Biomedical Signal Processing and Control 44 (2018) 110–126

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

A visual attention guided unsupervised feature learning for robust vessel delineation in retinal images

Chetan L. Srinidhi [a,*], P. Aparna [a], Jeny Rajan [b]

[a] Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, India
[b] Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal, India

ARTICLE INFO

Article history:
Received 8 January 2018
Received in revised form 2 April 2018
Accepted 15 April 2018

Keywords:
Retinal image
Vessel segmentation
Visual attention
Unsupervised feature learning

ABSTRACT

Background and objective: Accurate segmentation of retinal vessels from color fundus images play a significant role in early diagnosis of various ocular, systemic and neuro-degenerative diseases. Segmenting retinal vessels is challenging due to varying nature of vessel caliber, the proximal presence of pathological lesions, strong central vessel reflex and relatively low contrast images. Most existing methods mainly rely on carefully designed hand-crafted features to model the local geometrical appearance of vasculature structures, which often lacks the discriminative capability in segmenting vessels from a noisy and cluttered background.

Methods: We propose a novel visual attention guided unsupervised feature learning (VA-UFL) approach to automatically learn the most discriminative features for segmenting vessels in retinal images. Our VA-UFL approach captures both the knowledge of visual attention mechanism and multi-scale contextual information to selectively visualize the most relevant part of the structure in a given local patch. This allows us to encode a rich hierarchical information into unsupervised filtering learning to generate a set of most discriminative features that aid in the accurate segmentation of vessels, even in the presence of cluttered background.

Results: Our proposed method is validated on the five publicly available retinal datasets: DRIVE, STARE, CHASE_DB1, IOSTAR and RC-SLO. The experimental results show that the proposed approach significantly outperformed the state-of-the-art methods in terms of sensitivity, accuracy and area under the receiver operating characteristic curve across all five datasets. Specifically, the method achieved an average sensitivity greater than 0.82, which is 7% higher compared to all existing approaches validated on DRIVE, CHASE_DB1, IOSTAR and RC-SLO datasets, and outperformed even second-human observer. The method is shown to be robust to segmentation of thin vessels, strong central vessel reflex, complex crossover structures and fares well on abnormal cases.

Conclusions: The discriminative features learned via visual attention mechanism is superior to hand-crafted features, and it is easily adaptable to various kind of datasets where generous training images are often scarce. Hence, our approach can be easily integrated into large-scale retinal screening programs where the expensive labelled annotation is often unavailable.

1. Introduction

Retinal fundus image provides rich information about the early manifestation of various diseases related to the eye, cardiovascular and neurodegenerative diseases [1–3]. These images are often routinely acquired for non-invasive examination of its anatomical components such as vessel tree, optic disc and the fovea. Segmentation and quantification of retinal vessel tree provide important clinical biomarkers through the analysis of its geometrical properties, that aid in early diagnosis of various diseases such as diabetes [1], stroke [4], hypertension [5], arteriosclerosis [6] and cerebral small vessel diseases. Numerous longitudinal studies have been conducted in the past, that shows a strong and consistent link between retinal microvasculature with incident clinical stroke, hypertension, cardiovascular and various neurodegenerative diseases [4,5,7]. Changes in microvascular geometrical patterns such as vessel width, tortuosity, fractal dimension and branching angle provide an early insights into the progression of aforementioned diseases. Therefore, an accurate delineation of retinal vessel

* Corresponding author.
E-mail addresses: srinidhipy@gmail.com (C.L. Srinidhi), p.aparnadinesh@gmail.com (P. Aparna), jenyrajan@gmail.com (J. Rajan).

https://doi.org/10.1016/j.bspc.2018.04.016
1746-8094/

structures from color fundus image is of interest. Manual segmentation of vessel tree is often tedious, time-consuming and prone to large intra and inter-observer variability. In addition, manual delineation often requires careful interpretation of images, which is very cumbersome and painful especially for large population-based screening programs. Hence, there is a need for the automated segmentation of retinal vessels for the accurate quantification of vascular changes, along the entire disease course.

Segmenting retinal vessel tree is extremely challenging due to multi-scale nature of varying vessel caliber, the presence of strong central vessel reflex, the close proximity of pathological lesions, close parallel and highly curved vessels and complex crossover regions. In the past decades, many solutions have been proposed with techniques ranging from conventional matched filtering [8] to more recent convolutional neural network [9,10] based approaches. The existing techniques can be broadly divided into two main categories: unsupervised and supervised methods [11]. Supervised methods require a set of manually labelled training images for classifying a pixel in a previously unseen image. Whereas, unsupervised methods can segment the vessels without requiring any manual labelled annotations. In general, most of the existing techniques mainly rely on carefully designed hand-crafted filters to inherently model the local geometrical appearance of vessel structures. For example, the hand-crafted filters such as Gabor filters [12], multi-scale derivative of Gaussian [13], matched filters [8], ridge detector [14], line detector [15], wavelet transform [16,17], moment invariant features [18], first and second order derivatives of Gaussian [19] and response of COSFIRE filters [20,21] to name a few. These hand-crafted filters were designed based on complex domain knowledge and require careful parameter tuning to achieve optimal segmentation performance, across a wide variety of data. Besides, the response of these filters often poorly represents the appearance of thin vessel structures, crossover and bifurcation regions, highly curved tortuous vessels and susceptible to non-illumination present in an image. The other drawback of these methods is that the extracted features often lack the discriminative capability to predict the actual class label, even in the presence of similar looking cluttered objects.

To overcome the aforementioned limitations, various automatic feature learning algorithms [9,10,22,23] have been proposed to learn the feature representations directly from the training data. These approaches are primarily motivated by the success of deep learning (DL) methods, which is applied in various computer vision applications such as object recognition, scene classification, semantic segmentation, etc. The success of these methods is critically dependent on an enormous amount of labelled training data which is typically expensive in medical imaging applications. To address these shortcomings, various unsupervised feature learning (UFL) algorithms [24–28] have been proposed to automatically learn the feature representations, only from a set of unlabelled data. Automatic feature learning enables to encode rich hierarchical information that learns to map complex functions from input to output, directly from the data, without depending on hand-crafted features.

The main bottleneck for the accurate segmentation of retinal vessels comes from the multiscale nature of varying vessel caliber, poor visibility of low contrast thin vessels, the proximal presence of pathological structures and poor vessel connectivity at complex junction locations. The automatic features learned from these challenging locations often lack discriminative capability in accurately identifying vessel pixels from similar-looking cluttered objects. This is mainly because, the traditional UFL approaches encodes feature representation from a limited input patch size, which is often referred to as "*receptive field*". The selection of this receptive field mainly depends on the object of interest, which we are trying to encode. In case of retinal vessels, the structure of interest varies significantly, resulting in difficulty in choosing an appropriate receptive field size, that fits for a wide range of input data. Further, various size of receptive fields encodes different *contextual* information (inter-pixel dependencies), thereby resulting in different feature representation for the same pixel centred on a patch.

To alleviate this problem, we propose a novel UFL approach which is primarily inspired by the visual attention mechanism in the human visual system where we humans have the capability to pay attention *selectively* to the part of the image, instead of processing the whole scene in its entirety [29,30]. Such a selection mechanism is often referred to as "*visual attention prediction*". In this work, we leverage this idea of visual attention mechanism with the unsupervised feature learning approach, to automatically learn the most relevant hierarchical features from unlabelled data. The proposed visual attention guided unsupervised feature learning (VA-UFL) approach automatically learns to selectively pay attention to the most relevant part of the structure in a given input patch, and use this information to selectively encode features for subsequent classification. The feature learned via this approach offers an edge over tradition UFL methods – by exploring both the notion of *selection mechanism* and the *multi-scale contextual information* under a single framework. This allows us to learn the most relevant hierarchical features at multiple scales, which encodes sufficient information about the most salient region as well as the multi-scale context for a given input patch. The advantage of our proposed approach is that we encode a complex hierarchical level of information with such a simple approach, without relying on sophisticated feature learning architectures.

The key contributions of this paper can be summarized as follows.

1. We propose to explore the idea of visual attention mechanism that learns to selectively pay attention to the most relevant structure *and* capture multi-scale contextual information from a given local patch. This, in turn, drives the subsequent UFL approach to encode the rich hierarchical information for the automatic feature discovery in retinal images.
2. We show that the proposed visual attention model, when leveraged with the UFL approach, aims to automatically learn the most discriminative set of features that underscores the inter-class differences (between the vessel and background pixels) to be large, while keeping intra-class differences (between vessel pixels) to be small, resulting in the accurate segmentation of retinal vessels.
3. We validated our approach on five publicly available retinal datasets, including both the RGB and SLO images. The extensive experimental analysis demonstrates the effectiveness of the proposed approach in handling all the challenging cases, compared with the state-of-the-art methods.

The rest of this paper is organized as follows. In Section 2, relevant literature to this work is briefly reviewed. Section 3 explains in detail the proposed framework. In Section 4, we provide information about the dataset used, quantitative measures employed, parameter settings, followed by experimental results. Section 5 discusses the key findings, followed by future directions and finally, we conclude in Section 6.

2. Background

In past decades, many methods have been proposed for the automatic segmentation of vessels from retinal color fundus images. These methods can be broadly divided into two main