RETINAL VESSEL SEGMENTATION VIA DEEP LEARNING NETWORK AND FULLY-CONNECTED CONDITIONAL RANDOM FIELDS

Huazhu Fu, Yanwu Xu, Damon Wing Kee Wong, Jiang Liu

Ocular Imaging Department, Institute for Infocomm Research
Agency for Science, Technology and Research (A*STAR), Singapore.
{fuhz, yaxu, wkwong, jliu}@i2r.a-star.edu.sg

ABSTRACT

Vessel segmentation is a key step for various medical applications. This paper introduces the deep learning architecture to improve the performance of retinal vessel segmentation. Deep learning architecture has been demonstrated having the powerful ability in automatically learning the rich hierarchical representations. In this paper, we formulate the vessel segmentation to a boundary detection problem, and utilize the fully convolutional neural networks (CNNs) to generate a vessel probability map. Our vessel probability map distinguishes the vessels and background in the inadequate contrast region, and has robustness to the pathological regions in the fundus image. Moreover, a fully-connected Conditional Random Fields (CRFs) is also employed to combine the discriminative vessel probability map and long-range interactions between pixels. Finally, a binary vessel segmentation result is obtained by our method. We show that our proposed method achieve a state-of-the-art vessel segmentation performance on the DRIVE and STARE datasets.

Index Terms— Vessel segmentation, Conditional Random Fields, Convolutional Neural Networks.

1. INTRODUCTION

Retinal vessel has diagnostic significance, it is widely used in monitoring the disease progression, and evaluation of various ophthalmologic diseases [1]. However, manually vessel segmentation by trained specialists delineate is a repetitious and time-consuming task. In the last two decades, many approaches have been introduced to segment the retinal vessel automatically. For example, Marin *et al.* [2] extracted a 7-D vector composed of gray-level and moment invariants-based features for pixel representation, and classified each pixel by using a neural network scheme. Nguyen *et al.* [3] utilized a multi-scale line detection to produce the final vessel segmentation. Orlando *et al.* [4] proposed an automatic method for vessel segmentation based on fully-connected Conditional Random Fields (CRFs), which learned its configuration by using a structured output SVM. However, most of the existing methods are based on manual designed features, which lack the efficient discriminative and are easily affected by the pathological regions in the fundus images.

Deep learning (DL) architecture is formed by the composition of multiple linear and non-linear transformations of the data, with the goal of yielding more abstract and ultimately more useful representations, which has been demonstrated having the powerful ability in many computer vision tasks (e.g., image classification and object detection). Convolutional Neural Networks (CNNs) are DL architectures, which are not only improving for whole-image classification [5], but also making progress on local tasks with structured output [6]. Long *et al.* [7] showed that a fully CNNs trained end-to-end, pixels-to-pixels on semantic segmentation exceeded the most state-of-the-art methods without further machinery. Based on fully CNNs, Xie *et al.* [8] further proposed a holistically-nested edge detection (HED) method to automatically learn the rich hierarchical representations with deep supervision, and resolved the challenging ambiguity in edge and object boundary detection. In this paper, we formulate the retinal vessel segmentation to a boundary detection problem, and utilize the fully CNNs architecture to learn the discriminative features that better characterize the important hidden patterns related to vessels and backgrounds in the fundus images.

A similar DL-based vessel segmentation work is proposed by Melinscak*et al.* [9], which addressed the retinal vessel segmentation into a pixel-level binary classification task. They used a deep neural network [5] to classify each pixel in the fundus image. However, it has two disadvantages: first, the classification for each individual pixel is less the global smoothness correlation, which makes the method failed for some local pathological regions; second, the pixel-to-pixel classification strategy significantly spends plenty of running times on both training and testing phases. By contrast, our method is based on fully CNNs architecture [7, 8], which is an image-to-image training system and provides a multi-scale and multi-level visual responses. A vessel probability map is generated as the output of our fully CNNs architecture. And then a fully-connected CRFs is utilized further to take into account the global pixel correlation and output a binary vessel segmentation results.

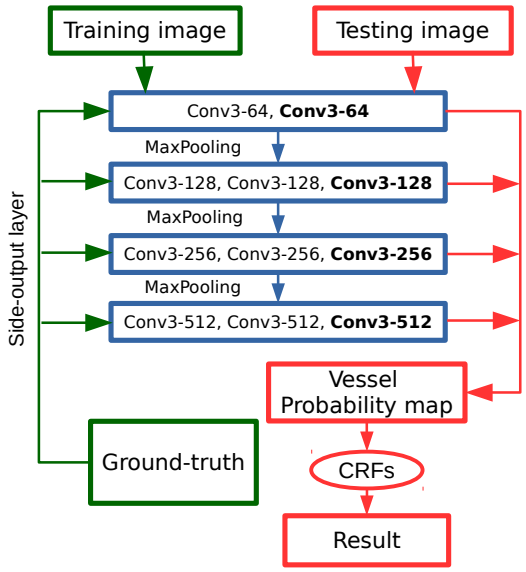

Fig. 1. The architecture of our retinal vessel segmentation method, which contains one fully CNNs network and one fully-connected CRFs. The fully CNNs network is a four-stages HED-like architecture, where side-output layers are inserted after the last convolutional layers in each stage (marked in **Bold**). The convolutional layer parameters are denoted as "Conv<receptive field size>-<number of channels>". The ReLU activation function is not shown for brevity.

2. PROPOSED METHOD

Fig. 1 give the illustration of our architecture for retinal vessel segmentation. Our proposed method includes two parts: one fully CNNs network used to learn the discriminative features and generate the vessel probability map; one fully-connected CRFs used to produce the binary vessel segmentation result with the dense global pixel correlation.

2.1. Vessel probability map via deep learning networks

Different from the existing DL-based method [9], which employed a pixel-to-pixel classification, our method treats vessel segmentation as a counter detection problem. To capture the inherent scales of vessel maps, we build our architecture on the top of HED system [8], which is based on the ideas of fully fully CNNs [7] and deeply-supervised net [10].

Fully convolutional networks: Suppose $\mathbf{x}_{ij}$ is the data vector at location (i, j) in a particular layer, and $\mathbf{y}_{ij}$ is for the following layer. The function of output $\mathbf{y}_{ij}$ is defined as:

$$\mathbf{y}_{ij} = f_{ks}(\{\mathbf{x}_{si+\delta i, sj+\delta j}\}_{0 \le \delta i, \delta j \le k}), \quad (1)$$

where k and s are the kernel size and the stride factor, respectively. f_{ks} denotes the layer manipulation (e.g., convolution, pooling, or activation function). This functional form is maintained under composition, with kernel size and stride obeying

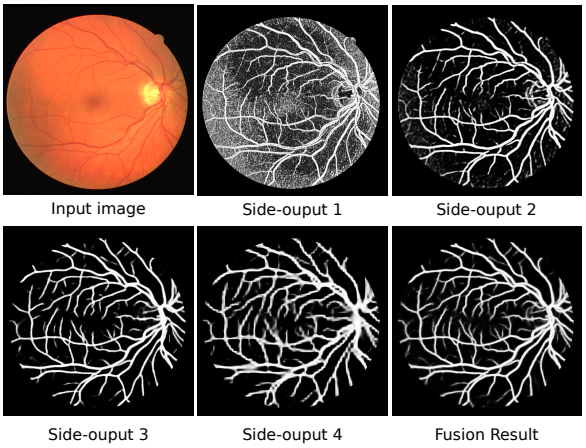

Fig. 2. The map for each side-output in our architecture.

the transformation rule:

$$f_{ks} \circ g_{k's'} = (f \circ g)_{k'+(k-1)s', ss'}. \quad (2)$$

A fully CNNs is that the whole networks are composed by this form layers to compute the nonlinear filter. A major advantage is that the fully CNNs can take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning.

Our architecture: The original HED has five stages. Each stage includes multiple convolutional and ReLU layers, and the side-output layers are connected to the last convolutional layer in each stage to suppose the deep layer supervision. In our architecture, we employ four stages to generate vessel probability map, as shown in Fig. 1. The main reason is that the retinal vessels in fundus images are different from the general object boundaries in natural images. The object boundary separates two regions with different appearances, which makes the boundary available in the small plane outputted from deeper layer. By contrast, the retinal vessel is a line-shape object, which is too-thin to output a meaningful response in the higher stride layer. Thus, we only employ a four-stage architecture. The generated map for each side-output in our architecture is shown in Fig. 2.

Implementation details: We implement our framework using the *Caffe* library[11] and build on top of implementation of HED [8]. The model parameters follow the configuration used in [8]. For the fine-tuning data, we employ the ARIA dataset [12], which is the largest vessel segmentation dataset containing 143 fundus images with 768×576 resolution. We a) rotate the images to 4 different angles and crop the largest rectangle in the rotated image; b) flip the image at each angle. Moreover, due to the fully convolutional network can take input of arbitrary size, we also add the dataset by scaling the images to 50% and 75% of its original size. The fine-tuning phase takes about two days on a single NVIDIA K20 GPU ($10,000$ iterations). For a 565×584 image, it takes about 200 ms to generate a vessel probability map.

2.2. Fully-connected CRFs segmentation:

Although the fully CNNs networks could produce the satisfactory vessel probability maps, they still have some problems. Firstly, traditional CNNs have convolutional filters with large receptive fields and hence produce coarse maps when pixel-level vessel segmentation (e.g., non-sharp boundaries and blob-like shapes). Secondly, fully CNNs lack smoothness constraints that may result in small spurious regions in segmentation output. Thus, we introduce a CRFs to obtain the binary vessel segmentation result. CRFs models pixel labels as random variables that form a Markov Random Field (MRF) when conditioned upon a global correlation. In the fully-connected CRFs, each node is a neighbor of every other [13]. Following this approach, the method is able to take into account long-range interactions in the whole image. We denote $\mathbf{v} = \{v_i\}$ as a labeling over all pixels of the image. The energy of a label assignment $\mathbf{v}$ is given by:

$$E(\mathbf{v}) = \sum_i \psi_u(v_i) + \sum_{i<j} \psi_p(v_i, v_j), \qquad (3)$$

where the unary energy components $\psi_u(v_i)$ measure the cost of the pixel i taking the label v_i, which is obtained by the vessel probability map P in our method as $\psi_u(v_i) = P(i)$. The pairwise energy components $\psi_p(v_i, v_j)$ measure the cost of assigning labels v_i, v_j to pixels i, j, and provide the data-dependent smoothing term. It is defined as:

$$\psi_p(v_i, v_j) = \mu(v_i, v_j) \sum_{(m=1)}^{M} w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j), \qquad (4)$$

where each $k^{(m)}$ is a Gaussian kernel applied on feature vectors. The feature vector of pixel i, denoted by $\mathbf{f}_i$, is derived from image features such as spatial location and RGB values. The mean-field approximation [13] to the CRFs distribution is used for approximate maximum posterior marginal inference.

3. EXPERIMENTS

We evaluate our method using two publicly available DRIVE [14] and STARE [15] datasets. These two datasets provide two manual segmentations generated by two different experts for each image. The selection of first observer is accepted as ground truth and used for performance evaluation in literature. We performed the evaluation in terms of Accuracy (Acc) and Sensitivity (Sen).

We compare our method with several state-of-the-art vessel segmentation methods, and also report the ground truth labeling of the second observer as the human performance. Table 1 lists the performance scores on the DRIVE and STARE datasets. Our method obtains the best Acc score than other methods, even including the other DL method [9]. Note that our method obtains a similar Acc performance to Human observer on DRIVE dataset and a better Acc on STARE dataset.

Table 1. Performances on the two datasets.

	DRIVE dataset		STARE dataset	
Methods	Acc	Sen	Acc	Sen
Martinez-Perez [16]	0.9344	0.7246	0.9410	0.7506
Nguyen [3]	0.9407	0.7429	0.9326	0.8014
You [17]	0.9434	0.7410	0.9497	0.7260
Orlando [4]	0.9437	0.7851	-	-
Staal [14]	0.9443	0.7345	0.9516	0.6970
Marin [2]	0.9451	0.7067	0.9526	0.6944
Mendonca [18]	0.9452	0.7344	0.9440	0.6996
Melinscak [9]	0.9466	0.7276	-	-
Side-output 1	0.8602	0.7698	0.9141	0.4633
Side-output 2	0.9370	0.7194	0.9397	0.7189
Side-output 3	0.9368	0.6965	0.9429	0.8143
Side-output 4	0.9109	0.6241	0.9261	0.7630
Our fusion	0.9470	0.7294	0.9545	0.7140
Human observer	0.9472	0.7761	0.9352	0.8951

We also provide the results produced by the individual side-output in Table 1. We observe that the second and third side-output obtain the better performances than other two layers. Fig. 3 shows some segmentation examples. It can be observed that in the optic disc boundary and pathological region (marked by red rectangle in the third column), there are many false positive pixels produced by the existing method (e.g., [4]). By contrast, our method deals with these challenge cases by using learned discriminative features.

4. CONCLUSION

In this paper, we have developed a retinal vessel segmentation based on deep learning architecture and fully-connected CRFs. The discriminative features learned by fully CNNs architecture deal with the challenging pathological regions in fundus images. A high quality vessel probability map is produced to obtain a binary segmentation result by using fully-connected CRFs segmentation. We have demonstrated that our system achieves a new state-of-the-art on DRIVE and STARE datasets. In our experiments, the training data and testing data come from the different datasets due to the limitation of dataset size, which may lead a cross-learning problem. How to deal with this cross-learning will be a direction for future work. Moreover, the variation of different deep learning architectures should be developed to further improve the vessel segmentation in the fundus images.

5. REFERENCES

[1] M. Abramoff, M. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[2] D. Marin, A. Aquino, M. Gegundez-Arias, and J. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment

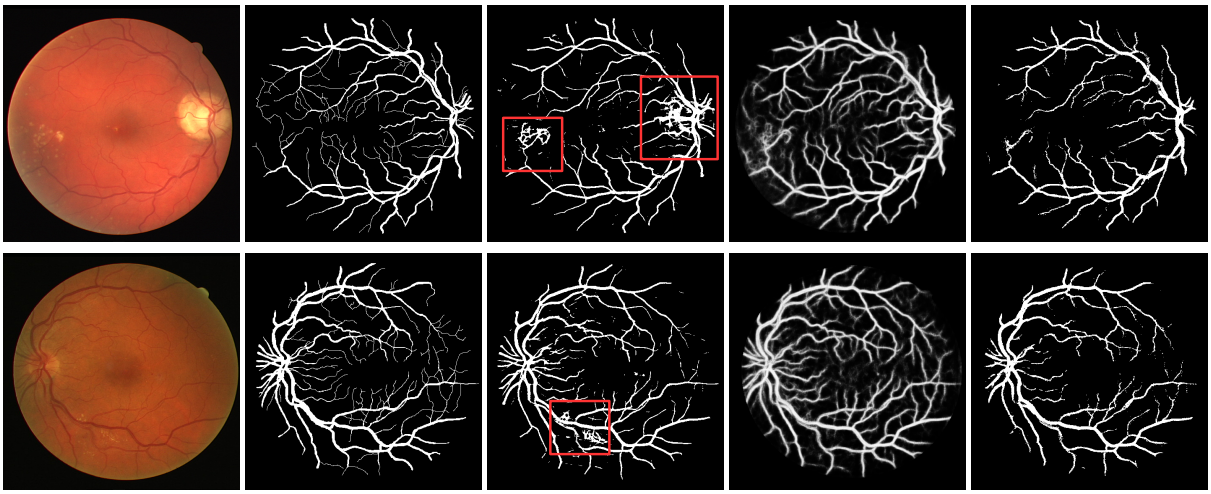

Fig. 3. Examples of results from the dataset. From left to right: original fundus images, ground truth, the result of [4], our vessel probability map and final vessel segmentation result. Note that the state-of-the-art vessel segmentation method (e.g., [4]) is limited by optic disc boundary and pathological region highlighted by the red rectangle.

invariants-based features," *IEEE TMI*, vol. 30, no. 1, pp. 146–158, 2011.

[3] U. Nguyen, A. Bhuiyan, L. Park, and K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection," *PR*, vol. 46, no. 3, pp. 703–715, 2013.

[4] J. Orlando and M. Blaschko, "Learning fully-connected crfs for blood vessel segmentation in retinal images," in *MICCAI*, 2014.

[5] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012.

[6] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs," in *ICLR*, 2015.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015.

[8] S. Xie and Z. Tu, "Holistically-nested edge detection," in *ICCV*, 2015.

[9] M. Melinscak, P. Prentasic, and S. Loncaric, "Retinal vessel segmentation using deep neural networks," in *VISAPP*, 2015.

[10] C. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *AISTATS*, 2015.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding," *ACM MM*, 2014.

[12] D. Farnell, F. Hatfield, P. Knox, M. Reakes, S. Spencer, D. Parry, and S.PHarding, "Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators," *Journal of the Franklin Institute*, vol. 345, no. 7, pp. 748–765, 2008.

[13] P Krähenbühl and Vladlen Koltun, "Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials," in *NIPS*, 2011.

[14] J. Staal, M. Abràmoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE TMI*, vol. 23, no. 4, pp. 501–509, 2004.

[15] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE TMI*, vol. 19, no. 3, pp. 203–210, 2000.

[16] M. Martinez-Perez, A. Hughes, S. Thom, A. Bharath, and K. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," *Medical Image Analysis*, vol. 11, no. 1, pp. 47–61, 2007.

[17] X. You, Q. Peng, Y. Yuan, Y. Cheung, and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," *PR*, vol. 44, no. 1011, pp. 2314 – 2324, 2011.

[18] A. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE TMI*, vol. 25, no. 9, pp. 1200–1213, 2006.