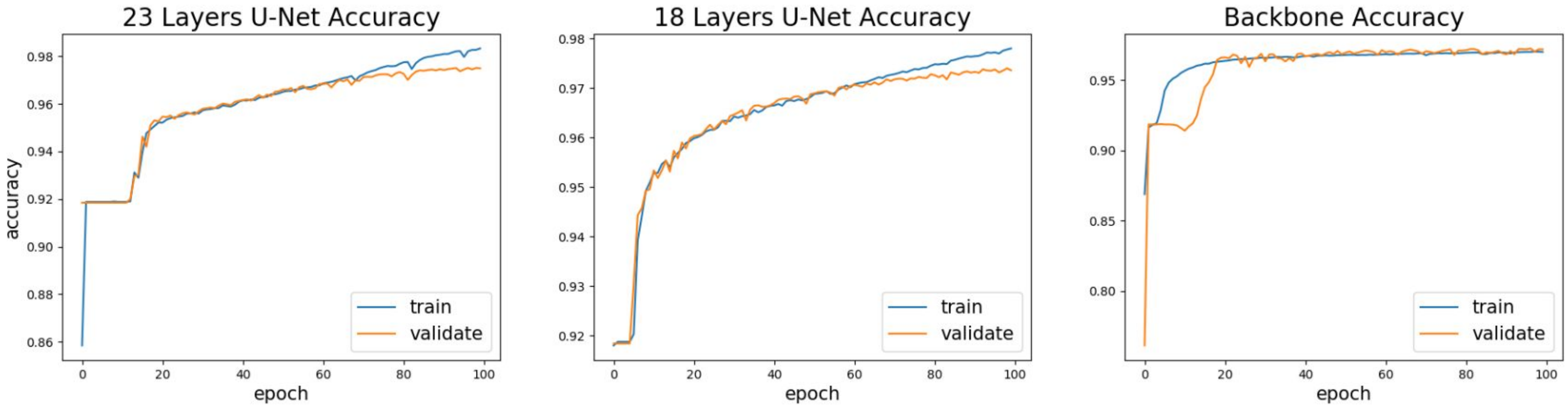
Figure 2.3: (a) A region of the original image,
(b) and (c) show the dropout situation in Dropout and DropBlock respectively.

Structured Dropout Convolutional Block



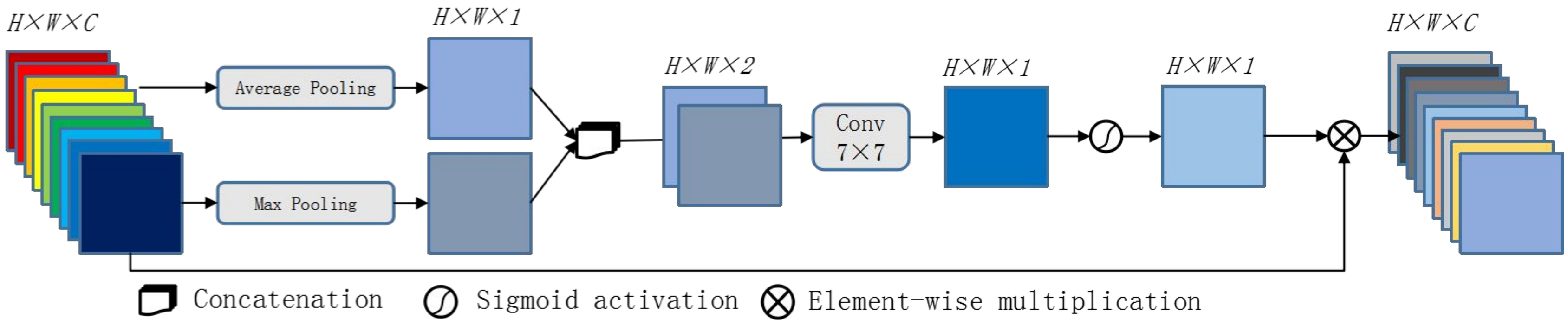
Original U-Net block (left), SD-Unet block (middle), Structured dropout convolutional block(right)

Batch normalization (BN) can improve the convergence speed of the network.



Compared to the 23 convolutional layers of the original U-Net, our Backbone has only 18 convolutional layers, and as shown in Fig. (left), the over-fitting problem is perfectly solved and accelerates the convergence of the network.

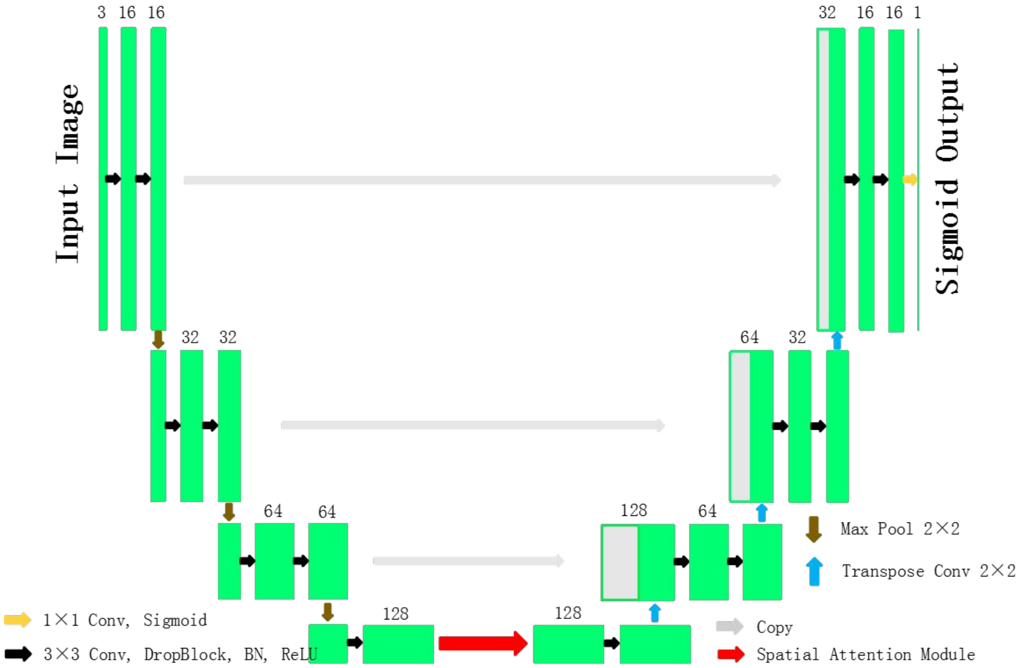
Spatial Attention Module



The spatial attention map enables the network to enhance important features (e.g. vascular features) and suppress unimportant ones

SA-UNet Architecture

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.



MŰEGYETEM 1782

I.

Expriments and Results

Datasets

Datasets	**DRIVE**	**CHASE_DB1**
Obtained from	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
Augmentation methods	(1) Random rotation; (2) adding Gaussian noise; (3) color jittering; (4) horizontal, vertical and diagonal flips.	

Ablation Experiments

(1) Spatial attention is effective. (2) 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) 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.

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**	**0.8097**

TABLE IV. 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

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**

Comparisons with state-of-the-art methods

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

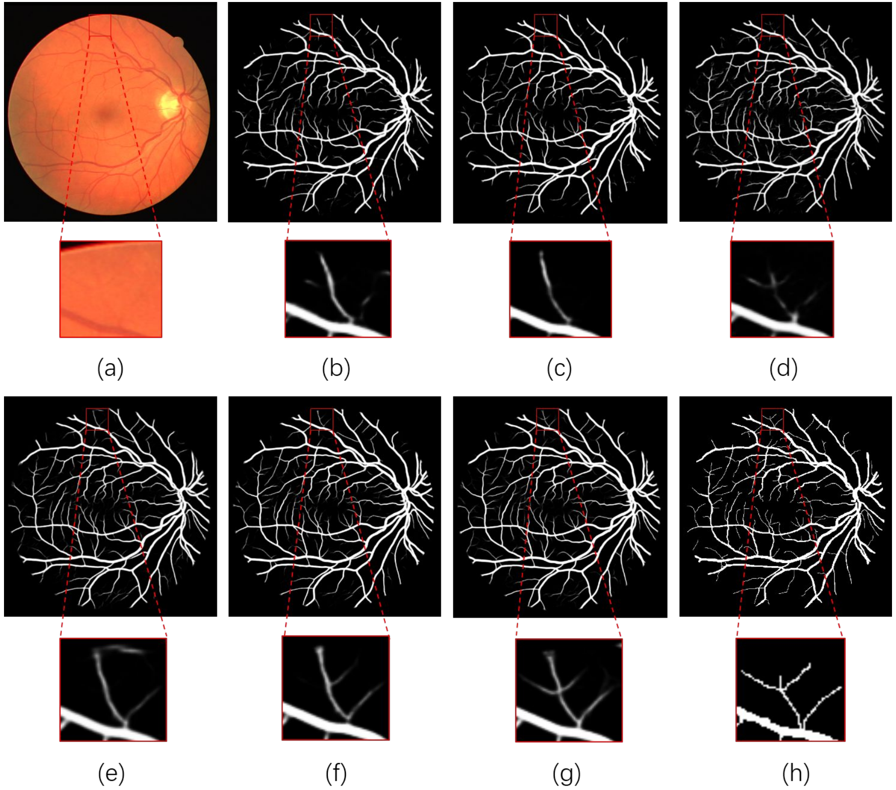
Dataset	DRIVE				
Metrics	Year	*SE*	*SP*	*ACC*	*AUC*
Liskowski et .al. [15]	2016	0.7811	0.9807	0.9535	0.9790
Orlando et. al. [16]	2017	0.7897	0.9684	0.9454	0.9507
Yan et. al. [17]	2018	0.7653	0.9818	0.9542	0.9752
MS-NFN [18]	2018	0.7844	0.9819	0.9567	0.9807
DEU-Net [7]	2019	0.7940	0.9816	0.9567	0.9772
Vessel-Net [8]	2019	0.8038	0.9802	0.9578	0.9821
AG-Net [9]	2019	0.8100	**0.9848**	0.9692	0.9856
SA-UNet	**2020**	**0.8212**	0.9840	**0.9698**	**0.9864**

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

Datasets	CHASE_DB1				
Metrics	Year	*SE*	*SP*	*ACC*	*AUC*
Liskowski et .al. [15]	2016	0.7816	0.9836	0.9628	0.9823
Orlando et. al. [16]	2017	0.7277	0.9712	0.9458	0.9524
Yan et. al. [17]	2018	0.7633	0.9809	0.9610	0.9781
MS-NFN [18]	2018	0.7538	0.9847	0.9637	0.9825
DEU-Net [7]	2019	0.8074	0.9821	0.9661	0.9812
Vessel-Net [8]	2019	0.8132	0.9814	0.9661	0.9860
AG-Net [9]	2019	0.8186	**0.9848**	0.9743	0.9863
SA-UNet	**2020**	**0.8573**	0.9835	**0.9755**	**0.9905**

Segmentation results

(a) 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.



I.

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 by comparing with other state-of-the-art methods for retinal vessel segmentation, our lightweight SA-UNet achieves state-of-the-art performance.

MŰEGYETEM 1782

Thank You!

clguo.ai@gmail.com