Computerized Medical Imaging and Graphics 55 (2017) 28–41

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation

Julian Zilly[a], Joachim M. Buhmann[b], Dwarikanath Mahapatra[b,*]

[a] Department of Mechanical Engineering, ETH Zurich, Switzerland
[b] Department of Computer Science, ETH Zurich, Switzerland

ARTICLE INFO

Article history:
Received 1 February 2016
Received in revised form 16 July 2016
Accepted 29 July 2016

Keywords:
Optic cup
Optic disc
Segmentation
Glaucoma
Ensemble learning
CNN
Boosting

ABSTRACT

We present a novel method to segment retinal images using ensemble learning based convolutional neural network (CNN) architectures. An entropy sampling technique is used to select informative points thus reducing computational complexity while performing superior to uniform sampling. The sampled points are used to design a novel learning framework for convolutional filters based on boosting. Filters are learned in several layers with the output of previous layers serving as the input to the next layer. A softmax logistic classifier is subsequently trained on the output of all learned filters and applied on test images. The output of the classifier is subject to an unsupervised graph cut algorithm followed by a convex hull transformation to obtain the final segmentation. Our proposed algorithm for optic cup and disc segmentation outperforms existing methods on the public DRISHTI-GS data set on several metrics.

1. Introduction

Glaucoma is one of the leading causes of irreversible vision loss in the world, accounting for 12% of such cases. It is estimated that almost 80 million people globally may be affected with glaucoma by 2020 (World Health Organization, 2006). Glaucoma is characterized by damage to the optic nerve through increasing degeneration of the nerve fibers. The progression of the disease is asymptomatic in the early stages but gradually leads to irreversible vision loss. Although there is no known cure, early treatment has been shown to decrease the rate of blindness by around 50% (Michelson et al., 2008). Hence it is essential to have a reliable early detection system for glaucoma onset. This work proposes a computationally efficient method using a ensemble learning based convolutional neural network (CNN) architecture for accurate and robust segmentation of the optic cup (OC) and optic disc (OD) from retinal fundus images. The segmented OC and OD are used to calculate the cup-to-disc ratio (CDR) which is an important indicator of glaucoma progression.

Ophthalmologists use three principal methods to detect onset of glaucoma (Cheng et al., 2013). The first approach is the assessment of increased intraocular pressure inside the eye. However, this is not sensitive enough for early detection and glaucoma can sometimes occur without increased eye pressure. The second approach identifies field of abnormal vision with specialized equipment which makes it unsuitable for a comprehensive screening of glaucoma except in sophisticated medical centers. The third approach is evaluation of damage to the optic nerve. This is most reliable but requires a trained professional, is time-consuming, expensive and highly subjective. Expert assessment may vary depending on experience and training (Cheng et al., 2013). CDR values from segmented optic cup and disc are an important indicator of damage to the optic nerve.

The use of automated diagnostic tools is desirable to minimize subjectivity and make the diagnosis robust and consistent. Color fundus imaging (CFI) has emerged as the preferred procedure for comprehensive large-scale retinal disease screening due to their ease of acquisition and good visibility of retinal structures (Singer et al., 1992). Glaucoma screening methods apply computer algorithms on color fundus retinal images for OD and OC segmentation and calculation of CDR values.

1.1. *Prior related work*

Automatic CDR measurement involves: (1) optic disc localization and segmentation; and (2) optic cup segmentation. Current state-of-the-art methods for disc segmentation use morphological features (Aquino et al., 2010) and active contours (Joshi et al.,

* Corresponding author.
E-mail address: dwarikanath.mahapatra@inf.ethz.ch (D. Mahapatra).

http://dx.doi.org/10.1016/j.compmedimag.2016.07.012
0895-6111/

2011). A OC segmentation method is proposed in Joshi et al. (2012) using depth maps computed from relatively displaced sequentially acquired images. Finally, a confidence measure is used to determine the boundary localization. Performance of these methods depends upon initialization and ability to identify weak edges. Chakravarty and Sivaswamy (2014) formulate a Markov Random Field on depth maps extracted from multiple shifted images of the same retina to model the relative depth and discontinuity at the cup boundary. This depth map is subsequently used for optic cup segmentation.

Of late machine learning methods have become popular as they provide a powerful tool for feature classification using learned models. Cheng et al. (2013) formulate a superpixel based classification method to segment the OD and OC. Center surround statistics from the super pixel neighborhood improve performance and a self-assessment reliability score indicates when a given segmentation might be less reliable. Bock et al. (2010) apply glaucoma specific preprocessing (including blood vessel removal), followed by the extraction of different generic features which are compressed using principal component analysis (PCA). A probabilistic two-stage classification scheme then combines these features types into a proposed glaucoma risk index. Mahapatra et al. use a field of experts model (Mahapatra and Buhmann, 2015) and consensus based methods (Mahapatra and Buhmann, 2015; Mahapatra, in press) for segmenting the optic cup and disc. Xu et al. (2014) focus on localizing the optic cup in fundus images and state an unsupervised closed form solution. Their technique estimates optic cup parameters from a code book and estimates the optic cup parameters through a weighted reconstruction based on training images. A prominent limitation of supervised learning methods is the definition of hand crafted features thought to be most relevant for the particular task. Such approaches do not generalize well for different datasets or application domains. Therefore many recent works on segmentation focus on learning the most discriminative features using deep learning and neural networks (Liao et al., 2013). CNNs are a general approach for learning discriminative features from training data in the form of convolutional filters.

Since we use CNNs for OC and OD segmentation we present related work on image segmentation. Mayraz and Hinton (2002) proposed a hierarchical learning procedure based on a probabilistic learning framework called the product of experts (Brown and Hinton, 2000) where the probability of an image is described by the normalized product of learned individual distributions. Kiros and Popuri (2014) use a hierarchical CNN at multiple scales for lung vessel segmentation by optimizing a 2-norm orthogonal matching pursuit problem. Given learned filters, new feature maps are extracted by convolving with the original images, and serve as input to the next layer of filter learning. Ciresan et al. (2012) use a deep neural network (DNN) to segment neuronal structures in electron microscopy (EM) images. Turaga et al. (2009) segment neuronal structures in EM images by learning an affinity graph using a CNN.

1.2. Our contribution

Previous approaches to OC and OD segmentation have used hand crafted features to segment the desired anatomy. However, it is not known whether the hand crafted features are optimal in their performance. As a result such methods do not perform equally well on a wide variety of datasets. An alternative approach is to use CNNs to learn the most discriminative features from the training data. However CNN training requires a large dataset as well as significant computing resources. Our work approaches the problem of learning feature representations from training data from a ensemble learning perspective. Our proposed method is inspired from CNNs with the learned output being a set of filters whose convolutional output provides the optimal representation of the training data. Hence there is no need to define hand crafted features since the CNN architecture learns the optimal representational features through the filters. An ensemble learning approach significantly improves the computational efficiency of training and can be used with limited training data.

The primary research contribution of our work is a hierarchical architecture of CNNs to segment the OC and OD from retinal fundus images. We introduce a novel learning procedure to construct a CNN architecture based on boosting and it shares characteristics with ensemble learning systems. Secondly, an entropy based sampling technique is presented to identify most informative samples from the training dataset and significantly reduce computational complexity. The entropy sampling method is shown to generally yield superior results when compared to uniform sampling. Overall, the proposed method is demonstrated to outperform several other state-of-the-art approaches on a public retinal image data set. Our proposed method differs from conventional CNNs in the following respects: (1) instead of backpropagation we adopt a greedy approach where each stage of filters is learned sequentially using boosting; (2) each stage considers the final classification error to update itself and not the error backpropagated through the next stages; (3) our method operates on patch level data instead of image level data used for traditional CNNs. In summary our proposed method is a ensemble learning system inspired from traditional CNNs and is an effective approach to learn convolutional filters in the absence of large numbers of training data. We describe different components of our method in Sections 2-6, present our results in Section 7, and conclude with Section 8.

2. Convolutional filters and networks

In this section we briefly describe the theory behind convolutional filters and networks. Hierarchical layers of convolutional filters that mimic the effects of visual receptive fields were inspired by Hubel and Wiesel's work on feedforward processing in the early visual cortex of cats (Hubel and Wiesel, 1963). Inspired by this CNNs use local spatial correlations in images and also exhibit robustness to natural transformations such as changes of viewpoint or scale.

2.1. Convolutional filtering

Convolution between functions f and g can be written as

$$y(t) = f(t) * g(t) = \int_{-\infty}^{\infty} f(t-\tau) * g(\tau) d\tau \tag{1}$$

where * denotes the convolution operation. The equivalent discrete formulation is

$$y[i, j] = (I * K)[i, j] = \sum_m \sum_n I[m, n] K[i-m, j-n] \tag{2}$$

where K is a two-dimensional matrix called kernel and I is a two-dimensional (gray-valued) image. Convolution can be extended to higher dimensional images and kernels. Generally kernels tend to be considerably smaller than the images.

Convolutional networks exploit three key ideas (Bengio et al., 2015): sparse interactions, parameter sharing and equivariant representations. Using convolutional filters as building blocks, complex convolutional networks can be constructed. Some of the important building blocks of CNNs are:

- **Convolutional filter**: Also called kernel, it is the fundamental building block of a CNN. Each filter is convolved with each point of an assigned input image and generates an output. Filters are generally of size $n_1 \times n_1 \times n_{maps}$ where n_1 is the specified filter