Ensemble of Deep Convolutional Neural Networks for Learning to Detect Retinal Vessels in Fundus Images

Debapriya Maji\textsuperscript{1}, Anirban Santara\textsuperscript{2}, Pabitra Mitra\textsuperscript{2}, Debdoit Sheet\textsuperscript{2}

Abstract—Vision impairment due to pathological damage of the retina can largely be prevented through periodic screening using fundus color imaging. However the challenge with large scale screening is the inability to exhaustively detect fine blood vessels crucial to disease diagnosis. In this work we present a computational imaging framework using deep and ensemble learning for reliable detection of blood vessels in fundus color images. An ensemble of deep convolutional neural networks is trained to segment vessel and non-vessel areas of a color fundus image. During inference, the responses of the individual ConvNets of the ensemble are averaged to form the final segmentation. In experimental evaluation with the DRIVE database, we achieve the objective of vessel detection with maximum average accuracy of 94.7% and area under ROC curve of 0.9283.

Index Terms—Computational imaging, deep learning, convolutional neural network, ensemble learning, vessel detection.

I. INTRODUCTION

Pathological conditions of the retina examined through regular screening [1], [2] can heavily assist prevention of visual blindness. Fundus imaging is the most widely used modality for early screening and detection of such blindness causing diseases like diabetic retinopathy, glaucoma, age-related macular degeneration [3], hypertension and stroke induced changes [4]. Imaging of fundus has largely improved with progress from the film based photography camera to use of electronic imaging sensors; as well as red free imaging, stereo photography, hyperspectral imaging, angiography, etc. [5], thereby reducing inter- and intra-observer reporting variability. Retinal image analysis has also significantly contributed to this technological development [5], [6]. Since fundus imaging is predominantly used for first level of abnormality screening, research focus includes: (i) detection and segmentation of retinal structures (vessels, fovea, optic disc), (ii) segmentation of abnormalities, and (iii) quality quantification of images acquired to assess reporting fitness [5].

Related Work: The process of clinical reporting of retinal abnormalities is systematic and lesions are reported with respect to their location from vessels or optic disc. Computer assisted diagnosis systems are accordingly being developed to improve the clinical workflow [5]. Some of the developments include, assessment of image quality [8]; blood vessel detection [7], branch pattern, diameter and vascular tree analysis [3], [4], [9], [10]; followed by reporting of lesions and their location with respect to the vessels [4], [10]. An important challenge in this context is robust and exhaustive detection of retinal vessels in color fundus images leveraging the potential of computer assisted diagnosis [5], [11] and assist in routine screening [3], [4].

Challenge: Methods for vessel detection and segmentation [5]–[7], [9], [10], [12] predominantly use image filters, vector geometry, statistical distribution studies, and machine learning of low-level features and photon distribution models for vessel detection. Such methods rely on use of handcrafted features or heuristic assumptions for solving the problem and are not generalized to learn pattern attributes from the data itself, thus making them vulnerable to performance subjectivity on account of the method’s inherent weaknesses. Recently fully data-driven deep learning based models have been proposed [13]. However, they are weaker in performance compared to the state of the art methods that use the conventional paradigm. The primary challenge here is to design an end-to-end framework which learns pattern representation from the data without any domain knowledge based heuristic information to identify both coarse and fine vascular structures and is at least at par if not better than the heuristic-based models.

Approach: This paper makes an attempt to ameliorate the issue of subjectivity induced bias in feature representation by training an ensemble of 12 Convolutional Neural Networks (ConvNets) [14] on raw color fundus images to discriminate

\textsuperscript{1} D. Maji is with Texas Instruments Inc.
\textsuperscript{2} A. Santara, P. Mitra and D. Sheet are with the Indian Institute of Technology Kharagpur, Kharagpur, WB 721302, India. anirban_santara@iitkgp.ac.in
* equal contribution
There are four kinds of layers in a typical ConvNet like structure [14], [15]. The ConvNet architecture is based on neural networks designed for processing data with a grid-like structure [14], [15]. The ConvNet architecture is based on sparse interactions and parameter sharing and is highly effective for efficient learning of spatial invariances in images [16], [17]. There are four kinds of layers in a typical ConvNet architecture: convolutional (conv), pooling (pool), fully-connected (affine) and rectifying linear unit (ReLU). Each convolutional layer transforms one set of feature maps into another set of feature maps by convolution with a set of filters. Mathematically, if \(W_i^l\) and \(b_i^l\) denote the weights and the bias of the \(i^{th}\) filter of the \(l^{th}\) convolutional layer and \(H_i^l\) be its activation-map, then:

\[
H_i^l = H_i^{l-1} \otimes W_i^l + b_i^l
\]

where \(\otimes\) is the convolution operator. Pooling layers perform a spatial downsampling of the input feature maps. Pooling helps to make the representation become invariant to small translations of the input. Fully-connected layers are similar to the layers in a vanilla neural network. Let \(W^l\) denote the incoming weight matrix and \(b^l\) the bias vector of a fully-connected layer, \(l\). Then:

\[
H^l = \text{flatten}(H^{l-1}) \ast W^l \oplus b^l
\]

where \(\text{flatten(.)}\) operator tiles the feature-maps of the input volume along the height, \(\ast\) is matrix multiplication and \(\oplus\) is element-wise addition. ReLU layers perform a pointwise rectification of the input and correspond to the activation function. For the \(i^{th}\) unit of layer \(l\):

\[
H_i^l = \max(H_i^{l-1}, 0)
\]

In a deep ConvNet, units in the deeper layers indirectly interact with a larger area of the input, thus forming a high level abstraction of the input data.

**Ensemble learning:** Ensemble learning is a technique of using multiple models or experts for solving a particular artificial intelligence problem [18]. Ensemble methods seek to promote diversity among the models they combine and reduce the problem related to overfitting of the training data. The outputs of the individual models of the ensemble are combined (e.g. by averaging) to form the final prediction. Concretely, if \(\{m_1, \ldots, m_k\}\) be \(k\) models of an ensemble and \(p(x = y_i|m_j)\) is the probability that the input \(x\) is classified as \(y_i\) under the model \(m_j\), then the ensemble predicts:

\[
p(x = y_i|m_1, \ldots, m_k) = \frac{1}{k} \sum_{j=1}^{k} p(x = y_i|m_j)
\]

Ensemble learning promotes better generalization and often provides higher accuracy of prediction than the individual models.

**Problem Statement**

Let \(I\) be an image acquired by the RGB sensor of a color fundus camera. The intensity observed at location \(x\) is denoted by \(g(x)\). Let \(N(x)\) be a set of pixels in the local neighborhood of \(x\). Let \(\omega = \{\text{vessel, non-vascular}\}\) be the set of class labels for the pixel at location \(x\). In a machine learning framework, the probability of finding a tissue of type \(\omega\) at location \(x\), \(p(\omega|I, x)\) is modelled by a class of functions \(\mathcal{H}(\omega|I, N(x), x, \theta_1, \theta_2)\) where \(\theta_1\) is a set of parameters which are learned from the training data and \(\theta_2\), a set of hyperparameters tuned using the validation data. In the proposed method, \(\mathcal{H}(\cdot)\) is an ensemble of ConvNets, the architecture of which is described next.
V. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents experimental validation of the proposed technique and its performance comparison with earlier methods [11].

Dataset: The ensemble of ConvNets is is evaluated by learning with the DRIVE training set (image id. 21-40) and testing over the DRIVE test set (image id. 1-20)

Learning mechanism: Each ConvNet is trained independently on a set of 60000 randomly chosen $3 \times 3 \times 3$ patches. Learning rate and annealing rate were kept constant across models at $5e-4$ and 0.95 respectively. Dropout probability, $L_2$ regularization coefficient and number of hidden units in the penultimate affine layer of the different models were sampled respectively from $U([0.5, 0.9])$, $U([1e-3, 2.5e-3])$ and $U([128, 256, 512])$ where $U(.)$ denotes uniform probability distribution over a given range. The models were trained using RMSProp algorithm [20] with minibatch size 200.

Performance assessment: Table 1 presents the accuracy and consistency of detection in comparison with those reported in earlier techniques. It is clearly evident that although our approach does not have the highest accuracy as compared with other methods, it does exhibit superior performance than the previously proposed deep learning based method for learning vessel representations from data [13]. The kappa score being a study of observer consistency indicates sensitivity of the technique to detect both coarse and fine vessels as desired. Typical response to detection of both coarse and fine vessels are presented in Fig. 3.

Fig. 4 gives the receiver operating characteristic (ROC) curve of the proposed method. Area under ROC curve obtained is 0.9283.

VI. CONCLUSION

This paper presents a ConvNet-ensemble based framework for processing color fundus images for detection of coarse and fine vessels. The method is evaluated experimentally on the DRIVE dataset. The remarkable ability of ConvNets to recognize images and that of ensemble learning at generalization is leveraged to design a heuristics independent, data driven approach to analyzing medical images. This presents a feasible solution to subjectivity induced bias in medical

1 DRIVE dataset: [http://www.isi.uu.nl/Research/Databases/DRIVE/](http://www.isi.uu.nl/Research/Databases/DRIVE/)
image analysis. This is an improvement of our previous work on data-driven analysis of fundus images [13]. This approach in general also provides a strong alternative approach to solve complex medical data analysis problem through deep learning combined with the power of ensemble learning.

REFERENCES

[1] A. Tuulonen, P. J. Airaksinen, A. Montagna, and H. Nieminen, “Screening for glaucoma with a non-mydriatic fundus camera,” Acta Ophthalmol., vol. 68, no. 4, pp. 445–449, 1990.

[2] E. Stefansson, T. Bek, M. Porta, N. Larsen, J. K. Kristinsson, and E. Agardh, “Screening and prevention of diabetic blindness,” Acta Ophthalmol., vol. 78, no. 4, pp. 374–385, 2000.

[3] C. Heneghan, J. Flynn, M. O-Keefe, and M. Cahill, “Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis,” Medical Image Analysis, vol. 6, no. 4, pp. 407 – 429, 2002.

[4] T.-Y. Wong, R. Klein, B. E. K. Klein, J. M. Tielsch, L. Hubbard, and F. J. Nieto, “Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease, and mortality,” Survey Ophthalm., vol. 46, no. 1, pp. 59 – 80, 2001.

[5] M.D. Abrámoff, M.K. Garvin, and M. Sonka, “Retinal imaging and image analysis,” IEEE Rev. Biomed. Engg., vol. 3, pp. 169 –208, 2010.

[6] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable, “Retinal image analysis: Concepts, applications and potential,” Progress in Retinal and Eye Research, vol. 25, no. 1, pp. 99 – 127, 2006.

[7] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” IEEE Trans. Med. Imaging, vol. 23, no. 4, pp. 501 –509, Apr. 2004.

[8] M.E. Gugundez-Arias, A. Aquino, J.M. Bravo, and D. Marin, “A function for quality evaluation of retinal vessel segmentations,” IEEE Trans. Med. Imaging, vol. 31, no. 2, pp. 231 –239, Feb. 2012.

[9] M. Sofka and C.V. Stewart, “Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures,” IEEE Trans. Med. Imaging, vol. 25, no. 12, pp. 1531 –1546, Dec. 2006.

[10] J. Jan, J. Odstrcilik, J. Gazarek, and R. Kolar, “Retinal image analysis aimed at blood vessel tree segmentation and early detection of neural-layer deterioration,” Comput. Med. Imaging and Graphics, vol. 36, no. 6, pp. 431 – 441, 2012.

[11] 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 SPIE Medical Imaging. Proc. SPIE, 2004, vol. 5370, pp. 648–656.