Boosting Convolutional Filters with Entropy Sampling for Optic Cup and Disc Image Segmentation from Fundus Images

Julian G. Zilly[1], Joachim M. Buhmann[2], Dwarikanath Mahapatra[2]

[1] Department of Mechanical Engineering, ETH Zurich, Switzerland
[2] Department of Computer Science, ETH Zurich, Switzerland

Abstract. We propose a novel convolutional neural network (CNN) based method for optic cup and disc segmentation. To reduce computational complexity, an entropy based sampling technique is introduced that gives superior results over uniform sampling. Filters are learned over several layers with the output of previous layers serving as the input to the next layer. A softmax logistic regression classifier is subsequently trained on the output of all learned filters. In several error metrics, the proposed algorithm outperforms existing methods on the public DRISHTI-GS data set.

1 Introduction

Glaucoma is one of the leading causes of irreversible vision loss in the world. Due to the aging world population, the WHO estimates that the number of people affected by glaucoma disease may increase to almost 80 million by 2020 [13]. Glaucoma progression is characterized by increase of optic cup area in color fundus images. Our work aims to develop a learning based algorithm using convolutional neural networks (CNN) to segment the optic disc (OD) and optic cup (OC) from retinal fundus images.

There exist numerous approaches for automatic optic cup and disc segmentation such as morphological features [1] and active contours [10]. Their performance depends upon contour initialization and ability to identify weak edges. Machine learning (ML) methods [4] have gained importance as they provide a powerful tool for feature classification. Success of ML methods depends on carefully hand designed features. However, hand crafted features limit their applicability to different datasets. This work proposes to learn the most discriminative features for OC and OD segmentation in the form of convolutional filters.

Mayraz and Hinton [12] proposed a hierarchical learning procedure based on a probabilistic learning framework called the product of experts [3], where the probability of an image is described by the normalized product of learned individual distributions. Another approach that also employs a hierarchical network and was evaluated on medical images is used in [11], where a CNN is learned from multiple scales by optimizing a 2-norm orthogonal matching pursuit problem.

X. Chen, M. K. Garvin, J. Liu, E. Trucco, Y. Xu (Eds.): OMIA 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, Iowa Research Online, pp. 153–160, 2015. Available from: http://ir.uiowa.edu/omia/2015_Proceedings/2015/

Ciresan et al. [5] used a Deep Neural Network (DNN) to segment neuronal structures in electron microscopy (EM) images and significantly outperform state of the art. Turaga et al. [17] segment neuronal structures in EM images by learning an affinity graph using a CNN. Our work proposes a novel CNN architecture for OC and OD segmentation without the need to define hand crafted features. This improves the algorithm's generalization ability. The primary contributions of this paper are: 1) a novel sampling strategy is introduced to identify landmarks that provide high information content to train our CNN architecture; and 2) a boosting framework is introduced to learn convolutional filters in our CNN architecture.

2 Method

Preprocessing: Each image is cropped with the optic disc or cup relatively central to the image and some background pixels around the OD and OC. This allows the algorithm to capture the essential characteristics of the image while focusing on the OC and OD. All images are downsampled by a factor of 4 to reduce computation complexity. The RGB images are converted to L*a*b color space. The intensity mean is subtracted from all pixel values and divided by the standard deviation. The intensities are then normalized to $[0, 1]$. Figures 1 (a),(b) show an example original image and the normalized image after preprocessing.

Entropy sampling: Entropy maps are calculated for each color channel using entropy filtering [8] and highly informative points selected to reduce the computational. They are used to training our CNN architecture. Figure 1 (c) shows the entropy filtered output of the original image. The informative points have higher value in the image.

2.1 Boosting convolutional filters:

Figure 1 (d) illustrates our method's architecture. Convolutional networks are composed of individual convolutional filters which can be regarded as classifiers in an ensemble. In that sense, the question may be posed how one could learn such filters in a principled manner. Different principled methods exist to arrive at an ensemble classifier of which the most popular are Boosting [7] and Bagging [2]. Kiros et al. [11] proposed to learn a generative convolutional network with two layers through bagging and solving an optimization problem for each filter using orthogonal matching pursuit. The key requirement is successive filters need to be orthogonal to previously learned filters. In contrast, in this work we propose a more direct way of exploring different filters through boosting to learn a discriminative convolutional network.

As a first step, 3×3 patches around each sampled point are extracted. To ensure that the filters do not need to learn superfluous patterns, such as similar patterns with different magnitude, all patches are subjected to *Local Contrast Normalization* as done in [11]. To this end, each patch is reshaped into a vector

x_{patch} and divided by the l_2-norm of the patch. The mean of the patch vector is then subtracted.

Using the extracted patch data, the following optimization problem is solved for each filter individually

$$\underset{w}{\text{minimize}} \quad \sum_{i=1}^{N} v_i \cdot |y_i - x_i w| \tag{1}$$

where $y_i \in Y = \{-1, +1\}$ is the label of a given training point i, $x_i \in X_{patch}$ represents the corresponding patch around point i. w is the convolutional filter in vector form that is to be learned, and v_i are the positive weights on an individual data point. The CVX optimization environment [9] was used. The architecture specifications can be summarized as:

1. Filters of size $3 \times 3 \times n_{maps}$ are trained, where n_{maps} corresponds to the number of channels of each input image, e.g. three for a regular RGB image.
2. 500 points are sampled to learn convolutional filters for each scale.
3. Filters are learned for five scales in the first layer and four scales in the second layer. This gives the local algorithm (3 × 3 filters) a more global understanding.
4. Two layers of convolutional filters are implemented. For the optic disc, five filters per scale are learned in the first layer and one filter per scale in the second layer. For the optic cup, six filters per scale are learned in the first layer and one filter per scale in the second layer. This makes for a total of 29 filters for the optic disc and 34 filters learned for the optic cup segmentation. Since optic cup segmentation is more challenging, more filters lead to better segmentation

Exploration of different filters is done through reweighting of data points based on *Gentle AdaBoost* [6] as it generalizes better by avoiding overfitting. The following reweighting is performed:

Initialize the weights as $v_i = \frac{1}{m}$, for $i = 1, \ldots, m$. For $n = 1, \ldots N$:

1. Estimate the "weak" hypothesis $h_n(x)$, i.e. learn filter w and bias b in the optimization problem 1.
2. Update weights
$$v_i \leftarrow \frac{v_i \cdot exp(-y_i h_n(\mathbf{x_i}))}{Z_n} \tag{2}$$
with Z_n chosen so that $\sum_{i=1}^{m} v_i = 1$.

α is always set to 1 and is determined by the error ϵ of the individual classifier, where a highly accurate classifier yields a high α factor and an inaccurate classifier yields low α factor as can be inferred from the equation defining α in this paragraph. The output of the convolution of each filter is passed through a *tanh* saturation function. As previously done for preprocessing, the mean of the image is subtracted and all values are divided by the standard deviation. Again, values are rescaled to lie in the range $[0, 1]$. These convoluted images are then passed through a max-pooling operation [15] to introduce further robustness into the system. In a second layer, the stacked output maps of the filters learned in the first layer are used as input. This second layer can be regarded as an extended ensemble classifier. The ensemble classifier is extended in the sense that not only values of one point for multiple individual classifiers are treated but the patch around these points as well. This reads as

$$H_{conv}(x) = \sum_{i=1}^{K} \sum_{p \in patch} w_{i,p} h_i(p) \tag{3}$$

where $H_{conv}(x)$ is a convolutional filter of the second layer (or any further layers), p denotes the positions of points in the patch around point x, $w_{i,p}$ are the learned weights for the convolutional filter in layer two and $h_i(p)$ describes the processed output of convolutional filter i of the previous layer. The processed output of the convolution of filters with the input images provide a good impression of which characteristics of the image a filter is focusing on. Figure 1 e) shows the output of the first filter in the first layer, while Figure 1 f) shows the output of the first filter in the second layer, i.e. an extended ensemble filter. Figure 2 (a) shows a subset of the learned filters for OC and OD.

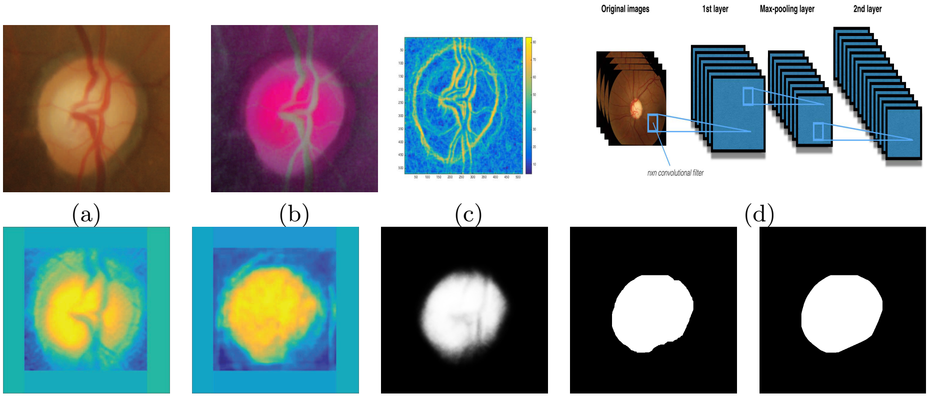
2.2 Classification

A classifier is trained on the extracted features of all sampled points, which are the output of each convolutional filter, the color values at these points and a "centricity" score. To extract the output of each convolutional filter, filters at smaller scales are upsampled to the original image size. For each image, 5000 points are sampled for which the L*a*b colors as well as the output of each convolutional filter at these points are extracted. Additionally, a "centricity number" of the sampled points is extracted which is meant to be a value measuring the "radius" from the center of the optic disc. The centricity is calculated by finding the weighted centroid c of the maximum intensity region of the L-color in L*a*b color space and calculating the outward radius from this point for a given sampled point as

$$C = \frac{(p_x - c_x)^2}{l_x} + \frac{(p_y - c_y)^2}{l_y} \tag{4}$$

where l_x, l_y is the width and height of a the given image, p is the position of the sampled point and c denotes the centroid's position. These are the features on which subsequently a softmax logistic regression classifier is trained as in [11]. The resulting probability map is shown in Figure 1 g).

Postprocessing: After classification, an unsupervised graph cut algorithm [14] is applied to the probability map in the previous subsection to smooth the results as demonstrated on the sample image in Figure 1 h). A convex hull transform is applied to the graph cut output. Given the oval shape of both the optic disc and cup it is apriori known that a convex shape is to be detected. Taking the convex hull of the graph cut output unites previously disjoint regions that



(a) (b) (c) (d)

(e) (f) (g) (h) (i)

Fig. 1. (a)Example original image; (b) image after preprocessing and normalization of (a); (c) entropy filtered output; (d) illustration of CNN architecture; (e) output of the first filter in the first layer; (f) output of the first filter in the second layer; (g) probability map from logistic regression classifier; (h) graph cut segmentation; (i) final output after convex hull fitting.

all belong to the optic disc or cup. The improved segmentation is demonstrated in Figure 1 i).

3 Experimental results

Our proposed method for optic disc and cup segmentation was validated on the DRISHTI-GS dataset [16] which consists of 50 patient images obtained using 30 degree FOV at a resolution of 2896 × 1944. We use a 5 fold cross validation scheme with 40 training images and 10 test images in each fold. The ground truth disc and cup segmentation masks were obtained by a majority voting of manual markings by 4 ophthalmologists. Quantitative evaluation is based on F-score ($F = 2\ P \times R/(P + R)$) to measure the extent of region overlap and absolute pointwise localization error B in pixels (measured in the radial direction); P is precision and R is recall. Additionally we report the overlap measure $S = Area(M \cap A)/Area(M \cup A)$. M is the manual segmentation while A is the algorithm segmentation. Our whole pipeline was implemented in MATLAB on a 2.66 GHz quad core CPU running Windows 7.

3.1 Segmentation Performance

Table 1 summarizes the segmentation performance of different methods. Our proposed method, CNN, outperforms all the competing methods as is evident from the higher F and S values, and lower B values. The difference is also statistically significant since $p < 0.01$ (from Student-t tests) for all methods

	Optic Disc					Optic Cup				
	CNN	[4]	[18]	[10]	[16]	CNN	[4]	[18]	[10]	[16]
F	94.7	93.0	92.2	90.8	95.0	83.0	80.8	78.4	78.9	80.7
S	89.5	87.3	86.8	85.0	85.2	86.4	82.1	80.1	82.5	84.2
B	9.1	9.4	9.9	12.1	11.7	16.5	19.3	20.6	17.2	16.2

Table 1. Segmentation performance for OC and OD segmentation using different methods.

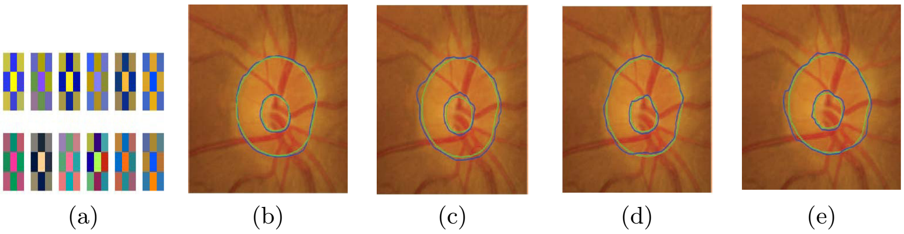


Fig. 2. (a) Subset learned filter for OD (top row) and OC (bottom row). Segmentation results for different methods: (b) our proposed CNN model; (c) integrated disc and cup segmentation method of [18]; (d) superpixel segmentation method of [4] and ; (e) [10]

compared to CNN. Segmenting the optic cup is more challenging than the disc due to absence of distinguishing depth information. While pallor is one factor, it is not always reliable due to similar intensity profiles of neighboring regions. CNN obtains high segmentation accuracy than hand crafted features by learning image priors for the cup region.

Since [4] is a superpixel based approach, pixels from different classes may be grouped in one superpixel which affects its performance. [10] uses a modified Chan-Vese model, which finds it challenging to segment the optic disc using only intensity information. [1] uses only morphological features which is good enough for disc segmentation, but does not perform as well for cup segmentation. However CNN outperforms all these methods. Figure 2 shows the comparative results of FoE and the combined disc and cup segmentation methods of [4],[18] and [10].

For optic disc segmentation all the methods perform almost at the same level since OD is much easy to segment. However CNN's advantages are prominent for cup segmentation. CNN outperforms the state-of-the-art approaches tested by Sivaswamy et al. [16] which achieve a maximal F-score of 0.80 on the training set. On the other hand CNN achieves a F-score of 0.83 with a smaller standard deviation than the other methods. CNN also has lower boundary localization error than other competing methods. We also show the best and worst case results for optic disc (Figure 3) and optic cup (Figure 4).

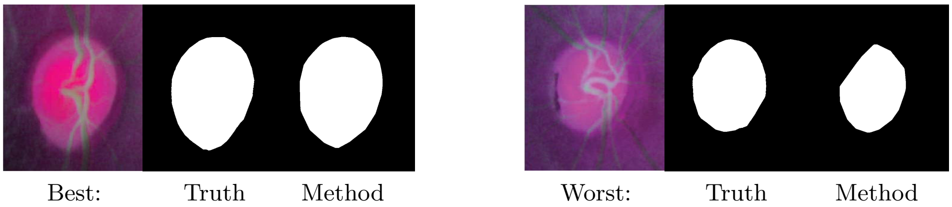


Fig. 3. Results of best/worst case segmentation for optic disc, respectively

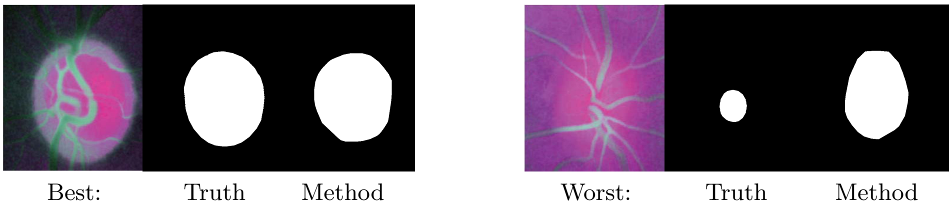


Fig. 4. Results of best/worst case segmentation for optic cup, respectively

4 Discussion and Conclusion

This paper introduced a novel entropy sampling method within a CNN architecture. Building upon this technique, an original framework for learning convolutional filters in a principled manner using boosting was described. The boosted network of convolutional filters was shown to outperform existing methods on the DRISHTI-GS data set. Entropy sampling competently finds more relevant points than uniform sampling. Boosting convolutional filters is able to learn very discriminative convolutional filters even for small data sets. Our proposed CNN architecture, helped by entropy sampling, focuses to learn discriminative features from more relevant points.

References

1. A.Aquino, Gegundez-Arias, M., Marin., D.: Detecting the optic disc boundary in digital fundus images using morphological edge detection and feature extraction techniques. IEEE Trans. Med. Imag 20(11), 1860–1869 (2010)
2. Breiman, L.: Bagging predictors. In: Machine Learning. pp. 123–140 (1996)
3. Brown, A., Hinton, G.: Products of hidden markov models. Technical report GCNU TR 2000-008, Gatsby Computational Neuroscience Unit, University College London (November 2000)
4. Cheng, J., Liu, J., et. al.: Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans. Med. Imag 32(6), 1019–1032 (2013)
5. Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (eds.) Advances in Neural Information Processing Systems 25, pp. 2843–2851. Curran Associates, Inc. (2012)

6. Doğan, H., Akay, O.: Using adaboost classifiers in a hierarchical framework for classifying surface images of marble slabs. Expert Syst. Appl. 37(12), 8814–8821 (Dec 2010)
7. Freund, Y., Schapire, R.E.: A short introduction to boosting (1999)
8. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing Using MATLAB. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (2003)
9. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx (Mar 2014)
10. Joshi, G., Sivaswamy, J., Krishnadas, S.: Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans. Med. Imag 30(6), 1192–1205 (2011)
11. Kiros, R., Popuri, K., Cobzas, D., Jagersand, M.: Stacked multiscale feature learning for domain independent medical image segmentation. In: Wu, G., Zhang, D., Zhou, L. (eds.) Machine Learning in Medical Imaging, Lecture Notes in Computer Science, vol. 8679, pp. 25–32. Springer International Publishing (2014)
12. Mayraz, G., Hinton, G.: Recognizing handwritten digits using hierarchical products of experts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(2), 189–197 (Feb 2002)
13. Organization, W.H.: Vision 2020. the right to sight. global initiative for the elimination of avoidable blindness. WHO Press (2006)
14. Salah, M., Mitiche, A., Ayed, I.: Multiregion image segmentation by parametric kernel graph cuts. Image Processing, IEEE Transactions on 20(2), 545–557 (Feb 2011)
15. Scherer, D., Müller, A., Behnke, S.: Evaluation of pooling operations in convolutional architectures for object recognition. In: Proceedings of the 20th International Conference on Artificial Neural Networks: Part III. pp. 92–101. ICANN'10, Springer-Verlag, Berlin, Heidelberg (2010)
16. Sivaswamy, J., Krishnadas, S., Datt Joshi, G., Jain, M., Syed Tabish, A.: Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation. In: Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on. pp. 53–56 (April 2014)
17. Turaga, S.C., Murray, J.F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., Seung, H.S.: Convolutional networks can learn to generate affinity graphs for image segmentation. Neural Computation 22(2), 511–538 (2015/05/17 2009)
18. Wong, D., et. al.: Level set based automatic cup to disc ratio determination using retinal fundus images in argali. In: Proc. IEEE EMBC. pp. 2266–2269 (2008)