Improved SA-UNet: SA-UNet Based Neural Network for Human Retinal Vessel Segmentation

Ahmatjan Mattohti (23020200156686), Jiayi Wei (31520211154090), Kaiwen Du (23020210156920), Qianji Di (31520210156934), Zhang Qian (31520211154113)

[1]Xia men University·school of information science and engineering
Haiyun Park, Xiamen University, West Zengcuo Aun Road, Siming District, Xiamen, Fujian, China

Abstract

Retinal vessel segmentation is an important step in early diagnosis of ocular diseases, which requires large spatial information and receiving field. There are some problems in traditional vessel segmentation methods, such as high storage overhead and low computational efficiency. U-Net is an excellent network for image segmentation, but it also has the problem that the receptive field is small and large spatial information cannot be obtained. Spatial Attention U-Net (SA-UNet) is a lightweight network, which introduces a spatial attention module to expand receptive field and obtain more spatial information. In this project, we utilize the SA-UNet as the baseline model, try to use depthwise over-parameterized convolutional layers instead of conventional convolutional layers in the SA-UNet. Furthermore, we add an attention module as a channel to exploit the inter-channel relationship of features. In terms of relevant experiments, DRIVE and CHASEDB1 are implemented as datasets to evaluate our model. Experimental results demonstrate the effectiveness of the improved SA-UNet.

Introduction

Human retina is a light-sensitive tissue with extremely rich vascular information, which is the only non-invasive and non-invasive visualization property. Physicians can diagnose patients' diseases by analyzing the number, angle, branching and curvature of retinal vessels, so the research and analysis of retina is very beneficial to biological sciences. Recently, some cutting-edge technologies are also used for retinal research. In these years, deep learning based feature learning methods are widely used for fundus retinal vessel segmentation, which is different from manual feature extraction methods, but requires another better classifier for the final vessel segmentation. Deep learning convolutional neural networks combine feature extraction and classifier with better generalization ability and robustness. Among them, U-Net is one of the early algorithms using full convolutional networks for semantic segmentation. The use of a symmetric U-shaped structure containing compressed and expanded paths, which is very innovative and has influenced to some extent the design of several segmentation networks that follow, and the name of the network is also taken from its U-shaped shape. U-Net is one of the earlier algorithms using multi-scale features for semantic segmentation tasks, and its structure has inspired many of the later algorithms. However, it also has certain drawbacks. Firstly, effective convolution increases the difficulty and universality of model design, and many current algorithms directly use same convolution, which can also eliminate the operation of cutting edges before Feature Map merging. Secondly, it is not symmetric with the Feature Map by cutting the edges. Therefore, for complex inputs such as the human retina, models are needed that are more robust and conform to the unique characteristics of the retina.

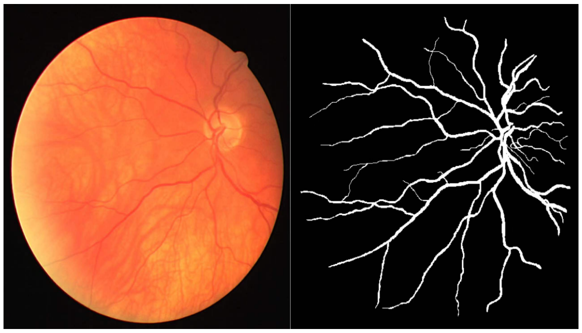

Figure 1: An example of Retinal vessel segmentation

In U-Net, the expansion and contraction paths in the network structure are essentially symmetrical, producing a U-shaped structure, which results in a U-Net. In medical image segmentation, U-NET has been widely used in recent years and has shown good performance. an important modification of U-Net is that there are a large number of feature channels in the upsampling part, which allows the network to propagate contextual information to higher resolution layers. Two identical U-Net models form the MS-NFN model for retinal fundus vascular segmentation (Wu et al. 2018). deU-Net uses large convolution kernels to preserve spatial information and multiscale features to obtain more semantic information, thus expanding the spatial and perceptual fields. SD-unet(Gadosey et al. 2020) proposed a mass normalization algorithm, which uses a normalization method to ensure accuracy, reduce computation time, shrink the model size, and reduce the parameters by nearly 8 times. Although U-Net variants perform well in the retinal vessel segmentation task, they inevitably make the network more complex, less interpretable, and not tailor-made for the retina as an input. so in this project, we use a lightweight network called Spatial Attention U-Net (SA-UNet) as a baseline. SA-UNet introduces a spatial attention module and structured discarded convolutional blocks instead of the original convolutional blocks of U-Net to prevent the network from overfitting. We use deep hyperparametric convolutional layers (DO-Conv) instead of some traditional convolutional layers in SA-UNet to improve the accuracy of retinal vessel segmentation. In addition, we added a channel attention module before the spatial attention module, which can exploit the inter-channel relationships of features. The evaluation of the project in this paper is based on the Children's Heart and Health Study (CHASE-DB1) dataset.

Related Work

In the past studies, automatic analysis of the retinal vascular system has become a hot topic in the field of medical imaging. Segmentation of retinal fundus vessels is mainly divided into manual segmentation and computer algorithm segmentation. Computerized algorithmic segmentation has gradually developed into a mainstream technique in the segmentation field because of its excellent performance in efficiency. Automatic segmentation algorithms can be divided into two categories. The first category is image processing algorithms, including pre-processing, segmentation and post-processing. For example, wavelet transform methods are used to enhance foreground and background for fast vessel detection(Peter et al. 2012). The other one is based on machine learning, which extracts feature vectors to train classifiers to determine whether pixels in retinal images belong to blood vessels or not.

In the last few years, emerging work has emerged using fully convolutional networks (FCN) to simultaneously segment and classify retinal vessels. Orlando(Orlando, Prokofyeva, and Blaschko 2017) successfully combined a dense conditional random field (CRF) model with CNN for retinal vessel segmentation, establishing remote links within images, thus addressing the problem of "systolic bias" , but the phenomenon of lesion mis-segmentation exists. Vessel segmentation, which better solves the problem of inadequate microvessel segmentation, still suffers from some microvessel breaks and easy chain-knotting of vessels.AlBadawi and Frazcite 2018Arterioles used FCN with encoder-decoder structure for pixel-by-pixel classification of arteries and veins. The deep learning based approach has demonstrated its potential for blood vessels. Modifying the FCN so that it has a large number of feature channels in the upsampling part allows the network to propagate contextual information to higher resolution layers. Most existing unsupervised and supervised retinal image segmentation methods rely on hand-crafted features to characterize the differences between vascular and non-vascular pixels. For example, the multiscale matched filter-based fundus segmentation method(Al-Rawi, Qutaishat, and Arrar 2007)) uses a segmented linear approximation of retinal vessels with a Gaussian-like intensity distribution to enhance the vessels before thresholding, and although it enhances most of the tiny vessels, there is still under-segmentation of vessel crossings and mis-segmentation of lesions. Soares(Soares et al. 2006) uses two-dimensional Gabor filters at different scales as an effective alternative to train classifiers for vascular pixel detection, but the method still suffers from broken microvessel segmentation. The network proposed by (Shi et al. 2015) et al. can extract most of the vascular features better, but the resistance to noise is poor and the heterogeneity causes microvessel segmentation breakage.

Although the above methods achieve better segmentation results, these artificially selected features are still not robust enough in solving the two problems of vascular change trends and invariance of vascular information, causing problems such as under-segmentation of microvessels, lesion and optic disc segmentation errors.

Proposed Solution

In this project, we use a lightweight network named Spatial Attention U-Net (referred to as SA-UNet) (Guo et al. 2021) as the baseline, and this network 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. The SA-UNet introduces a spatial attention module (Woo et al. 2018) which infers the attention map along the spatial dimension, and multiplies the attention map by the input feature map for adaptive feature refinement. In addition, the SA-UNet employs structured dropout convolutional blocks instead of the original convolutional blocks of U-Net (Ronneberger, Fischer, and Brox 2015) to prevent the network from overfitting.

The SA-UNet is a U-shaped network architecture with encoder-decoder structure. In encoder, every step contains a structured dropout convolutional block and a 2×2 max pooling operation. The convolutional layer of each convolutional block is followed by a DropBlock, a batch normalization (BN) layer and a rectified linear unit (ReLU), and then the max pooling operation is utilized for down-sampling with a stride size of 2. At the same time the number of feature channels is doubled every down-sampling step. In decoder, each step contains a 2×2 transposed convolution operation for up-sampling and halves the number of feature channels. The spatial attention module is added between the encoder and the decoder, which infers the attention map along the spatial dimension. At the final layer, a 1×1 convolution and a Sigmoid activation function are used to obtain the output segmentation map. The SA-UNet also implement the fusion of features at different scales, which improves the accuracy of the model. The coarse feature map captures the context information and highlights the classification and position of the foreground object. In order to link the coarse-level and fine-level dense predictions, the feature maps distilled from different scales are merged by a skip connection.

In order to effectively prevent over-fitting of the network, the SA-UNet adopts DropBlock (Ghiasi, Lin, and Le 2018) to regularize the network, and as a structured form of dropout, DropBlock can effectively prevent over-fitting problems in convolutional networks (Guo et al. 2019). Its primary difference from dropout is that it discards contiguous areas from a feature map of a layer instead of dropping independent random units. Based on this, the SA-UNet construct a structured dropout convolutional block, and each

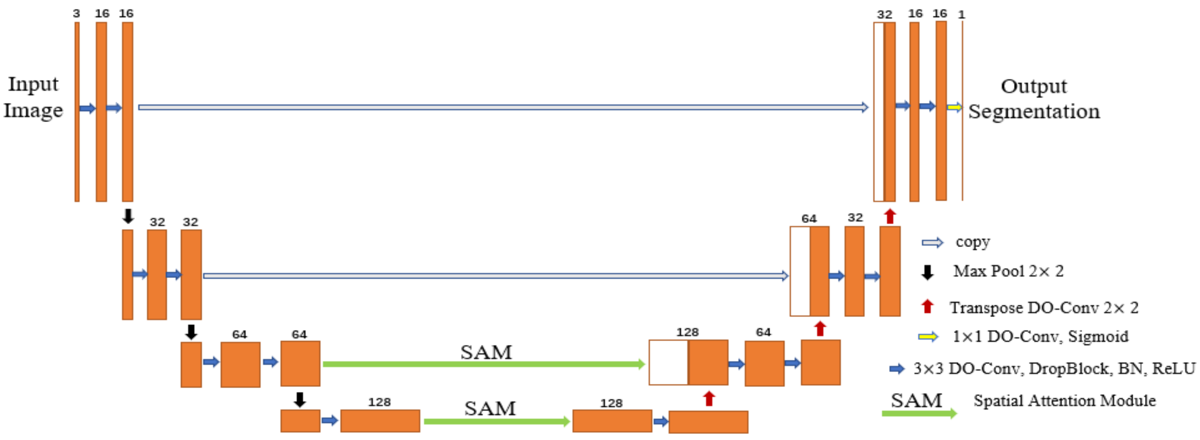

Figure 2: The improved network architecture of SA-UNet with a U-shaped encoder-decoder structure.

convolutional layer is followed by a DropBlock, a layer of batch normalization and a ReLU activation unit. Unlike the convolutional block of SD-Unet (Guo et al. 2019), the structured dropout convolutional block introduces batch normalization to accelerate network convergence. The results show that the overfitting problem is perfectly solved and accelerates the convergence of the network. The another important components of SA-UNet is the spatial attention module (SAM), which infers the attention along the spatial dimension, and multiplies the attention map by the input feature map for adaptive feature refinement (Woo et al. 2018). The spatial attention uses the spatial relationship between features to produce a spatial attention map. To calculate spatial attention, the spatial attention first applies maxpooling and average-pooling operations along the channel axis and concatenate them to produce an efficient feature descriptor. Then a convolutional layer followed by the Sigmoid activation function on the concatenated feature descriptor is used to generate a spatial attention map. The spatial attention can help the network focus on important features and suppress unnecessary ones to improve the network's representation capability.

We achieved some improvements based on the SA-UNet. Fig. 2 shows the improvement network architecture of SA-UNet with a U-shaped encoder (left side)-decoder (right side) structure. Firstly, we used depth-wise over-parameterized convolutional (referred to as DO-Conv) (Cao et al. 2020) layers instead of ordinary convolutional layers in the SA-UNet. The DO-Conv is a novel and generic way for boosting the performance of CNNs, and it is helpful for many image segmentation tasks. In addition, DO-Conv does not introduce extra computation at the inference phase. Secondly, we used another spatial attention module to instead the skip connection top of the the encoder and the decoder. The experimental results show that the DO-Conv and the added spatial attention module improve the performance of retinal vessel segmentation.

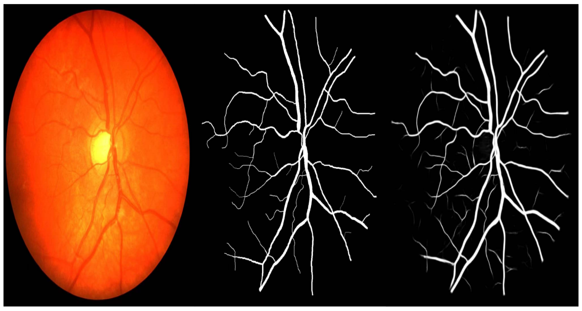

Figure 3: Demonstration of our model: the images from left to right are the original image, ground truth, and our output.

Experiments

A.Dataset

We used two data sets to evaluate our SA-UNet model, namely DRIVE and CHASE DB1. Both of these two data sets are related to the public retina, see Table 1 for details.

Table 1: DRIVE AND CHASE DB1.

Datasets	DRIVE	CHASE DB1
Total number	40	28
Train/Test number	20/20	20/8
Resolution	584*565	999*960
Resize	592*592	1008*1008

We can notice that the size of the two data sets cannot be directly input to the network. Therefore, we used zero padding to change its size.We used data augmentation methods, such as random rotation, adding Gaussian noise, color jittering, horizontal, vertical and diagonal flips. We augment the two original datasets from the original 20 training images to 256 images.

B. Implementation Detail

In order to test whether the training of our model is over-fitting, we randomly selected 26 and 13 images in the DRIVE and CHASE augmented datasets as auxiliary tests in the validation set.We used the following method to retrain SA-UNet using the augmented datasets. For the two datasets, we used the Adam optimizer and the classified cross-entropy loss function. In order to reduce the complexity and training time, we are the first The filters of each convolutional layer are only 16. The epoch is 100, the learning rate in the first 60 is set to 0.001, and the learning rate in the last 40 of the epoch is set to 0.0001.The size of the discard blocks of DropBlock is set to 7. Respectively, for DRIVE dataset, the batch size of the training is set to 8 and the dropout rate of DropBlock is set to 0.18. For CHASE DB1, the batch size is set to 4 and the dropout rates is 0.13. We use Keras-based Tensorflow for training, and all training is performed in Google Colab.

C. Evaluation Metrics

In order to better evaluate model performance, we use true positive (TP), false positive (FP), false negative (FN), and true Negative (TN) as the division result, by comparing the division result and the comparison of each pixel. Then, the sensitivity (SE), specificity (SP), F1- score (F1), and accuracy (ACC) are used to evaluate the performance of the model. Regarding the retinal blood vessel segmentation task, only about ten percent of the pixels belong to blood vessels, and the others are considered background.The Matthews Correlation Coefficient (MCC) is suitable for observing binary classification problems of different sizes.Therefore, the MCC value can help find the optimal setting for the vessel segmentation algorithm. MCC is defined as:

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

The area under the ROC curve (AUC) can be used to measure the performance of the segmentation. If the AUC value is 1, it means perfect segmentation.

D. Overall Results

By comparing our method with other segmentation methods on the CHASE and DRIVE test set, we demonstrate that our method can achieve favorable segmentation performance, as shown in Table 2 and Table 3. Particularly, compared with the baseline SA-UNet, our improved SA-UNet achieves superior results.

Table 2: Results on CHASE DB1.

Method	SE	SP	ACC	AUC	F1	MCC
U-net	0.7677	0.9857	0.9666	0.9789	0.8012	0.7839
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
SA-Unet	0.8572	0.9834	0.9755	0.9904	0.8152	0.8033
Ours	0.8613	0.9822	0.9796	0.9910	0.8188	0.8042

Table 3: Results on DRIVE.

Method	SE	SP	ACC	AUC	F1	MCC
U-net	0.7842	0.9861	0.9733	0.9838	0.7875	0.7733
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
SA-Unet	0.8234	0.9828	0.9689	0.9859	0.8226	0.8056
Ours	0.8285	0.9819	0.9749	0.9881	0.8235	0.8074

Conclusion

The analysis of human retina helps to help diagnose related conditions, and machine learning for retinal vessel segmentation has also been proven to be efficient. In previous image segmentation algorithms, neural networks often appear to be less robust as well as overly complex, and this paper proposes some improvements to the retinal vessel segmentation algorithm based on SA-UNet. In the present algorithm, the model takes into account the spatial attention problem in the image, but also ensures that the algorithm is robust and lightweight, based on this, we add an attention channel to make the model more suitable for the retinal blood vessel characteristics. The experimental results show that our modifications improve the effectiveness of the model for retinal vessel segmentation.

References

Al-Rawi, M.; Qutaishat, M.; and Arrar, M. 2007. An improved matched filter for blood vessel detection of digital retinal images. *Computers in Biology Medicine*, 37(2): 262–267.

Cao, J.; Li, Y.; Sun, M.; Chen, Y.; Lischinski, D.; Cohen-Or, D.; Chen, B.; and Tu, C. 2020. Do-conv: Depthwise over-parameterized convolutional layer. *arXiv preprint arXiv:2006.12030*.

Gadosey, P. K.; Li, Y.; Agyekum, E. A.; Zhang, T.; Liu, Z.; Yamak, P. T.; and Essaf, F. 2020. Sd-unet: Stripping down u-net for segmentation of biomedical images on platforms with low computational budgets. *Diagnostics*, 10(2): 110.

Ghiasi, G.; Lin, T.-Y.; and Le, Q. V. 2018. Dropblock: A regularization method for convolutional networks. *arXiv preprint arXiv:1810.12890*.

Guo, C.; Szemenyei, M.; Pei, Y.; Yi, Y.; and Zhou, W. 2019. SD-UNet: A structured dropout U-Net for retinal vessel segmentation. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 439–444. IEEE.

Guo, C.; Szemenyei, M.; Yi, Y.; Wang, W.; Chen, B.; and Fan, C. 2021. Sa-unet: Spatial attention u-net for retinal vessel segmentation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, 1236–1242. IEEE.

Orlando, J. I.; Prokofyeva, E.; and Blaschko, M. B. 2017. A Discriminatively Trained Fully Connected Conditional Random Field Model for Blood Vessel Segmentation in Fundus Images. *IEEE Transactions on Biomedical Engineering*.

Peter, B.; Norman, S. C.; Graham, M.; Curtis, T. M.; and Teresa, S. G. 2012. Fast Retinal Vessel Detection and Measurement Using Wavelets and Edge Location Refinement. *PLoS ONE*, 7(3): e32435–.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Shi, C.; Guo, J.; Azzopardi, G.; Meijer, J.; Jonkman, M. F.; and Petkov, N. 2015. Automatic Differentiation of u- and n-serrated Patterns in Direct Immunofluorescence Images. In *Springer International Publishing*.

Soares, J.; Leandro, J.; Cesar, R. M.; Jelinek, H. F.; and Cree, M. J. 2006. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9): 1214–1222.

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

Wu, Y.; Yong, X.; Yang, S.; Zhang, Y.; and Cai, W. 2018. Multiscale Network Followed Network Model for Retinal Vessel Segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*.