Computer Methods and Programs in Biomedicine 148 (2017) 13–25

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement

Lei Zhou[a], Qi Yu[b], Xun Xu[b], Yun Gu[a], Jie Yang[a,*]

[a] *Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, SEIEE Building 2-427, No. 800, Dongchuan Road, Minhang District, Shanghai, 200240 China*
[b] *Shanghai General Hospital, Shanghai Jiao Tong University, Shanghai, China*

ARTICLE INFO

Article history:
Received 13 July 2016
Revised 28 May 2017
Accepted 23 June 2017

Keywords:
Retinal vessel segmentation
Dense conditional random field
Convolutional neural network
Feature learning
Image enhancement

ABSTRACT

Background and objectives: As retinal vessels in color fundus images are thin and elongated structures, standard pairwise based random fields, which always suffer the “shrinking bias” problem, are not competent for such segmentation task. Recently, a dense conditional random field (CRF) model has been successfully used in retinal vessel segmentation. Its corresponding energy function is formulated as a linear combination of several unary features and a pairwise term. However, the hand-crafted unary features can be suboptimal in terms of linear models. Here we propose to learn discriminative unary features and enhance thin vessels for pairwise potentials to further improve the segmentation performance.
Methods: Our proposed method comprises four main steps: firstly, image preprocessing is applied to eliminate the strong edges around the field of view (FOV) and normalize the luminosity and contrast inside FOV; secondly, a convolutional neural network (CNN) is properly trained to generate discriminative features for linear models; thirdly, a combo of filters are applied to enhance thin vessels, reducing the intensity difference between thin and wide vessels; fourthly, by taking the discriminative features for unary potentials and the thin-vessel enhanced image for pairwise potentials, we adopt the dense CRF model to achieve the final retinal vessel segmentation. The segmentation performance is evaluated on four public datasets (i.e. DRIVE, STARE, CHASEDB1 and HRF).
Results: Experimental results show that our proposed method improves the performance of the dense CRF model and outperforms other methods when evaluated in terms of F1-score, Matthews correlation coefficient (MCC) and G-mean, three effective metrics for the evaluation of imbalanced binary classification. Specifically, the F1-score, MCC and G-mean are 0.7942, 0.7656, 0.8835 for the DRIVE dataset respectively; 0.8017, 0.7830, 0.8859 for STARE respectively; 0.7644, 0.7398, 0.8579 for CHASEDB1 respectively; and 0.7627, 0.7402, 0.8812 for HRF respectively.
Conclusions: The discriminative features learned in CNNs are more effective than hand-crafted ones. Our proposed method performs well in retinal vessel segmentation. The architecture of our method is trainable and can be integrated into computer-aided diagnostic (CAD) systems in the future.

1. Introduction

Color fundus images are acquired by making photographs of the back of the eye, where blood vessels of humans can be directly visualized in such a non-invasive way [1]. As one of the main features in fundus images, the appearance of retinal blood vessels can provide useful information for the early diagnosis of various diseases, including diabetes, glaucoma, hypertension, arteriosclerosis and cardiovascular diseases [2]. Accurate delineation and measurement of retinal vessels is therefore an important prerequisite for a number of clinical applications. Moreover, knowledge on retinal vessel location also facilitates the localization of the optic disc and fovea, the registration of multimodal images, and the detection of retinal pathologies such as exudates and red lesions [3]. Nowadays, manual segmentation of retinal vessels by trained specialists becomes infeasible due to the massive images produced by the large-scale fundus screening, so accurate automated retinal vessel segmentation turns to be quite valuable in the development of CAD systems.

* Corresponding author.
E-mail addresses: minizon@sina.com (L. Zhou), jieyang@sjtu.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.cmpb.2017.06.016

During the past decades, various kinds of methods for automatic retinal vessel segmentation have been proposed in the literature [3–10]. As summarized by Fraz et al. [3], these methods can be generally classified into two categories: unsupervised and supervised. Unsupervised methods mostly rely on thresholding filter responses, vessel tracking techniques, or other rule-based approaches, while supervised methods manage to train a classifier or a model to achieve the discrimination between vessel and non-vessel pixels [10].

On one hand, most existing methods in either of the two categories depend on hand-crafted features to characterize the difference between vessel and non-vessel pixels. One of the most popular techniques used for extracting vessels is the matched filtering [11]. This methodology exploits the piecewise linear approximation and the Gaussian-like intensity profile of retinal vessels to enhance them before thresholding, and a number of variants [12–16] have been proposed to further improve the performance. Wavelet transformation, especially the multiscale Gabor transformation [17], is an effective alternative for training classifiers to achieve vessel pixel detection. And responses of line operators have also been proved to be simple and effective features for supervised classification [18]. Other carefully designed features include ridge-based features [19], moment-invariant features [20], local phase [21], Hessian-based features [22], responses of COSFIRE filters [6], and so on. However, vessels varying within a wide range of width can have different profiles. Thin vessels often have the lowest contrast, and their intensity difference from wide vessels may be greater than background variations. Central reflex is very common on wide vessels in high-resolution images. Besides, structures like optic disc, retinal boundary and pathologies may be close to vessels in certain feature spaces. Although several methods combining the advantages of individual features have been proposed to overcome such problems [23–26], these hand-crafted features still can be suboptimal in both expressing the discrimination information and keeping robust to vessel variations.

In recent years, feature learning [5] and especially deep learning methods [8,27] are applied in retinal vessel segmentation. As introduced in [28], deep learning methods are able to model higher-level abstractions from data by multiple layers. Among various deep architectures, deep CNNs [29,30] have shown the high capacity in learning rich features that both accommodate within-class variance and possess discriminative information. More recently, fully convolutional networks (FCNs) have also been proposed for image segmentation with a significant reduction in computation cost [31]. Two special FCNs, named holistically-nested edge detection net (HED) [32] and U-net [33], are then proposed with the ability of distinguishing object boundaries or edges, which are similar to the thin and elongated vessels.

On the other hand, most retinal vessel segmentation methods can be considered as pixel-wise processing, such as [5,8,27]. The identification of retinal vessels is achieved by inspecting the fundus image pixel by pixel, without considering the whole structure of the retinal vasculature. Random field models are good at modeling the structural information of objects, and they have been extensively used in image segmentation. But standard pairwise based models suffer the "shrinking bias" problem, so they are not competent in segmenting the thin and elongated vessels. Although several kinds of modifications have been proposed for segmenting fine structures in natural images [34–37], only a few attempts have been made for retinal vessel segmentation [4,10].

Orlando et al. [10] successfully applied the dense CRF model in retinal vessel segmentation, with the help of efficient inference and parameter learning techniques [37,38]. This model builds long-range connections within the image, and therefore it avoids the "shrinking bias" problem, leading to a satisfactory segmentation of retinal vessels. The corresponding energy function of their model is formulated as a linear multiplication of several unary features and a pairwise term. However, the unary features are selected from the commonly used hand-crafted ones, so they can be suboptimal with respect to the linear discrimination between vessels and non-vessels.

In this paper, we propose to learn unary features for linear discrimination and enhance thin vessels for pairwise potentials, in order to further improve the segmentation performance of the dense CRF model. First of all, a set of image preprocessing techniques is applied to the fundus image, that is, smooth expansion of the image to eliminate strong edges near the FOV, and luminosity and contrast normalization to alleviate the issue of uneven illumination and contrast variability across both intra- and inter-images. Then discriminative features are generated from a properly trained CNN, which is a modified version of CNN in MatConvNet [39] inspired by Wang et al. [8]. Meanwhile, we apply a combo of filters to the green channel of the preprocessed image in order to enhance thin vessels. The intensity difference between thin and wide vessels in the enhanced image is then reduced. Finally, the dense CRF model, which takes discriminative features for unary potentials and the thin-vessel enhanced image for pairwise potentials, is established to achieve the final retinal vessel segmentation. Parameters of the dense CRF model are learned from the training images, with the help of structured support vector machine (SSVM) and fast inference [10,37,38]. The main contributions in our paper can be summarized as follows:

1. We propose to learn discriminative unary features via CNNs in order to improve the dense CRF model for retinal vessel segmentation. Compared to the selection of hand-crafted features in [10], these high-order CNN features can be intrinsically linearly combined as unary potentials. To avoid the severe imbalanced training of Wang et al. [8], we devise a superpixel-level balanced sample selection strategy to properly train our CNN.
2. We propose a new vessel enhancement process, which aims at enhancing the thin vessels to reduce the intensity difference between thin and wide vessels and further improving the dense CRF model for retinal vessel segmentation.
3. We've made experiments on four public retinal image datasets, demonstrating that our proposed method outperforms other methods in terms of evaluation metrics such as F1-score, Matthews correlation coefficient and G-mean and that our proposed method indeed improves the dense CRF for retinal vessel segmentation.

The rest of this paper is organized as follows. Section 2 introduces four public retinal image datasets (i.e. DRIVE, STARE, CHASEDB1 and HRF) that are tested by our method. Section 3 describes the details of our methodology. In Section 4, the performances of our proposed method are presented, and important factors are discussed to further verify the effectiveness of our method. Finally, in Section 5, we make conclusions of the proposed method.

2. Materials

Four publicly available datasets of fundus images are used for the evaluation of retinal vessel segmentation algorithms:

DRIVE dataset [19] contains 40 color fundus images with a 45° FOV and a 565 × 584 pixel resolution. This dataset is equally divided into a training and a test set. The ground truth and FOV mask is provided for each image, and the diameter of the FOV mask are approximately 539 pixels. An additional human annotation of retinal vessels is provided for each test image.

STARE dataset [40] includes 20 images, 10 of them containing pathologies, captured at a 35° FOV and with a resolution of 700 × 605 pixels. Two human annotations are provided for each image. The FOV mask is obtained by the software provided by Hoover