PATCH-BASED FULLY CONVOLUTIONAL NEURAL NETWORK WITH SKIP CONNECTIONS FOR RETINAL BLOOD VESSEL SEGMENTATION

Zhongwei Feng, Jie Yang, Lixiu Yao

Institute of Image Processing and Pattern Recognition,
Shanghai Jiao Tong University,
Shanghai, China 200240
{fengzhongwei}{jieyang}{lxyao}@sjtu.edu.cn

ABSTRACT

Automated segmentation of retinal blood vessels plays an important role in the computer aided diagnosis of retinal diseases. The paper presents a new formulation of patch-based fully Convolutional Neural Networks (CNNs) that allows accurate segmentation of the retinal blood vessels. A major modification in this retinal blood vessel segmentation task is to improve and speed-up the patch-based fully CNN training by local entropy sampling and a skip CNN architecture with class-balancing loss. The proposed method is experimented on DRIVE dataset and achieves strong performance and significantly outperforms the-state-of-the-art for retinal blood vessel segmentation with 78.11% sensitivity, 98.39% specificity, 95.60% accuracy, 87.36% precision and 97.92% AUC score respectively.

Index Terms— Computer Aided Diagnosis, Convolutional Neural Networks, Retinal Blood Vessel Segmentation, Local Entropy Sampling, Class-balancing Loss

1. INTRODUCTION

Segmentation of blood vessels plays an important role in the diagnosis of ophthalmological diseases such as diabetes and hypertension. However, manual segmentation of retina blood vessels by ophthalmologist is both time consuming and lack accuracy. Consequently, automated segmentation systems can play an important part in increasing the segmentation accuracy as well as diagnosis accuracy.

Previous attempts for addressing the task of blood vessel segmentation through vessel enhancement techniques [1] and hand crafted filters like line detectors [2]. In [3], the authors use a combination of co-linearly aligned difference-of-Gaussian filters to detect vessels. Approaches that rely on efficient machine learning methods have raised over the last years. In [4], vessels are detected by utilizing different kinds of features with fully connected conditional random fields. The method presented in [5] augments the unsupervised line detection technique by training an SVM classifier on manually designed features.

Recently, Deep Learning has gained a lot of interest due to its highly discriminative representations, which outperforms many the-state-of-the-art techniques in computer vision. Obviously, it has also attracted medical imaging research field. In the domain of retinal image understanding, In [6] CNNs have been used for retinal vessel segmentation to classify patch features into different vessel classes. In 2016, Liskowski et al. [7] proposed a CNN architecture for vessel segmentation in retina images. Lahiri et al. [8] proposed an architecture that based on an ensemble of stacked denoising autoencoders and the final decision combines all autoencoders outputs. Maji et al. [9] proposed an ensemble of 12 deep CNNs and take the mean of the outputs of all networks as the final output. Deep Retinal Image Understanding [10] (DRIU) uses a base network architecture on which two set of specialized layers are trained to solve both the retinal vessel and optic disc segmentation.

In this paper, we propose a patch-based fully CNN architecture for blood vessel segmentation. The rest of the paper is organized as follows: Section 2 describes the proposed methodology in detail. In Section 3 shows the experimental results on publicly available DRIVE [11] dataset. Finally, in Section 4 we conclude our paper with a summary.

2. MATERIALS AND METHODS

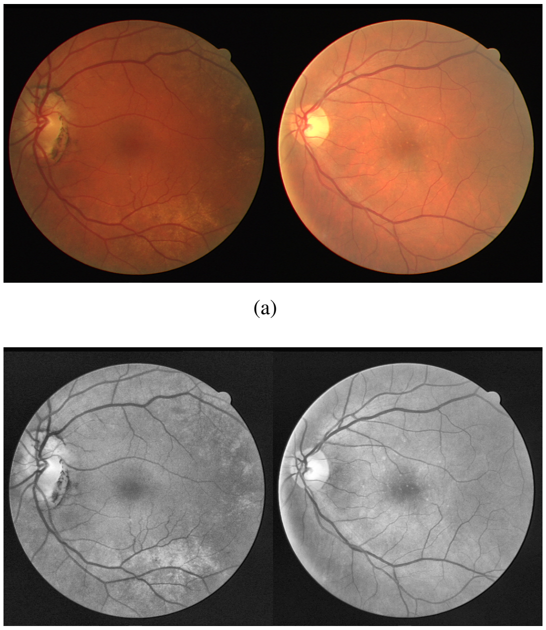
2.1. Dataset Description

In this paper, DRIVE dataset is used to assess the performance of the proposed methodology. It contians 40 images of size 565*584 with two manual segmentations for each image and the first expert observer segmentation is used as the gold standard (ground truth). The dataset is already divided into two parts for training and testing.

2.2. Preprocessing

Retianl images usually comprise noise and uneven illumination, thus image enhancement is necessary before postprocessing. Given a RGB fundus image, I, we convert it to gray

978-1-5090-2175-8/17/$31.00 ©2017 IEEE



(a)

(b)

Fig. 1. (a) Original RGB images, (b) Preprocessed images.

image, I_{gray}, by using the following formula:

$$I_{gray} = 0.299 \times I_r + 0.587 \times I_g + 0.114 \times I_b \tag{1}$$

where I_r, I_g and I_b denote the red, green and blue channel of the fundus image.

Then, normalizing the images by using the following formula:

$$I_{norm} = \frac{I_{gray} - \mu}{\sigma} \tag{2}$$

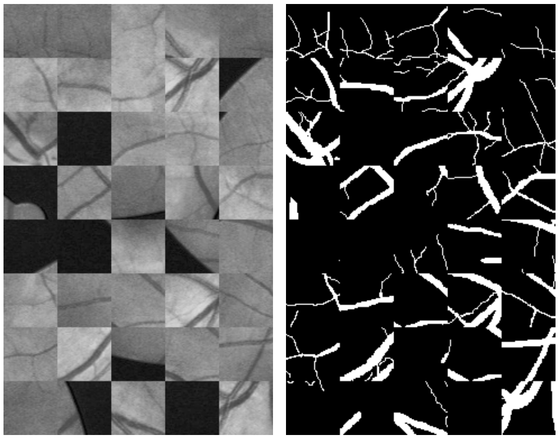
where μ and σ denote the mean and standard deviation of the data.

After 3×3 median filter is applied on normalized images to reduce noise followed by Contrast limited adaptive histogram equalization. Finally, the intensity values are scaled to [0, 1] to get the preprocessed image. Fig. 1 shows some preprocessed images along with the original image from DRIVE dataset.

2.3. Local Entropy Sampling

As for retinal blood vessel segmentation tasks there is very little training images available. However, the training data in terms of patches is much larger than the number of training images. In this paper, patch training is used to segment vessels.

In retinal images, approximately 90% of the pixels belong to the background class, with the remaining 10% of pixels belong to the foreground class. So, another important factor in patch selection is to make sure the retina blood vessels contained in the patches. Otherwise, the net will be overwhelmed with background images and fail to classify any of the minority classes.



(a) (b)

Fig. 2. (a) Examples of training patch, (b) Corresponding patch in ground truth.

A method we proposed to fix this involves selecting a certain subset of the training data from the highest entropy patches in the preprocessed image. Entropy is defined as follow:

$$H = -\sum_{i=0}^{255} p_i \log p_i \tag{3}$$

where p$_i$ is the probability that intensity value equal i.

High entropy patches have much context represented in them, so the model will have more precise vessels to learn from. A sliding window method is applied on the preprocessed images to obtain corresponding entropy images. In this paper, vessel width is an important consideration for patch selection, which is a bit larger than vessel. Here we select 48×48 as patch size. For each retinal image from the training set, we sample 10000 training patches totally, which consists of 5000 training patches from highest entropy area and the other patches at random. Fig. 2 shows some patches along with the ground truth from DRIVE dataset.

2.4. CNN Architecture

In the last five years, deep CNNs have outperformed the-state-of-the-art in many visual tasks. The basic architecture of CNN consists with a number of convolutional layers which are followed by a activiation layer and a pooling layer. Finally, a fully-connected layer is applied on the last layer to map the feature maps to the desired number of classes. As for the classification task, CNNs prominently performs the-state-of-the-art. Furthermore, CNNs have been applied successfully to a large variety of general recognition tasks such as object detection [12], semantic segmentation [13], and contour detection [14].

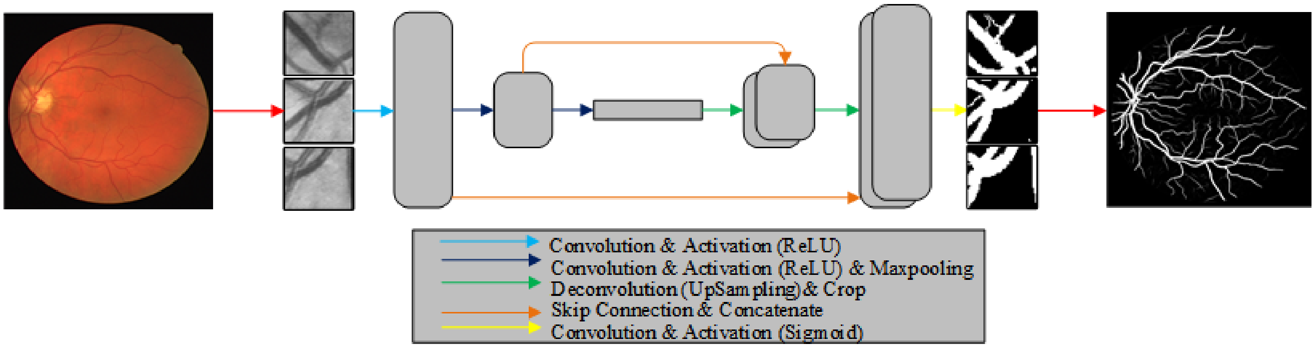

Convolution & Activation (ReLU)
Convolution & Activation (ReLU) & Maxpooling
Deconvolution (UpSampling) & Crop
Skip Connection & Concatenate
Convolution & Activation (Sigmoid)

Fig. 3. Proposed CNN architecture for vessel segmentation. Each box corresponds to a multi-channel feature map.

Inspired by the fully convolutional networks [13], which show good performance for nature image segmentation, a skip CNN architecture is illustrated in Fig. 3. Its input to the first convolutional layer in the proposed architecture is a $1 \times 48 \times 48$ patch from the training set. The main path follows the typical architecture of a convolutional network. It consists of the repeated application of 3×3 convolutions with padding for same size, each followed by a rectified linear unit and a 2×2 max pooling operation with 2 pixels stride for downsampling to reduce the amount of parameters and computation. At each downsampling step we double the number of feature channels. In order to concate different feature maps through the skip connection, an upsampling of the feature map followed by a 2×2 deconvolution [13] is used. This important modification in our architecture is that allow the network to propagate context information to higher resolution layers for more precise vessels segmentation. After a concatenation, each followed by a rectified linear unit. The dropout [15] layer with probability 0.2 is a regularization technique for reducing overfitting in neural networks by preventing complex co-adaptations on training data. At the final convolutional layer a 1×1 convolution with no padding is used to map multi-channel feature maps to the desired number of classes. Finally, a sigmoid operation is applied on the last convolutional layer to scale the output segmentation map. The output is of dimension $1 \times 48 \times 48$. In totally, the network has 11 convolutional layers.

For training the network, the vessel segmentation task is learnt by class-balancing loss function originally proposed in [16] for the task of contour detection in natural images. The class-balancing loss function is then defined as:

$$\begin{aligned} L(y_i, \hat{y}_i) = & -\beta \sum_{i \in Y_+} (y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)) \\ & - (1 - \beta) \sum_{i \in Y_-} (y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)) \\ = & -\beta \sum_{i \in Y_+} \log \hat{y}_i \\ & - (1 - \beta) \sum_{i \in Y_-} \log(1 - \hat{y}_i) \end{aligned} \tag{4}$$

where $y_i \in \{0, 1\}$ is the ground truth while $\hat{y}_i$ is the predicted segmentation map, which obtained from the last activation layer. The multiplier β is used to achieve the balance of the large number of background compared to foreground pixels, which are the vessels. Y_+ and Y_- denote the foreground and background sets of the ground truth Y, respectively. In this case, we use $1 - \beta = |Y_+|/|Y|$, $\beta = |Y_-|/|Y|$.

3. EXPERIMENT AND DISCUSSIONS

3.1. Training Parameters

Training consists in an iterative propagation of patches through the network and modification of its weights, which are initialized by xavier method. The cycle of presenting all training examples, an epoch, is split into smaller units called batches. In this approach, we use mini-batch stochastic gradient descent with momentum at 0.3 and fixed the learning rate to be 10^{-4}. The model is trained for 100 epochs with a batch size of 256. The implementation is based on the Caffe framework [17], which performs all computation on GPUs in single precision arithmetic. The experiments are conducted on Intel Core i7-4770K CPU with a NVIDIA Tesla K40c card.

3.2. Results

At testing phase, in order to achieve balance between accuracy and computation, we orderly extract patches with 5 pixels stride from the testing set. Then, for each pixel, the vessel probability is obtained by averaging probabilities over all the predicted patches covering the pixel. Fig. 4 shows exemplary segmentations produced by our proposed method.

In Table 1, we present the performance of networks tested on the testset in terms of area under ROC curve AUC, accuracy Acc, sensitivity Sen, specificity Spe and precision Pre, defined as:

$$\begin{aligned} Acc = & \frac{TP + TN}{TP + TN + FP + FN}, Sen = \frac{TP}{TP + FN} \\ Spe = & \frac{TN}{TN + FP}, Pre = \frac{TP}{TP + FP} \end{aligned} \tag{5}$$

Method	Sen	Spe	Acc	Pre	AUC
Fraz et al. [1]	.7302	.9742	.9422	.8112	-
Azzopardi et al. [3]	.7655	.9704	-	-	-
Orlando et al. [4]	**.7897**	.9684	-	.7854	-
Liskowski et al. [7]	.7819	.9748	.9472	-	.9719
Lahiri et al. [8]	.7500	.9800	.9480	-	-
proposed method	.7811	**.9839**	**.9560**	**.8736**	**.9792**

Table 1. Comparison of proposed method on the DRIVE dataset with other state-of-the-art methods.

where TP, FP, FP and FN are the numbers of true positive, false positive, true negative and false negative respectively. In this case, positive decision is made when the output of the unit associated with the positive class in the sigmoid output layer is greater than the 0.5; otherwise, negative decision is made. In fact, this threshold can be arbitrarily selected from [0, 1], which leads to different result.

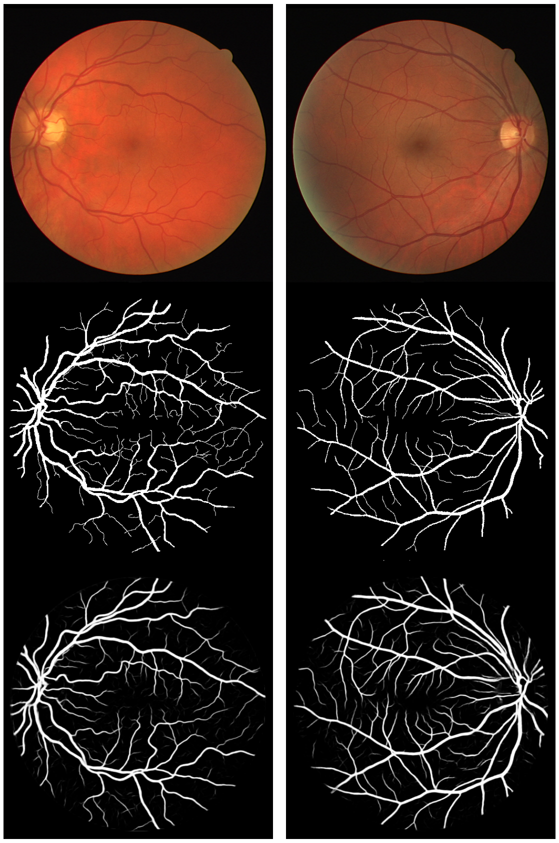


Fig. 4. Output map produced by our proposed method on the DRIVE dataset:top row, original retinal images; middle row, expert annotations; bottom row, results obtained by our method

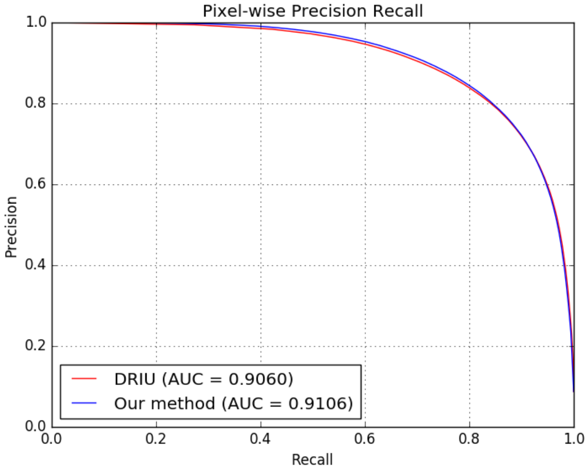


Fig. 5. Pixel-wise precision recall curves: the blue curve represents our method, another is DRIU.

Through binarizing the segmentation map at multiple confidence values and the pixel-wise precision-recall is computed between the obtained mask and the ground truth. Fig. 5 shows comparison of proposed method on the DRIVE dataset with the other state-of-the-art method:DRIU [10]. As for F1 score, the proposed system equals 0.8183 compared with 0.7382 in DRIU. All results of DRIU are obtained with pre-computed DRIU segmentations.

4. CONCLUSION

Combining with local entropy sampling, we presented a patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation. The proposed architecture has high-capability to learn hierarchical features and context information from raw pixel data without any prior domain knowledge.

In this paper, the proposed method has strong performance and significantly outperforms the-state-of-the-art for retinal blood vessel segmentation on the DRIVE dataset. Taking advantage of deep learning, our method has full potential of carrying out more robust than traditional methods based on statistical analyses. So, we currently work on verifying this claim for exudates segmentation on retinal images.

5. ACKNOWLEDGEMENT

This research is partly supported by NSFC, China (No: 81600776) and Committee of Science and Technology, Shanghai, China (No: 16411962100).

6. REFERENCES

[1] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R Rudnicka, Christopher G Owen, and Sarah A Barman, “Blood vessel segmentation methodologies in retinal images–a survey,” *Computer methods and programs in biomedicine*, vol. 108, no. 1, pp. 407–433, 2012.

[2] Uyen TV Nguyen, Alauddin Bhuiyan, Laurence AF Park, and Kotagiri Ramamohanarao, “An effective retinal blood vessel segmentation method using multi-scale line detection,” *Pattern recognition*, vol. 46, no. 3, pp. 703–715, 2013.

[3] George Azzopardi, Nicola Strisciuglio, Mario Vento, and Nicolai Petkov, “Trainable cosfire filters for vessel delineation with application to retinal images,” *Medical image analysis*, vol. 19, no. 1, pp. 46–57, 2015.

[4] José Ignacio Orlando and Matthew Blaschko, “Learning fully-connected crfs for blood vessel segmentation in retinal images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 634–641.

[5] Elisa Ricci and Renzo Perfetti, “Retinal blood vessel segmentation using line operators and support vector classification,” *IEEE transactions on medical imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.

[6] Yaroslav Ganin and Victor Lempitsky, “N4-fields: Neural network nearest neighbor fields for image transforms,” vol. 9004, pp. 536–551, 2014.

[7] P Liskowski and K Krawiec, “Segmenting retinal blood vessels with deep neural networks.,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1–1, 2016.

[8] Avisek Lahiri, Abhijit Guha Roy, Debdoot Sheet, and Prabir Kumar Biswas, “Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography,” in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 1340–1343.

[9] Debapriya Maji, Anirban Santara, Pabitra Mitra, and Debdoot Sheet, “Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images,” *arXiv preprint arXiv:1603.04833*, 2016.

[10] K.K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, “Deep retinal image understanding,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016.

[11] Joes Staal, Michael D Abrmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[12] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, “Region-based convolutional networks for accurate object detection and segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 142–158, 2016.

[13] Jonathan Long, Evan Shelhamer, and Trevor Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[14] Jimei Yang, Brian Price, Scott Cohen, Honglak Lee, and Ming-Hsuan Yang, “Object contour detection with a fully convolutional encoder-decoder network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 193–202.

[15] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting.,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[16] Saining ”Xie and Zhuowen” Tu, “Holistically-nested edge detection,” in *Proceedings of IEEE International Conference on Computer Vision*, 2015.

[17] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, “Caffe: Convolutional architecture for fast feature embedding,” *arXiv preprint arXiv:1408.5093*, 2014.