Dual Encoding U-Net for Retinal Vessel Segmentation

Bo Wang[1,2], Shuang Qiu[2], and Huiguang He[1,2,3(✉)]

[1] School of Artifical Intelligence, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[2] Research Center for Brain-inspired Intelligence and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
huiguang.he@ia.ac.cn
[3] Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

Abstract. Retinal Vessel Segmentation is an essential step for the early diagnosis of eye-related diseases, such as diabetes and hypertension. Segmentation of blood vessels requires both sizeable receptive field and rich spatial information. In this paper, we propose a novel Dual Encoding U-Net (DEU-Net), which have two encoders: a spatial path with large kernel to preserve the spatial information and a context path with multiscale convolution block to capture more semantic information. On the top of the two paths, we introduce a feature fusion module to combine the different level of feature representation. Besides, we apply channel attention to select useful feature map in a skip connection. Furthermore, low-level and high-level prediction are combined in multiscale prediction module for a better accuracy. We evaluated this model on the digital retinal images for vessel extraction (DRIVE) dataset and the child heart and health study (CHASEDB1) dataset. Results show that the proposed DEU-Net model achieved the state-of-the-art retinal vessel segmentation accuracy on both datasets.

Keywords: Retinal vessel segmentation · Spatial path · Context path · Attention mechanism

1 Introduction

Segmentation of blood vessels plays an important role in the diagnosis of eye-related diseases such as diabetics, hypertension and retinopathy of prematurity [3]. However, manual annotation of retina blood vessels by ophthalmologist is time consuming task and requires training and skill. Therefore, automatic segmentation of retinal blood vessels from funds images is particularly significant.

Many automatic retinal vessel segmentation algorithms have been reported in the past decades and can be divided into two broad categories. The first

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 84–92, 2019.
https://doi.org/10.1007/978-3-030-32239-7_10

category is image processing algorithms, including pre-processing, segmentation and postprocessing. For example, Bankhead et al. [2] developed wavelet transform approach to enhance the foreground and background for fast vessel detection. The other category is machine learning based algorithms, which utilizes extracted feature vectors to train a classifier to determine whether a pixel from retinal image belong to vessel or not. As a typical example, Lupascu et al. [5] constructed 41-D feature vector for each pixel, and an AdaBoost classifier was trained for classifying each pixel in retinal image.

Recently, deep learning method has shown its excellence in many computer version tasks. Since U-Net [8] has the encoder-decoder structure with skip connections which allows efficient information flow, it provides state-of-the-art performance in many medical image analysis. Wu et al. [11] proposed the multiscale network followed network (MS-NFN) model for retinal vessel segmentation and each submodel consists of two identical U-Net models. Zhuang et al. [12] reported a chain of multiple U-Nets (LadderNet), which has multiple pairs of encoder-decoder branches. However, in order to incorporate more spatial information of pixels into the pixel classification, all these variants just stack various U-Net, with low interpretability and high computational complexity.

This paper proposed a Dual Encoding U-Net (DEU-Net) model that greatly enhances deep neural networks' capability of segmenting vessels with an end-to-end and pixel-to-pixel manner. The main uniqueness of this model includes: (1) a spatial encoding path, which has a small stride and large kernel to preserve the spatial information; a context path, which has multiscale convolution modules to capture more semantic information; (2) channel attention mechanism is applied for skip connection to flow and select useful feature map. (3) low-level and high-level prediction are generated by decoding path and a multiscale predict module is introduced to fusion different scale feature for a better prediction. The proposed model has been evaluated on DRIVE and CHASEDB1 dataset. Results show that the proposed DEU-Net model achieved the state-of-the-art retinal vessel segmentation accuracy on both datasets.

2 Method

In this paper, we propose a Dual Encoding U-Net approach for retinal vessel segmentation. The proposed method for retinal vessel segmentation mainly consists of the following steps: (1) retinal image preprocessing, (2) patch extraction, (3) feeding each patch into Dual Encoding U-Net for segmentation, (4) segmentation result reconstruction. The overview of the retinal vessel segmentation framework is shown in Fig. 1.

2.1 Retinal Image Preprocessing and Patch Extraction

Retinal images usually comprise noise and uneven illumination, thus it is necessary to make an image enhancement before postprocessing. Firstly, we converted an RGB fundus image to a gray image, and the gray imaged was normalized.

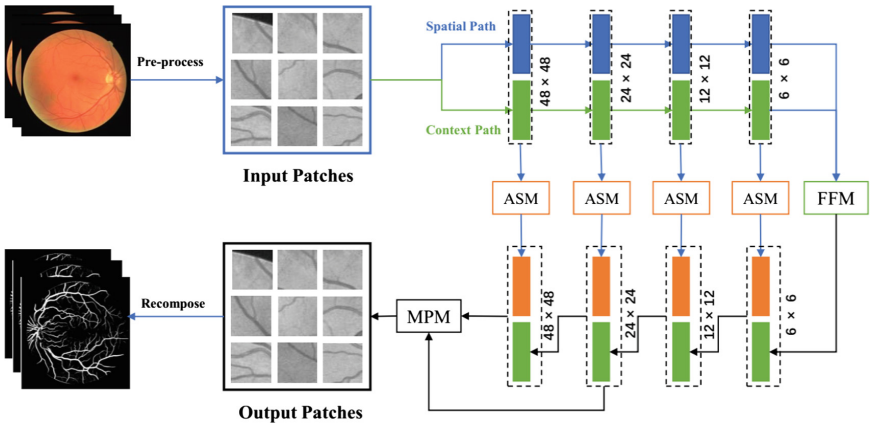


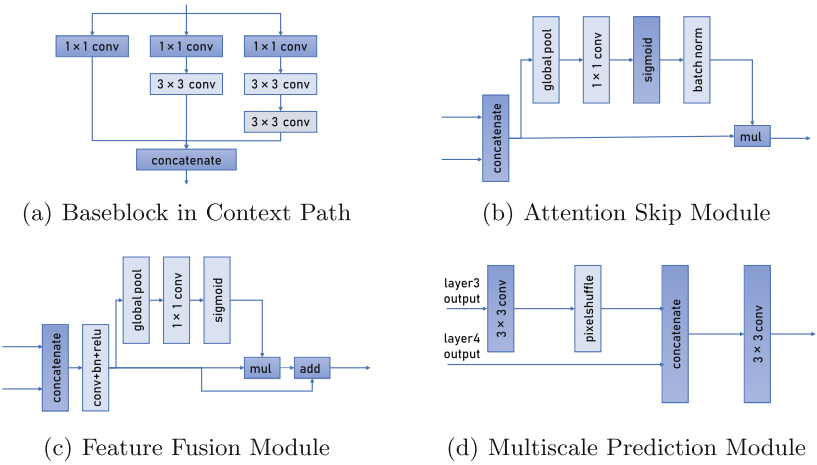
Fig. 1. Illustration of the Dual Encoding U-Net model-based retinal vessel segmentation. Components of the network architecture contains: Attention Skip Module(ASM), Feature Fusion Module (FFM) and Multiscale Predict Module (MPM).

Secondly, contrast limited adaptive histogram equalization and gamma adjustment was applied to improve the image contrast and the suppress noise. Finally, the intensity values are scaled range from 0 to 1. We randomly sampled 190,000 patches of size 48 × 48 from the DRIVE dataset and 760,000 patches from the CHASEDB1 dataset for training our model.

2.2 The Proposed Architecture

Inspired by U-Net [8], we proposed a Dual Encoding U-Net (DEU-Net) for retinal vessel segmentation. Figure 1 illustrates the network architecture. The proposed network has a U-shaped architecture with encoder and decoder. In encoder, a spatial path was designed to preserve the spatial information and a context path was used to capture more semantic information. Furthermore, channel attention mechanism was applied for skip connection to transfer information and select useful feature map. Finally, in decoder, low-level and high-level prediction were combined to predict a better accuracy.

Spatial Path. In semantic task, the spatial information and the receptive field are crucial to achieving high accuracy. However, it is hard to meet these two demands simultaneously. Refer to the Global Convolutional Network [7], we propose a spatial path with large kernels to encode the affluent spatial information from original input image. The spatial path contains four layers and each layer includes a convolution with stride = 7,12,9,6 in order, followed by batch normalization and ReLU. Since the input size of network is 48 × 48, these relatively large convolution kernels encode rich spatial information with large spatial size of feature maps.



(a) Baseblock in Context Path

(b) Attention Skip Module

(c) Feature Fusion Module

(d) Multiscale Prediction Module

Fig. 2. Structure of components in Dual Encoding U-Net

Context Path. The spatial path is used to encode spatial information, also, we design a context path to provide sufficient semantic information. The context path also contains four layers. Each layer includes a baseblock with numbers of size convolution filters (see Fig. 2(a)), which can improve the expressive ability of the convolution layer and raise efficiency of the network parameters.

Attention Skip Module. In skip layers, we propose a specific channel attention mechanism to select necessary feature map. As shown in Fig. 2(b), attention skip module employs global average pooling on the combined output of spatial and context path, which computes an attention vector to guide the feature map learning. This module is a self-attention mechanism, which can increase the network's sensitivity to informative features which is important in decoder without any supervisory information. Efficient information of encoder can make a better prediction.

Feature Fusion Module. The features of the two paths are different in level of feature representation. The spatial information captured by the spatial path encodes mostly rich detail information, which is known as low level. Moreover, the output feature of the context path mainly encodes context information, which is known as high level. Therefore, we introduce a feature fusion module to fuse these features as shown in Fig. 2(c).

Multiscale Predict Module. Given the different level of the prediction, a multiscale predict module is introduced to fusion different scale feature. PixelShuffle [9] is applied to upsample low level prediction feature to high resolution, which can preserve spatial information. Finally, low-level and high-level predictions with the same size are combined and followed by a convolution process (see Fig. 2(d)).

3 Experiments

3.1 Datasets

The DRIVE dataset consists of 40 color images of the retina, 20 of which were used for training and the remaining 20 images for testing. Each image has 584 × 565 pixels. The binary field of view (FOV) mask and segmentation ground truth are provided for each image in the dataset.

The CHASEDB1 dataset has 28 color images of the retina and the size of each image is 999 × 960. Usually, the first 20 images were used for training and the other 8 images were used for testing. The segmentation ground truth is provided for all 28 images in CHASE DB1, and FOV mask were obtained using manual method [10].

3.2 Training of the Neural Network

We randomly sampled 190,000 patches of size 48 × 48 from the training images in DRIVE, and used 10% of the training samples as validation data. For the CHASEDB1 dataset, we sampled 760,000 patches of size 48 × 48 from the training images, and used 10% of the training samples as validation. Instead of only cross-entropy loss for segmentation, we use a hybrid loss function that is a weighted sum of two terms. The first is a binary-class cross-entropy term that encourages the segmentation model to predict the right class label at each pixel location independently, which is defined as:

$$Loss_{ce}(y, \hat{y}) = -\sum y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \tag{1}$$

where both y_i is ground truth and $\hat{y}_i$ is predicted vectors. The second loss term is based on Intersection over Union (Iou), named the jaccard loss, which is defined as:

$$Loss_{jaccard}(y, \hat{y}) = 1 - \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|} \tag{2}$$

where y is ground truth image and $\hat{y}$ is predicted image. The jaccard loss can detect and correct higher-order inconsistencies between ground truth segmentation maps and the ones produced by the segmentation net, nor measured by a per-pixel cross-entropy loss. Finally, the joint loss function is:

$$Loss = \lambda_1 Loss_{ce} + \lambda_2 Loss_{jaccard} \tag{3}$$

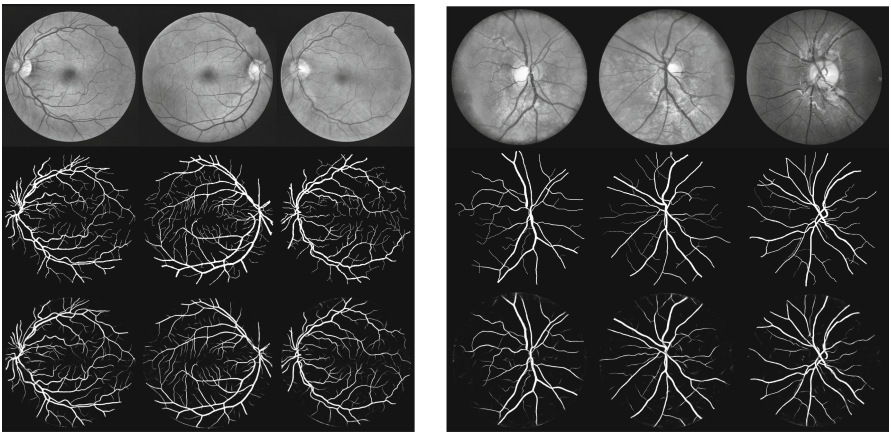
where λ_1 and λ_2 are the weight of two terms, satisfied with $\lambda_1 = 0.8$ and $\lambda_2 = 0.2$ in this paper. Adam optimizer with default parameters is applied to train the model and batch size is 256. Besides, we use "reduce learning rate on plateau" strategy, and set the learning rate as 0.01, 0.001, 0.0001 on epochs 0, 25 and 50 respectively, and set the total learning epochs as 200.

3.3 Evaluation Metrics

In retinal vessel segmentation, we divide the pixels in the vessel map into true positive (TP), false positive (FP), negative (FN) and true negative (TN) by comparing them with the corresponding ground truth labels. Then, accuracy (AC), sensitivity (SE), specificity (SP) and F1-score are used to evaluate the performance of DEU-Net. To further evaluate the performance of different neural networks, we also calculated the area under receiver operating characteristics curve (AUC).

4 Results

To evaluate the proposed Dual Encoding U-Net framework, we conduct experiments on the DRIVE and CHASEDB1 datasets. The retinal vessel segmentation results of DEU-Net are shown in Fig. 3, where in each column the input image, the ground truth, and the result of our segmentation method are shown from top to bottom. From Fig. 4, it can be observed that DEU-Net produces more distinct vessel segmentation results and preserve more details than Laddernet [12].



(a) Test results on DRIVE dataset. (b) Test results on CHASE-DB1 dataset.

Fig. 3. Test results on DRIVE and CHASE DB1 dataset. From top to bottom: input image, ground truth and predictions.

A quantitative evaluation of the results obtained on our experiments is presented in Tables 1 and 2. The experimental result shows that the proposed method is competitive with other existing methods by achieving the highest F1-score, sensitivity, specificity and accuracy for both tasks. DEU-Net also generates high AUC on two tasks. So the proposed method is a robust tool for vessel segmentation and the comparison demonstrate that the proposed method is very competitive with the stage-of-the-art methods.

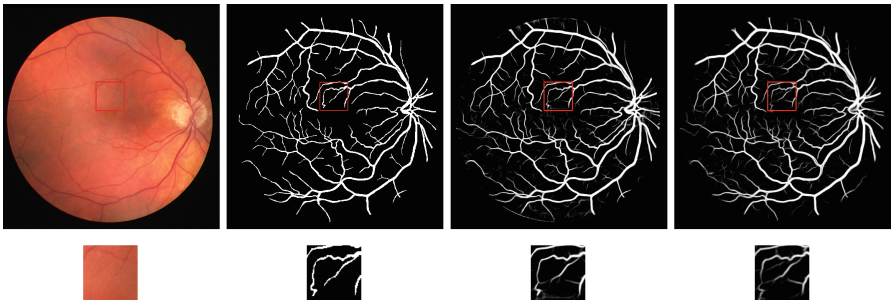


Fig. 4. A test image from the DRIVE dataset (1st column), ground truth (2th column), the segmentation results obtained by using our DEU-Net (3nd column) and Laddernet [12] (4rd column).

Table 1. Results of DEU-Net and other methods on DRIVE datasets.

Methods	Year	F1-score	SE	SP	AC	AUC
Li et al. [4]	2016	N.A	0.7569	**0.9816**	0.9527	0.9738
Orlando et al. [6]	2017	0.7857	0.7897	0.9684	N.A	0.9507
R2U-Net [1]	2018	0.8171	0.7792	0.9813	0.9556	0.9784
Recurrent UNet [1]	2018	0.8155	0.7751	**0.9816**	0.9556	0.9782
LadderNet [12]	2018	0.8202	0.7856	0.9810	0.9561	**0.9793**
Dual Encoding U-Net	2019	**0.8270**	**0.7940**	**0.9816**	**0.9567**	0.9772

Table 2. Results of DEU-Net and other methods on CHASE-DB1 datasets.

Methods	Year	F1-score	SE	SP	AC	AUC
Li et al. [4]	2016	N.A	0.7507	0.9793	0.9581	0.9716
Orlando et al. [6]	2017	0.7332	0.7277	0.9712	N.A	0.9524
R2U-Net [1]	2018	0.7928	0.7756	0.9712	0.9634	0.9815
Residual U-Net [1]	2018	0.7800	0.7726	0.9820	0.9553	0.9779
LadderNet [12]	2018	0.8031	0.7978	0.9818	0.9656	**0.9839**
Dual Encoding U-Net	2019	**0.8037**	**0.8074**	**0.9821**	**0.9661**	0.9812

We also analyzed the speed of our algorithm. It took more than 13 hours to train the DEU-Net on the DRIVE dataset and more than 20 hours on the CHASEDB1 dataset (Intel Xeon CPU E5-2650 v4 CPU, NVIDIA GTX 1080Ti GPU, 125 GB Memory, and pytorch 0.4.0). However, it only took 6.75 s to segment a 584 $\times$ 565 retinal image on DRIVE and 12.82 seconds to segment a 999 $\times$ 960 retinal image on CHASEDB1 dataset. The speed is faster than MS-NFN [11], thus, our proposed network has low computational complexity.

5 Conclusions

Dual Encoding U-Net is proposed in this paper to improve the accuracy of retinal vessel segmentation. Our proposed DEU-Net contains two encoders: a spatial path and a context path which is designed to preserve spatial information and capture semantic information, separately. Besides, channel attention mechanism is applied to select necessary feature map in skip connection. All the designs improve the interpretability of U-Net. Our results indicate that the proposed method has significantly outperforms the-state-of-the-art for retinal blood vessel segmentation on both datasest.

Acknowledgements. This work was supported in part by The National Key Research and Development Program of China (2017YFB1302704) and the Chinese Academy of Sciences (CAS) Scientific Equipment Development Project under Grant YJKYYQ20170050, the Beijing Municipal Science and Technology Commission under Grant Z181100008918010, Youth Innovation Promotion Association CAS and Strategic Priority Research Program of CAS.

References

1. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)
2. Bankhead, P., Scholfield, C.N., McGeown, J.G., Curtis, T.M.: Fast retinal vessel detection and measurement using wavelets and edge location refinement. PloS One **7**(3), e32435 (2012)
3. Kanski, J.J., Bowling, B.: Clinical Ophthalmology: a Systematic Approach. Elsevier, Amsterdam (2011)
4. Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., Wang, T.: A cross-modality learning approach for vessel segmentation in retinal images. IEEE Trans. Med. Imaging **35**(1), 109–118 (2016)
5. Lupascu, C.A., Tegolo, D., Trucco, E.: FABC: retinal vessel segmentation using adaboost. IEEE Trans. Inf. Technol. Biomed. **14**(5), 1267–1274 (2010)
6. Orlando, J.I., Prokofyeva, E., Blaschko, M.B.: A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. IEEE Trans. Biomed. Eng. **64**(1), 16–27 (2017)
7. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters-improve semantic segmentation by global convolutional network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361 (2017)
8. 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). https://doi.org/10.1007/978-3-319-24574-4_28
9. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)
10. Soares, J.V., Leandro, J.J., Cesar, R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification. IEEE Trans. Med. Imaging **25**(9), 1214–1222 (2006)

11. Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W.: Multiscale network followed network model for retinal vessel segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 119–126. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_14
12. Zhuang, J.: Laddernet: multi-path networks based on u-net for medical image segmentation. arXiv preprint arXiv:1810.07810 (2018)