Automatic Retinal Vessel Segmentation via Deeply Supervised and Smoothly Regularized Network

Yi Lin, Honggong Zhang, and Guang Hu

***Abstract*—In recent years, retinal vessel segmentation technology has become an important component for disease screening and diagnosing in clinical medicine. However, retinal vessel segmentation is a challenging task due to complex distribution of blood vessels, relatively low contrast between target and background, and potential presence of illumination and pathologies. In this paper, we propose an automatic retinal vessel segmentation network using deep supervision and smoothness regularization, which integrates holistically-nested edge detector (HED) and global smoothness regularization from conditional random fields (CRFs). It is an end-to-end and pixel-to-pixel deep convolutional network, can perform better results than HED-based methods and the methods where CRF inference is applied as a post-processing method. With co-constraints between pixels, the proposed DSSRN obtains better results. Finally, we show that our proposed method obtains a sate-of-the-art vessel segmentation performance on all three benchmarks, DRIVE, STARE and CHASE_DB1.**

***Index Terms*—Vessel segmentation, deep learning, medical image analysis, deep supervision, conditional random field.**

I. Introduction

The delineation of morphological attributes of retinal blood vessels has a certain connection with cardiovascular diseases, such as diabetes and hypertension [1], [2]. Retinal vessel segmentation technology is becoming a fundamental and important component for disease screening and diagnosing in clinical medicine. On the one hand, it helps speed up automatic retinal disease screening [3], releasing the labor of doctors and specialists; on the other hand, retinal vessel segmentation will help to early detect and control related diseases, has good effect of diabetic retinopathy [4], hypertensive retinopathy [5], cardiovascular diseases [6] and arterial narrowing [7].

At present, the existing retinal vessel segmentation algorithms can automatically classify blood vessels from background, but there are still two main challenges: First, the contrast between retinal blood vessels and background in different retinal image databases is different; it is easier to good segmentation results where the region has a striking contrast; otherwise. it is harder. Second, illumination and pathologies have a great influence, as is known, the vessels are dark and background is bright in retinal images, which is the most important factor to make a distinction, but the brightness can be affected by illumination reflection and pathologies in the form of cotton wool spots, bright and dark lesions.

Due to the importance of the retinal vessel segmentation problem, there are many methods have been proposed to address this challenging task [8], [9], [10], [11], [12] by utilizing different image features, such as textual features and wavelet features, and different machine learning methods, such as supervised classifiers and unsupervised matching methods. However, their performance of vessel segmentation is not satisfactory enough.

It is natural to apply deep learning for this vessel segmentation problem. Inspired by other semantic segmentation tasks, most research of using Fully Convolutional Networks (FCNs) have successfully achieve remarkable result. Long et al. [13] showed that a FCN trained end-to-end, pixel-to-pixel map on semantic segmentation, and exceeded the most existing methods. Later, Xie et al. [14] further proposed the Holistically-Nested Edge Detector (HED) model, which is to automatically learn the rich hierarchical representations with deep supervision, and resolved the challenging ambiguity in edge and object boundary detection. Besides, most researchers have given attention to pixel-level labelling in image understanding. The HED model obtains very impressive on the problem of edge detection in natural images, which is getting close human's performance. Since the retinal vessel segmentation problem is similar to the edge detection problem, applying holistically-nested edge detection network for vessel segmentation seems to be a very good solution.

However, thought the vessel segmentation problem and the edge detection problem is similar, there are clear differences. Notices that not all edge pixels in retina images are vessels and not all vessel pixels appear on boundaries. Besides, different from edges which have a width of one pixel, vessel have variant width. Thus, we need to develop new techniques to address these differences. Our observation is that non-maximum suppression is applied for removing duplicated edge pixels and only the skeleton of edge is preserved. But the vessel widths are different, we should have a model to reasoning the contextual information to obtain the true region of vessels. In addition, contextual information is also important to avoid false positives and recall false negatives. To accomplish this task, we recall a classical statistical model which is good modeling contextual information, which the conditional random field (CRF) method [15]. CRF acts as a smoothness regularizer and can use the spatial contextual information in image to obtain better segmentation results.

In the recent research of deep learning based semantic segmentation works, Zheng et al. [16] proposed the CRFasRNN,

Yi Lin and Honggong Zhang are with the School of Information and Communication Engineering,Beijing University of Posts and Telecommunications, Xitucheng Road 10, Beijing 100876 China. Emails: linyi092@bupt.edu.cn and zhhg@bupt.edu.cn

Guang Hu is with the School of Electronic Information and Communications, Huazhong University of Science and TechnologyLuoyu Road 1037, Wuhan,430074 China.

integrated CRF modelling with CNN, trained the whole deep end-to-end network, which is efficient in semantic segmentation task. By taking the advantages in the previous literatures, in this paper, we propose a deep retinal vessel segmentation network based on the holistic-nested edge detection model which provides deep supervision, and combine CRF to provide the global smoothness correlation. The deep network is termed as deeply supervised and smoothly regularized network (DSSRN).

Compared to widely used HED network structure, the proposed DSSRN has the following advantages: 1) shallower side outputs capture rich detailed information; 2) deeper side outputs have better global high level knowledge but fuzzy. So, we combined CRF with HED, built an end-to-end network, put the smooth correlation into the deeply supervised net. With co-constraints between pixels, results from DSSRN method perform better: 1) shallower side contains more detailed information; 2) segmentation result retains more tiny vessels. Our method produces state-of-the-art results on three widely used retinal vessel segmentation datasets.

The details of the proposed DSSRN will be given section III. In the experiments, we have tested three public available datesets, the DRIVE dataset [8], the STARE dataset [17] and the CHASE_DB1 dataset [18]. The retinal images in the three datasets are captured under different conditions. The proposed DSSRN retinal vessel segmentation method can obtain the state-of-the-art performance on all three datasets, which confirms the effectiveness and superiorities of the proposed method. Besides of the advanced the results, the main contribution of the paper is a deeply supervised and smoothly regularized network for retinal vessel segmentation and beyond, as other vessel segmentation tasks will also be beneficial from the proposed network design.

In the rest of the paper, Section II will briefly introduce related works; Section III will give the details of the proposed deeply supervised and smoothly regularized network for retina vessel segmentation; Section IV will gives the details of the experiments, compare the proposed method to other methods and carry out ablation studies; Finally, Section V will concludes the paper.

II. Related Works

Since the last century, many algorithms have been published relating to segment the retinal vessels. For example, Chaudhuri et al. [19] utilized matched filtering for the detection of retinal vessels, and proposed a two-dimensional linear kernel with a Gaussian profile. Staal et al. [8] proposed a ridge based vessel segmentation, the ridge line is regarded as the structure of blood vessels; the image is partitioned into convex set regions by assigning each pixel to the closest line element, then each pixel is classified by a k-NN classifier. Marin et al. [20] proposed a neural network based supervised algorithm, which used a 7-D feature vector composed of gray-level and moment invariant-based features for representation, and classified each pixel by a multi-layer network. Most of the existing algorithms for retinal vessel segmentation are based on manual designed features, which are not efficient to describe or discriminate the target(vessels) and the background. Moreover, manual designed features are not robust to noise and lack of spatial constraint information, cut and burr appears easily in the segmentation results.

Compared with manual designed features [19], [8], [21], [22], [23], CNN-based methods have an impressive record of applications in image analysis and interpretation, including medical imaging. Liskowski et al. [24] proposed a deep learning based method for the problem of detecting blood vessels in fundus imagery, in which they regarded retinal vessel segmentation as a image detection task, assumed that the decision on pixel's class can be made based on its neighborhood, defined as a square window (patch) centered on the pixel to be classified. The supervised segmentation technique used a deep neural network (two stack of convolutional layers and three fully-connected layers) trained on a large sample of examples pre-processed with global contrast normalization, zero-phase whitening, and augmented using geometric transformations and gamma corrections. The CNNs based method significantly outperform the previous algorithms on the accuracy of classification, but it has some disadvantages as follows: first, the method selects size-fixed patch centered on each image pixel, which brings about large storage; second, each neighbouring image patch will be calculated in the network, which spends plenty of training time; third, the method lack of global smoothness correlation, that is, each training patch is individual, there is less global spatial relationship between pixels.

Then, fully convolutional networks (FCN) become more and more popular. The first work was proposed by Long et al. in [13] which reveals the relation between image classification and semantic segmentation and obtains pixel to pixel prediction. Different from the previous deep learning based semantic segmentation methods which rely on a sliding window strategy to classify every region in a highly inefficient way, FCN works very fast and runs in real-time on GPU. The reason is that all pixels in a testing image sharing features only the last simple linear classifier requires individual calculations. Our DSSRN method is based FCN and enjoys the fast testing speed.

Deep supervision does not add much extra computation in training and testing the network. In directviewing, it only adds some shortcut connections between the intermediate feature maps in the deep network and the final ground-truth [14]. However, theoretically, it is very important, since it helps to alleviate the problem of gradient vanishing. Thus, it helps to learn better deep features and produce better semantic segmentation performance. Our work adopt this deep supervision network structure and confirms the effectiveness of deep supervision in the task of retinal vessel segmentation. Besides, we add smoothness term in the work to further boost the performance.

CRF [25] is one of the most applied machine learning models. Xu et. al applied it for semantic segmentation in [15]. In deep convolutional neural networks, it is still useful since FCN predicts the label of pixels individually and pay less attention on modeling the pairwise/contextual information in images. CRF helps and be integrated into FCN as a module. The proposed DSSRN enjoys the contextual modeling power

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2844861, IEEE Access

of CRF. One related work proposed by Fu et. al [26] which also use CRF and FCN for retinal vessel segmentation. However, the proposed method uses CRF to regularize intermediate feature map in the work and optimize them in a end-to-end manner, and our superior results confirms the effectiveness of this design.

In all, we propose the deeply supervised and smoothed regularized network (DSSRN) for retinal vessel segmentation. DSSRN is constructed on fully convolutional network (FCN), which is deeply supervised, the label information is used in all intermediate layers; then we add global smoothness correlation into the net via conditional random field (CRF) to build an end-to-end and pixel-to-pixel segmentation network.

III. Approach

To overcome the aforementioned problem, we propose a deep learning based vessel segmentation method, named as deeply supervised and smoothly regularized network (DSSRN), which is efficiently discriminative and has global spatial constraint information to the whole network. Specifically, DSSRN is an end-to-end network for pixel-to-pixel image segmentation network, which combines the strengths of both fully convolutional network (FCN) and conditional random field (CRF). The FCN is deeply supervised as the label information is used in all intermediate layers, and the CRF introduces global smoothness regularization into the network.

DSSRN: The proposed method is built on holistically-nested edge detection (HED), which is based on fully CNNs and deeply-supervised net; then combined with conditional random field (CRF) to construct an end-to-end segmentation system, the idea of CRF layer is based on CRF-RNN, which formulates mean-field approximate inference for the dense CRF with Gaussian pairwise potentials as a recurrent neural network (RNN). DSSRN is an end-to-end net, when properly trained, the network should outperform a system where CRF inference is applied as a post-processing method [26], such as the method proposed by Fu et al. Fig. III gives the illustration of our architecture (DSSRN) for retinal vessel segmentation. The following subsections are dedicated to a detailed description of the proposed approach.

A. Deeply Supervised Network

We first start out with the HED architecture, which has five blocks, and each block includes multiple convolutional and ReLU layers. Most importantly, the side-output from each intermediate layer will be connected to the last convolutional layer, thus the network is deeply supervised. HED is designed for resolving the challenging ambiguity in edge and object boundary detection, the trained images are mostly natural and scene images, usually are multi-scale and multi-object. Unlike retinal images, the composition is simpler, one pixel is either vessel target or background. The blood vessel is a line-shape object, which has fine structure. Thus, we adjust the convolution_param of the first convolutional stage to retain more details. More implementation details can be found in the next section.

Let $T = \{(X_n, Z_n, n = 1, ..., N)\}$ denote the training data set, where $X_n = \left\{x_j^n, j = 1, ..., |X_n|\right\}$ is the input image and $Z_n = \left\{z_j^n, j = 1, ..., |X_n|\right\}, z_j^n \in [0, 1]$ denotes the corresponding ground truth binary vessel segmentation map (label information) for image X_n. We omit the subscript n for notational convenience since we assume the inputs are all independent of one another. We denote the collection of all standard network layer parameters as $\mathbf{W}$. Suppose in the network we have totally M side-output layers. Each side-output layer is associated with a classifier, in which the corresponding weights can be represented as $\mathbf{w} = \left(\mathbf{w}^1, \mathbf{w}^2, ..., \mathbf{w}^M\right)$. Thus, the side objective function of HED can be given by

$$L_{side}\left(\mathbf{W}, \mathbf{w}\right) = \sum_{m=1}^{M} \alpha_m l_{side}^{(m)}\left(\mathbf{W}, \mathbf{w}^{(m)}\right) \tag{1}$$

where α_m is the weight of the mth side loss and $l_{side}^{(m)}$ denotes the image-level class-balanced cross-entropy loss function for the mth side output. Besides, a weighted-fusion layer is added to better capture the advantage of each side output. The fusion loss at the fusion layer can be expressed as

$$L_{fuse}(\mathbf{W}, \mathbf{w}, \mathbf{f}) = \sigma(Z, h(\sum_{m=1}^{M} f^m A_{side}^{(m)})) \tag{2}$$

where $\mathbf{f} = (f_1, ..., f_M))$ is the fusion weight, $A_{side}^{(m)}$ are activations of the mth side output, $h(\cdot)$ denotes the sigmoid function, and $\sigma(\cdot, \cdot)$ denotes the distance between the ground truth map and the fused predictions, which is set to be image-level class-balanced cross-entropy lass here. Therefore, the final loss function can be give by

$$L_{fical}(\mathbf{W}, \mathbf{w}, \mathbf{f}) = L_{fuse}(\mathbf{W}, \mathbf{w}, \mathbf{f}) + L_{side}(\mathbf{W}, \mathbf{w}) \tag{3}$$

HED connects each side output to the last convolutional layer (fuse layer) in each stage of the VGGNet. Each intermediate layer and the last fuse layer can predict a learned segmentation result from the trained net model.

B. Smoothly Regularized Network

Conditional Random Field (CRF) usually is used in the context of pixel-wise label prediction, which models pixel labels as random variables that form a Markov Random Field (MRF) when conditioned upon a global observation (the image). Note that, different from the previous methods which use CRF as a post-processing tool, the proposed DSSRN integrates CRF as a module of the deep network and optimize the parameters in an end-to-end manner.

Let X_i be the random variable associated to pixel i, which represents the label assigned to the pixel i and can take any value from a predefined set of labels $\mathcal{L} = \{1_1, 1_2, ..., l_L\}$. Let $\mathbf{X}$ be the vector formed by the random variables $X_1, X_2, ..., X_N$, where N is the number of pixels in the image. Given a graph $G = (V, E)$, where $V = \{X_1, X_2, ..., X_N\}$, and a global observation (the image) $\mathbf{I}$, the pair $(\mathbf{I}, \mathbf{X})$ can be modeled as a CRF characterized by a Gibbs distribution of the form $P(\mathbf{X} = \mathbf{x}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})} \exp\left(-E(\mathbf{x}|\mathbf{I})\right)$. Here $E(\mathbf{x})$ is called the energy of the configuration $\mathbf{x} \in \mathcal{L}^N$ and $Z(\mathbf{I})$ is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2844861, IEEE Access

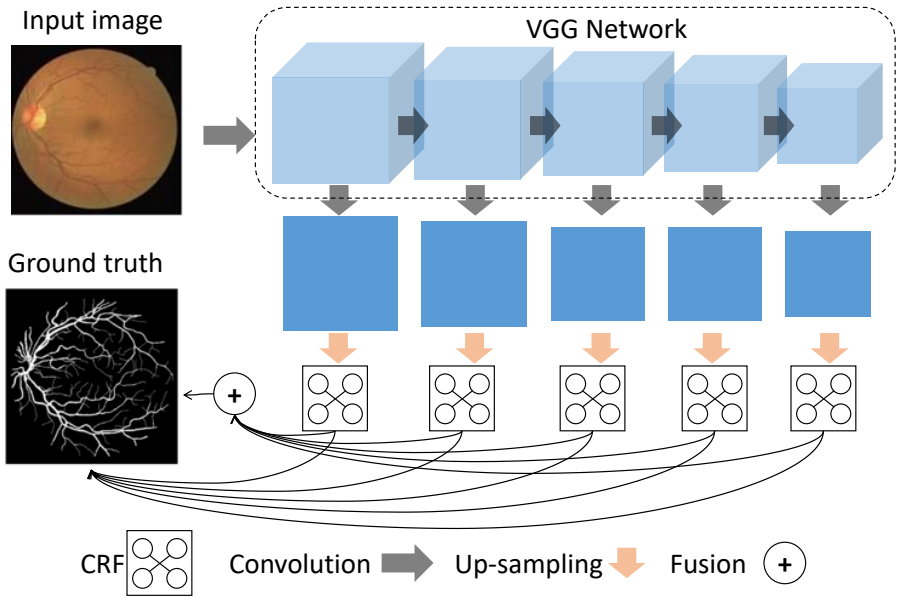

Fig. 1. The architecture of the proposed DSSRN. The network combines deep supervision of VGG network and smoothness regularization of CRF, is an end-to-end and pixel-to-pixel network for retinal vessel segmentation.

the partition function (We drop the conditioning on $\mathbf{I}$ in the notation for convenience).

Krähenbühl et al. [27] proposed the fully connected pairwise CRF model, the energy of a label assignment $\mathbf{x}$ is given by

$$E(\mathbf{x}) = \sum_i \varphi_u(x_i) + \sum_{i<j} \varphi_p(x_i, x_j) \tag{4}$$

where the unary energy components $\varphi_u(x_i)$ measure the inverse likelihood of the pixel i taking the label x_i, and pairwise energy components $\varphi_p(x_i, x_j)$ measure the cost of assigning labels x_i, x_j to pixels i, j simultaneously. In our model, unary energies are obtained from HED, which, roughly speaking, predicts labels for pixels without considering the smoothness and the consistency of the label assignments. The pairwise energies provide an image data-dependent smoothing term that encourages assigning similar labels to pixels with similar properties. We model pairwise potentials as a weighted Gaussian

$$\varphi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{M} w^m k_G^{(m)}(\mathbf{f}_i, \mathbf{f}_j) \tag{5}$$

where each $k_G^{(m)}$ for $m = 1, ..., M$, 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 function $\mu(\cdot, \cdot)$, called the label compatibility function, captures the compatibility between different pairs of labels as the name implies.

Minimizing the above CRF energy $E(\mathbf{x})$ yields the most probable label assignment $\mathbf{x}$ for the given image. Since this exact minimization is intractable, a mean-field approximation to the CRF distribution is used for approximate maximum posterior marginal inference. The mean-field CRF is based on CRFasRNN, which referred that filter-based approximate mean-field inference approach for dense CRFs relies on applying Gaussian spatial and bilateral filters on the mean-field approximates in each iteration. Unlike the standard convolutional layer in a net, in which filters are fixed after the training stage, the method uses edge-preserving Gaussian filters, coefficients of which depend on the original spatial and appearance information of the image.

Combined mean-field CRF into HED, we construct an end-to-end deep network for retinal vessel segmentation. On the one hand, the method inherits the deeply supervised information from fully convolutional network; on the other hand, it contains the global smoothness correlation via conditional random field. In the experiments, we will show the effectiveness of end-to-end training by comparing the methods which use CRF as a post-processing tool.

IV. Experiments

We evaluate our method on three publicly available DRIVE [8], STARE [17] and CHASE_DB1 [18] datasets. These three datasets all provide two manual segmentation results generated by two different experts for each image, we select the first observer's as the ground truth. Here, we perform the evaluation in terms of Accuracy (Acc) and Sensitivity (Sen), their definitions are shown as follows, and the evaluation is computed in the field of view (FOV, the circular active area) of retinal images.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Sen = \frac{TP}{TP + FN} \tag{7}$$

where TP, TN, FP, FN are respectively the numbers of true positive, true negative, false positive and false negative decisions. To put it simply, we consider P as the predicted

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2844861, IEEE Access

pixel is vessel, and N means the predicted pixel is background, exactly corresponding to the two objects in retinal images. Acc is measured by the ratio of the total number of correctly classified pixels to the number of all pixels in the image field of view. Sen represents the ability of detecting the vessel pixels correctly.

Examples and detailed messages of the three datasets are illustrated in Fig IV and Table I respectively. We choose one half of the datasets as training images, the other half as test images, which is the standard setting has been applied in previous works [26], [24], [20], [22], [21], [8].

TABLE I
DETAILED INFORMATION OF THE THREE STUDIED DATASETS.

Dataset	Image size	Training images	Testing images
DRIVE	564×584	20	20
STARE	700×605	10	10
CHASE_DB1	999×960	14	14

A. Implementation details

In our architecture, DSSRN is based on HED and CRF. HED is a five-stage deeply supervised network, the segmentation results form each layer as well as the the last fuse layer then connects mean-field CRF, which provides the global smoothness information for the whole net.

The hyper-parameters used in this work contain: the learning rate is initially set to 10^{-8}, and then decreased by a factor of 10 every 10000 iterations. The momentum is 0.9, weight_decay is 0.0002. Our target, the blood vessel is a line-shape and fine-structure object, as the vessel width can range from one pixel to twenty pixels. To get accurate segmentation result, we adjust the convolution_param of HED, make the receptive field be close to the original input. The convolution_param(pad) of the first convolutional stage is changed from 35 to 1. The fusion layer weights are initialized with 0.2, 0.28, 0.28, 0.2, 0.04. We set the number of mean-field iterations in the CRF layer to 5 during training, to avoid exploding gradient problems and to reduce the training time; and increase the iteration count to 10 during testing.

B. Running time

The implementation is based on the Caffe framework [28], and conducted on NVIDIA GTX Titan X GPU. We first prepare one half, 44 images from three datasets (DRIVE, STARE and CHASE_DB1) as the training images, and then raise the data using flip transform. We use the random crop data augmentation by randomly cropping 500×500 sub-images during training. Finally, we obtain the trained model after 10000 iterations. Training of a model can take up to 12 hours on a single GPU, and the testing of one retinal image only needs 0.3 seconds in average.

C. Results and Comparison with the State-of-the-arts

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. The details of the three tested datasets are given separately as follows.

TABLE II
RESULTS ON THE THE DRIVE DATASET. THE RESULTS ARE EVALUATED USING THE ACCURACY (ACC) AND SENSITIVITY (SEN) CRITERIA.

Methods	Acc	Sen
Human observer	0.9473	0.7482
Unsupervised methods		
Zana et al [29]	0.9377	0.6971
Jiang et al [30]	0.9212	-
Mendonca et al [31]	0.9452	0.7344
Miri et al [32]	0.9458	0.7352
Fraz et al [33]	0.9430	0.7152
You et al [34]	0.9434	0.7410
Supervised methods		
Niemeijer et al [35]	0.9416	-
Cheng et al [36]	0.9474	0.7252
Staal et al [8]	0.9442	-
Fraz et al [21]	0.9480	0.7406
Marin et al [20]	0.9452	0.7067
Liskowski et al [24]	0.9495	**0.7763**
Fu et al [26]	0.9470	0.7294
Ours (DSSRN)	**0.9536**	0.7632

The results of proposed DSSRN are shown and compared to the previous state-of-the-art methods in Table II, Table III and Table IV on the DRIVE dataset, the STAR dataset and the CHASE_DB1 dataset, respectively. In each table, there are three sections: the first section shows the performance of human observer; the second section shows the results of some unsupervised methods; the third section shows the results of some supervised methods. The symbol "-" means the corresponding result is not available in the original paper.

TABLE III
RESULTS ON THE THE STARE DATASET. THE RESULTS ARE EVALUATED USING THE ACCURACY (ACC) AND SENSITIVITY (SEN) CRITERIA.

Methods	Acc	Sen
Human observer	0.9323	0.8517
Unsupervised methods		
Zana et al [29]	0.9009	-
Mendonca et al [31]	0.9440	0.6996
Fraz et al [21]	0.9442	0.7311
You et al [34]	0.9497	0.7260
Supervised methods		
Staal et al [8]	0.9516	-
Fraz et al [21]	0.9534	0.7548
Marin et al [20]	0.9526	0.6944
Liskowski et al [24]	0.9566	**0.7867**
Fu et al [26]	0.9545	0.7140
Ours (DSSRN)	**0.9603**	0.7423

From the results, we can observe that the supervised learning methods are generally work better than the unsupervised methods. Due to the difficulties of problem, the progress on this problem is not fast. However, it is exciting that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2844861, IEEE Access

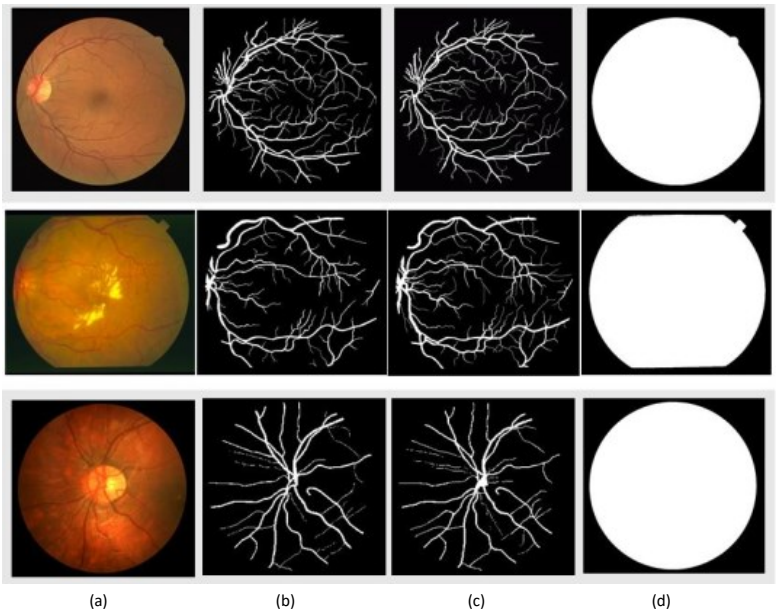

Fig. 2. Example images in the three tested datasets. Each row shows an example from a dataset. (a) shows the original retinal image; (b) and (c) show two annotated ground truth images; (c) shows the mask of the retinal region.

many methods have outperform the human observer, which is calculated on the second observer (GT_2nd).

Among the compare methods, our method can obtain better Acc performance than others on the three datasets, and get the best results on the STARE and CHASE_DB1 datasets. Note that these two datasets contains more pathological retinal images, which shows that our method is efficient and robust. Also, our method outperforms than the human observer.

TABLE IV
RESULTS ON THE THE CHASE_DB1 DATASET. THE RESULTS ARE EVALUATED USING THE ACCURACY (ACC) AND SENSITIVITY (SEN) CRITERIA.

Methods	Acc	Sen
Human observer	0.9520	0.7038
Unsupervised methods		
Azzopardi et al [37]	0.9387	0.7585
Supervised methods		
Fraz et al [21]	0.9469	0.7224
Roychowdhury et al [22]	0.9530	0.7259
Ours (DSSRN)	**0.9587**	**0.7815**

Compared with the DeepVessel work proposed by Fu et al [26], where CRF inference is applied as a post-processing method, according to Table II and Table III, the proposed DSSRN consistently obtains better results; especially, the end-to-end training strategy obtains a significant better performance when considering sensitivity. Overall, the proposed DSSRN method obtains the state-of-the-art performance on the three tested datasets.

D. Ablation studies

The effectiveness of data augmentation: Data augmentation is widely used in deep learning for natural image object recognition/segmentation. In this retinal vessel segmentation task, data augmentation should be also applied since the number of training image is relatively small considering there are usually tens of thousands of natural images for training. As mentioned above, we perform random crop data augmentation on training images. For each training, we randomly crop a 500×500 sub-image in each iteration during training. In testing, image is resized to 500×500 before it is fed into the network. To verify the effectiveness of the data augmentation, we carry out experiments on the DRIVE dataset without using data augmentation, we obtain 0.9342 accuracy and 0.7028 sensitivity. The results show that the data augmentation is essential to obtain an excellent performance on this task.

The effectiveness of the smoothness regularization: By comparing the proposed DSSRN with the DeepVessel method, we can observe that the end-to-end learning of CRF with deep CNN is helpful. Furthermore, we conduct experiments on the STARE using CRF; thus, we use test the HED network on the retinal vessel segmentation task. The results are 0.9415 accuracy and 0.7142 sensitivity. The results confirms the effectiveness of the CRF smoothness regularization. To qualitative show the contribution of the smoothness regularization, Fig IV-D shows some vessel segmentation results using and without using the regularization term. From the images, we can observe that the output the proposed DSSRN is adaptively smoothed compared to the HED results.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2844861, IEEE Access

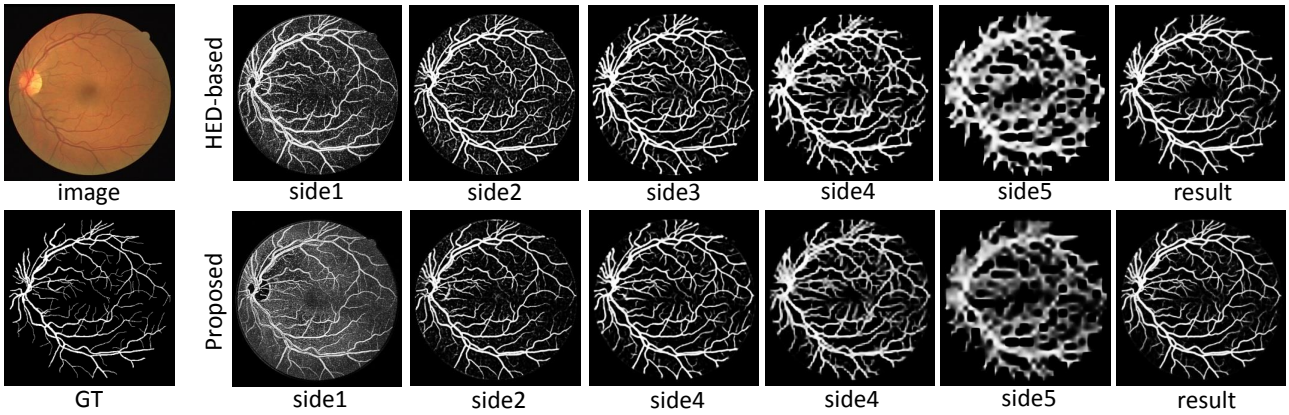

Fig. 3. Visualization of the retinal vessel segmentation result by comparing with HED. The first column is the original image and corresponding ground truth, the rest columns are the side-output and final segmentation result of network. (The original image is from the DRIVE dataset: 01_test.jpg.)

V. Conclusion

In this paper, we developed a new deep learning based method for retinal vessel segmentation. The method combines deep supervision of HED architecture and global smoothness regularization of CRF, is an end-to-end and pixel-to-pixel segmentation network. It outperforms HED-based method, and outperforms a system where CRF inference is applied as a post-processing method. Our experiments show that our proposed method achieves the state-of-the-art results on DRIVE, STARE and CHASE_DB1 datasets, is efficient to retain more vessels in retinal images.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful suggestions.

References

[1] M. L. Baker, P. J. Hand, J. J. Wang, and T. Y. Wong, "Retinal signs and stroke," *Stroke*, vol. 39, no. 4, pp. 1371–1379, 2008.

[2] N. Patton, T. Aslam, T. MacGillivray, A. Pattie, I. J. Deary, and B. Dhillon, "Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: a rationale based on homology between cerebral and retinal microvasculatures," *Journal of anatomy*, vol. 206, no. 4, pp. 319–348, 2005.

[3] M. Niemeijer, B. Van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu *et al.*, "Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 185–195, 2010.

[4] T. Teng, M. Lefley, and D. Claremont, "Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy," *Medical and Biological Engineering and Computing*, vol. 40, no. 1, pp. 2–13, 2002.

[5] M. Foracchia, E. Grisan, and A. Ruggeri, "Extraction and quantitative description of vessel features in hypertensive retinopathy fundus images," in *Book Abstracts 2nd International Workshop on Computer Assisted Fundus Image Analysis*, vol. 6, 2001.

[6] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010.

[7] E. Grisan and A. Ruggeri, "A divide et impera strategy for automatic classification of retinal vessels into arteries and veins," in *Engineering in medicine and biology society, 2003. Proceedings of the 25th annual international conference of the IEEE*, vol. 1. IEEE, 2003, pp. 890–893.

[8] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. 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.

[9] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification," *IEEE Transactions on medical Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006.

[10] E. Ricci and R. 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.

[11] L. M. Lorigo, O. D. Faugeras, W. E. L. Grimson, R. Keriven, R. Kikinis, A. Nabavi, and C.-F. Westin, "Curves: Curve evolution for vessel segmentation," *Medical image analysis*, vol. 5, no. 3, pp. 195–206, 2001.

[12] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1706–1721, 2014.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[14] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1395–1403.

[15] X. He, R. S. Zemel, and M. Á. Carreira-Perpiñán, "Multiscale conditional random fields for image labeling," in *Computer vision and pattern recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE computer society conference on*, vol. 2. IEEE, 2004, pp. II–II.

[16] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

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

[18] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program," *Investigative ophthalmology & visual science*, vol. 50, no. 5, pp. 2004–2010, 2009.

[19] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on medical imaging*, vol. 8, no. 3, pp. 263–269, 1989.

[20] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on medical imaging*, vol. 30, no. 1, pp. 146–158, 2011.

[21] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.

[22] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Blood vessel segmentation of fundus images by major vessel extraction and subimage classification," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 1118–1128, 2015.

[23] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Transactions on image Processing*, vol. 25, no. 1, pp. 53–64, 2016.

[24] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with¡? pub _newline?¿ deep neural networks," *IEEE transactions on medical imaging*, vol. 35, no. 11, pp. 2369–2380, 2016.

[25] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[26] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 698–701.

[27] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.

[28] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[29] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE transactions on image processing*, vol. 10, no. 7, pp. 1010–1019, 2001.

[30] X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 131–137, 2003.

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

[32] M. S. Miri and A. Mahloojifar, "Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1183–1192, 2011.

[33] M. M. Fraz, S. A. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A. R. Rudnicka, and C. G. Owen, "An approach to localize the retinal blood vessels using bit planes and centerline detection," *Computer methods and programs in biomedicine*, vol. 108, no. 2, pp. 600–616, 2012.

[34] X. You, Q. Peng, Y. Yuan, Y.-m. Cheung, and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2314–2324, 2011.

[35] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *Medical Imaging 2004: Image Processing*, vol. 5370. International Society for Optics and Photonics, 2004, pp. 648–657.

[36] E. Cheng, L. Du, Y. Wu, Y. J. Zhu, V. Megalooikonomou, and H. Ling, "Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features," *Machine vision and applications*, vol. 25, no. 7, pp. 1779–1792, 2014.

[37] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical image analysis*, vol. 19, no. 1, pp. 46–57, 2015.

Yi Lin received the B.S. degree in microelectronics from Nanjing University Of Posts And Telecommunications, NanJing, China, He is currently pursuing the Ph.D. degree in information and communication engineering at Beijing University Of Posts and Telecommunications. He worked as a Visiting Scholar in School of Electrical and Computer Engineering,Georgia Institute of Technology from 2015 to 2017. His research interest includes image retrieval, image annotation, medical image analysis, and deep learning.

Honggang Zhang received the B.S degree from the department of Electrical Engineering, Shandong University in 1996, the Master and Ph.D degrees from the School of Information Engineering, Beijing University of Posts and Telecommunications (BUPT) in 1999 and 2003 respectively. He worked as a Visiting Scholar in School of Computer Science, Carnegie Mellon University (CMU) from 2007 to 2008. He is currently an Associate Professor and Director of web search center at BUPT. His research interests include image retrieval, computer vision and pattern recognition. He published more than 30 papers on TPAMI, SCIENCE, Machine Vision and Applications, AAAI, ICPR, ICIP. He is a senior member of IEEE.

Guang Hu received the B.S. degree in communication engineering from Huazhong University of Science and Technology, China, in 1999. From 1999 to 2018, he was a lecturer with the School of Electronic and Information and Communications, Huazhong University of Science and Technology. His research areas include deep learning, human behavior study and network security.