Detection of Neovascularization in Retinal Images using Semi-Supervised Learning

by

Pujitha Appan K., Jahnavi Gamalapati S., Jayanthi Sivaswamy

in

IEEE International Symposium on Biomedical Imaging 2017
(*ISBI-2017*)

Melbourne, Australia

Report No: IIIT/TR/2017/-1

Centre for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
April 2017

DETECTION OF NEOVASCULARIZATION IN RETINAL IMAGES USING SEMI-SUPERVISED LEARNING

Appan K. Pujitha, Gamalapati S. Jahnavi* and Jayanthi Sivaswamy*

Center for Visual Information Technology, IIIT Hyderabad, India

ABSTRACT

Retinal Neovascularization (NV) is a critical stage of Diabetic Retinopathy (DR) and its detection is important to prevent blindness. Existing fully supervised frameworks typically take a patch-based approach and report good results only on limited number of images due to sparsity of annotated data. We propose a patch-based semi-supervised framework which paves the way for including unlabeled data in training. In this framework, NV patches are modeled using oriented energy and vesselness based features. These features are fused within a co-training based semi-supervised framework by using neighborhood information in feature space. Rule-based criteria on patch-level neovascularity scores is used to derive the final image-level decision. The proposed approach was evaluated on 1 private and 3 public datasets, both at patch and image level detection on nearly 200,000 patches. An AUC of 0.985 with sensitivity of 96.2% at specificity of 92.6% was obtained for abnormality detection at patch-level, while at the image-level, a sensitivity of 96.76% at a specificity of 91.85% were obtained. The achieved performance on a large number of patches indicates the robustness of our approach.

***Index Terms*—** Neovascularization; Gradient Boost; Co-training; Diabetic Retinopathy; Consistency in feature space.

1. INTRODUCTION

Advanced stages of Diabetic Retinopathy (DR) is marked by the formation of new, weak and thin microvascular networks, a phenomenon known as Neovascularization (NV). NV increases the risk of fibrosis, bleeding and ultimately loss of vision and its detection is therefore of interest. However, existing literature is predominantly devoted to detection of lesions that arise in the earlier stage of DR.

Existing approaches for NV detection typically follow a classification route, where the features are either extracted from the segmented vessel map [1] [2] [3] or directly from the raw image [4] [5] [6] [7] [8]. A wide variety of features and their combinations have been used in both approaches. These are extracted in all but one method [4] at a patch level. Features considered in the first approach include a combination of shape, intensity and gradient features [1]; multi-scale AM-FM [2]; shape, position, orientation, intensity and line density [3]. In the second approach, features that have been explored include morphological [4], morphological and GLCM [5], multiscale AM-FM [6]; vesselness, power spectrum distribution [7]; LBP and multi-scale Counterlet transform [8]. The final classification is done using LDA [1], SVM [2] [3] [4] , hierarchical clustering [5], random forest classifier [6] [7] and Artificial Neural Network [8].

Accurate vessel segmentation, specifically of thin vessels, is critical to the first type of approach while adequate training data and hand crafting of features is required for both approaches. Large public access datasets for NV detection provide image-level annotation whereas the existing detection methods require patch level labels whose generation is laborious and hence not scalable. Consequently, most methods depend on locally sourced annotations and report only on a selected set of images from public datasets. We propose a patch-based NV detection method which i) uses generic features thereby eliminating the need for hand crafting and ii) leverages the availability of large amount of unlabelled data by employing a co-training based semi-supervised framework. We show that co-training with generic features such as vesselness and oriented local energy leads to consistently good performance on nearly 3000 images from 4 public datasets.

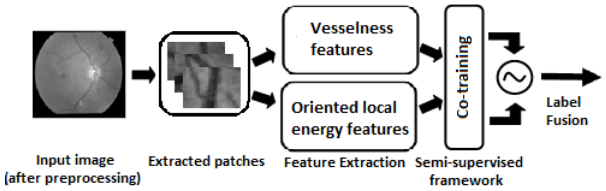


Fig. 1: Proposed system for NV detection

2. METHOD

The proposed method consists of four stages: pre-processing, feature extraction, co-training and label fusion as shown in Fig. 1. All processing was restricted to the green plane of the given image. The stages are described in detail next. As Retinal images suffer from non-uniform illumination, this is

* Equal contribution

This work was supported by the Dept. of Electronics and Information Technology, Govt. of India under Grant: DeitY/R&D/TDC/13(8)/2013

corrected using luminosity and contrast normalization [9] as a pre-processing step.

2.1. Feature extraction

Co-training requires two independent feature spaces to represent a patch. These features also have to be discriminative for NV and Non-NV patches. Two types of features are chosen to satisfy these requirements: (i) A Hessian based vesselness feature which is popular for vessels and (ii) A more generic one for oriented structures to capture coarse texture.

2.1.1. Vesselness based features

The vesselness is computed on the basis of eigenvalues of the Hessian as described in [10]. The probability of an image region to contain vessels or other ridges is found based on the eigenvalues and the vesselness at every point X: (x,y) in a patch at a scale s is computed as:

$$v(X,s) = \begin{Bmatrix} 0 & if \quad \lambda_2 > 0, \\ exp(-\frac{\lambda_1^2}{2\lambda_2^2\alpha^2})(1-exp(-\frac{\lambda_1^2+\lambda_2^2}{2c^2})) & o/w \end{Bmatrix} \quad (1)$$

Here, α and c are thresholds; $\lambda_i; i \in \{1,2\}$, are eigenvalues of the Hessian matrix computed at X.

The final vesselness map is obtained by taking the maximum response across all scales. A sample NV patch and the derived vessel map are shown in Fig. 2(a,b). This map is row vectorized to form the feature vector g_{vm}.

2.1.2. Oriented local energy features (OLE)

NV is characterized by texture at multiple scales and orientations. This texture can be represented by local energy which is defined as a sum of squared responses of a pair of conjugate symmetric filters. A Gabor filter bank is a standard way to compute local energy as it aids determining responses at different orientation and scales. We choose the log Gabor kernel in the frequency (f) domain:

$$K(f,\theta) = exp(-\frac{(log(\frac{f}{f_0}))^2}{2(log(\frac{\sigma_f}{f_0}))^2})exp(-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}) \quad (2)$$

Here f_0 is the central radial frequency, θ_0 is the orientation of the filter, σ_θ and σ_f are the angular and radial bandwidths.

The OLE at every point X in the patch is computed as:

$$E_{\theta_0}^{f_0}(x,y) = \sqrt{(R_{\theta_0}^{f_0,even}(x,y))^2 + (R_{\theta_0}^{f_0,odd}(x,y))^2} \quad (3)$$

Here $R_{\theta_0}^{f_0,even}$ and $R_{\theta_0}^{f_0,odd}$ are the responses of even and odd symmetric log gabor filters. The total energy of entire patch of size m xn, at a specific θ_0 is found as:

$$\hat{E}(\theta_0,f_0) = \sum_{x=1}^{m}\sum_{y=1}^{n} E_{\theta_0}^{f_0}(x,y) \quad (4)$$

This can be considered as the histogram function of energy map which expresses the oriented energy at different scales. The histogram is normalized by dividing it by maximum energy over all orientations at particular scale. The final feature vector g_{ole} is derived by concatenating the energy histogram at different orientations and scales. Fig. 2 shows sample Non-NV (d) and NV patches (e) and their energy histograms (c). The two types of patches are clearly distinguishable with the energy plots for NV patches (in red) having higher energy on average.

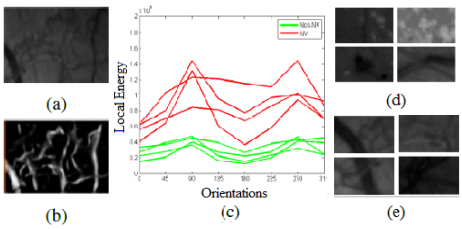


Fig. 2: Feature representation for patches. (a) NV patch and its (b) vesselness map; (c) Oriented local energy histogram for (d) Non-NV and (e) NV patches.

2.2. Co-training

Availability of annotations is limited in the medical domain as it is tedious to generate them and training with this limited data has the potential to over-fit and lacks robustness. Hence, the proposed method uses co-training [11] to utilise unlabeled images which are more widely available. Co-training works best with independent features. Table.1 shows the normalized values obtained after applying PCA on the features extracted from sample patches. The sample patches consists of all possible combinations of following structures relevant to DR: vessels, NV, hemorrhages, hard exudates and background.The last three are irrelevant to NV detection and can be seen to be marked by low values of g_{vm}. Our feature choice satisfies the requirement for co-training as, the pure vessel (denoted as V) and NV (denoted as NV+V) patches are adequately separated from each other, as well as from other patches.

Table 1: Feature Space Representation Values

	NV+V	**V**	**HEM+V**	**HE+V**	**HEM**	**HE**	**BG**
g_{vm}	1	0.8	0.65	0.25	0.02	0.01	0.01
g_{ole}	0.9	0.02	0.7	0.17	0.37	0.7	0.02

Hence, OLE and vesselness based features can be used to independently predict the confidence score for a patch with NV. Two classifiers C_1 and C_2 are trained separately on these features using gradient boosting [12] which is a way to design a strong classifier by fusing weak classifiers. The resultant strong classifier is used to predict the class probabilities.

In the co-training framework, let L be a set of labeled data (-1 for Non-NV and +1 for NV patches) and U be the set of

unlabeled data. A classifier h_1 is trained using g_{vm} and classifier h_2 is trained using g_{ole}. The trained classifier h_1 is used to label the unlabeled data. The most confidently predicted patches (p positive and n negative) are fed back to the training set to update h_2. Similarly, the most confidently predicted patches by h_2 are fed back to update h_1. This process is repeated until all the unlabeled patches are labeled.

2.3. Label fusion

The labels obtained from co-training are fused using nearest neighbour information in feature space . A rule-based voting system is considered for fusion as described. In a test patch, for each feature, the K nearest features are extracted from the updated training set. Each retrieved similar feature, along with its labels and weights, contributes a vote. The weights are computed as :

$$\widehat{w}_{g_i} = exp(\frac{-d(x_{g_i}, \widehat{y}_{Kg_i})}{\sigma^2}) \quad (5)$$

Here $g_i \in \{g_{vm}, g_{ole}\}$, $d(x_{g_i}, \widehat{y}_{Kg_i})$ is the Euclidean distance between the test feature x_{g_i} and K nearest training features $\widehat{y}_{Kg_i}$. The aggregation of votes results in a probabilistic decision as given:

$$P(x, y_K) = \frac{\widehat{w}_{Kg_{vm}} \widehat{l}_{Kg_{vm}} + \widehat{w}_{Kg_{ole}} \widehat{l}_{Kg_{ole}}}{\sum_{k=1}^{K} (\widehat{w}_{kg_{vm}} \widehat{l}_{kg_{vm}} + \widehat{w}_{kg_{ole}} \widehat{l}_{kg_{ole}})} \quad (6)$$

Here $\widehat{l}_{Kg_i}$ denotes K training labels of feature g_i. The probabilities obtained are thresholded to get the final class labels.

3. DATASETS

Four datasets (1 private and 3 public) are considered for the evaluation of the proposed method. KPHDR [8] is a private dataset with local annotations (only for NV). The patch level ground truth needed for evaluation was obtained such that atleast 30% of the patch was annotated as having NV. All 3 public datasets for DR , namely, MESSIDOR [13], Kaggle [14] (a challenge set) and DIARET-DB0 [15] provide only image-level annotations. MESSIDOR and Kaggle have 4-level annotations regarding DR severity and albeit different, which are roughly: No DR, Mild, Moderate, Severe and PDR. DIARET-DB0 provides annotations in terms of various abnormalities present in an image. Images from all datasets were re-sized to $\sim$0.8 times the given image size (maintaining the aspect ratio) prior to patch extraction for computational efficiency. A window of 100x150 dimensions was used to divide the image into patches with a stride of 75 pixels.

Training and Testing datasets: KPHDR has a total of 2575 abnormal patches. Training/testing sets were constructed for this dataset with 3000/55765 patches with 1:1 ratio of NV:Non-NV for training to address class imbalance problem. The data samples chosen fortraining, testing and unlabelled data are disjoint sets and the details are reported in Table. 2.

4. EXPERIMENTS AND RESULTS

4.1. Experiments

Log-Gabor filters were considered at 12 scales and 24 orientations as given in (2). Vesselness maps were computed at 5 scales with sigma varying between 1 to 10 (step size 2). Given an image patch, oriented local energy feature (of size 288) and vesselness based feature (of size 15000) are extracted to derive patch level predictions.

Training was done on features extracted from a set of 3000 patches (L) selected randomly from KPHDR dataset, 1:1 ratio of Non-NV and NV patches. The unlabeled (U) dataset of 4000 patches is taken from the remaining 3 datasets (1:1:1 ratio from each dataset). Gradient boosting for training was done with exponential loss function and decision trees were used as base-learners. The shrinkage factor which reduces the impact of potentially unstable regression coefficients was varied for different values and finally set to 0.1. For each iteration, the unlabeled dataset was tested using the obtained models, from which samples with top 10 confidence score of both models were used for updating the training set.

Table 2: Total number of data used for evaluation

Dataset	Training	Unlabelled $N_I : N_P$	Testing $N_I : N_P$
KPHDR	3000*	-	55765*
MESSIDOR	-	12 : 1296	469 : 46431
DIARET-DB0	-	7 : 1309	34 : 6358
Kaggle	-	29 : 1305	2081 : 93645
Total	3000*	48 : 3910	2584 : 202199

*number of patches; N_P denotes the number of patches extracted from N_I images

Image level decision of NV/Non-NV was done by considering patch level predictions on the entire image and thresholding the number of NV patches detected out of total patches in the image. Experimentally this threshold was determined to be 2%-4% of total number of patches.

4.2. Results :

The performance was assessed using the following evaluation metrics: Sensitivity (SN), Specificity (SP), Receiver Operating characteristics (ROC) and Area Under Curve (AUC). To underscore the contribution from co-training, the obtained values are reported without/with co-training (I/II) in Table. 3.

The tabulated results show that the average SN and SP values *without* co-training are comparable to other methods demonstrating the effectiveness of the selected features which are generic in nature. Overall, we can note that co-training results in an improvement of 5%-11% in performance.

The performance metrics reported by 4 other methods are listed in Table 4. As mentioned in the introduction, the testing results have been reported in literature not on the entire dataset, but on a selected number of patches/images as indicated in the table. Hence, a direct comparison is not possible.

Table 3: NV detection results for the proposed method, without (I) and with co-training (II) on 4 datasets.

Dataset	SN(%)		SP(%)		AUC	
	(I)	(II)	(I)	(II)	(I)	(II)
KPHDR**	89.8	96.2	92.7	92.6	0.9378	0.9850
MESSIDOR*	97.56	97.56	82.5	93.4	0.9659	0.9877
DIARET-DB0*	100	100	86.38	90.91	0.9605	0.9868
Kaggle*	92.91	92.73	74.8	91.25	0.9274	0.9725
Average*	96.82	96.76	81.22	91.85	0.951	0.982

**at patch level; *at Image level

The proposed method has SN/SP of 96.2/92.6% on KPHDR, which appears to be lower than the best results of 99.62/96.61% reported in [8]. However, the test set sizes for these results differ by several orders (200). Likewise, on the MESSIDOR dataset, the proposed method achieves a SN/SP of 97.56/93.4% against 98/97% reported in [1], however there is roughly a 5-fold difference in the test set sizes. Even though Kaggle has high inter-image variations, the proposed method is able to perform well. Thus, we can say that this method is robust as well as superlative in performance.

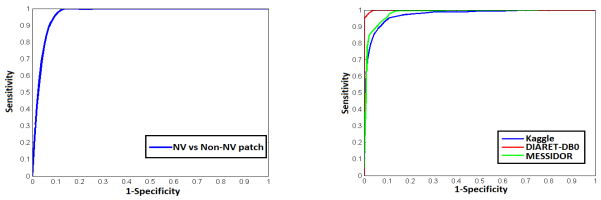
Table 4: NV detection results of existing methods

Method	Dataset	Test Samples	SN(%)	SP(%)	AUC
[16]	**KPHDR**	200 patches	94	85	0.92
[8]	**KPHDR**	322 patches	99.62	96.61	-
[7]	**MESSIDOR**	98 images	100	87	0.98
[1]	**MESSIDOR**	130 images	98	97	-

Patch level ROC was computed by varying the threshold on the probabilities obtained after label fusion and is shown in Fig. 3(a). Image-level NV detection was also analyzed and the ROC was derived by varying thresholds on the number of abnormal patches detected in a given image (see Fig. 3(b). The proposed method achieves an average AUC of 0.982 over 4 datasets which indicates consistency and robustness.

5. CONCLUSION

Automatic detection of NV is a difficult task as it is characterised by complex texture changes and learning is impeded by limited availability of annotated data. These constraints are overcome in our proposed method with the use of vesselness and Gabor features along with co-training. Co-training was seen to improve the AUC value from 0.95 to 0.98 which is significant. Further improvements are possible with appropriate selection of unlabeled data during co-training. Consistent good performance of the method across datasets (over nearly a quarter million patches), at both patch and image levels, depicts its robustness to changes in resolution, illumination, tissue type (due to population difference) and noisy conditions. These results demonstrate that the proposed system can be used in automated detection and grading of DR.



(a) (b)

Fig. 3: ROC curve for predictions at the (a) patch level (KPHDR) and (b) image level (other 3 datsets).

6. REFERENCES

[1] M.U. Akram et al., "Detection of neovascularization in retinal images using multivariate m-mediods based classifier," *Computerized Med. Imaging and Graphics*, vol. 2013, pp. 346–357, Jun 2013.

[2] C. Agurto et al., "A multiscale decomposition approach to detect abnormal vasculature in the optic disc," in *Computerized med. imaging and graphics*, July 2015, pp. 137–149.

[3] K. A. Goatman et al., "Detection of new vessels on the optic disc using retinal photographs," *IEEE Trans. Med. Imaging*, vol. 30, pp. 972–979, Apr 2011.

[4] U. Acharya et al., "Computer based detection of diabetes retinopathy stages using digital fundus images," *Journal of Engineering in Medicine*, pp. 502–512, Jun 2009.

[5] J. Nayak et al., "Automated identification of diabetic retinopathy stages using digital fundus images," *Jour. of Medical Systems*, vol. 32, pp. 107–115, Oct 2008.

[6] Agurto et al., "Multiscale am-fm methods for diabetic retinopathy lesion detection," *IEEE Trans. Med. Imaging*, vol. 29, pp. 502–512, Jun 2010.

[7] G. Gupta et al., "Computer-assisted identification of proliferative diabetic retinopathy in color retinal images," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 5642–5645.

[8] M. Vatanparast and A. Harati, "A feasibility study on detection of neovascularization in retinal color images using texture," in *Computer and Knowledge Engineering, 2012 International Conf.*, Oct 2012, pp. 221–226.

[9] G. D. Joshi and J. Sivaswamy, "Colour retinal image enhancement based on domain knowledge," in *Computer Vision, Graphics Image Processing, 2008. ICVGIP '08. Sixth Indian Conference on*, Dec 2008, pp. 591–598.

[10] A. F. Frangi and W. J. Niessen, "Multiscale vessel enhancement filtering," in *MICCAI*, October 1998, pp. 130–137.

[11] A. Blum and T. Mitchell, "Combining labelled and unlabelled data with co-training," in *Proc. Conf. on Computational learning theory*, July 1998, pp. 92–100.

[12] J. H.Friedman, "Greedy function approximation: A gradient boosting machine," in *Proceedings of the eleventh annual conference on Computational learning theory*, February 2002, pp. 1189–1232.

[13] "Feedback on a publicly distributed database: the messidor database," *Image Analysis & Stereology*, vol. 33, pp. 231–234, aug 2014.

[14] "Kaggle challenge diabetic retinopathy detection," https://www.kaggle.com/c/diabetic-retinopathy-detection.

[15] T. Kauppi et al., "Diaretdb0: Evaluation database and methodology for diabetic retinopathy algorithms," *Technical report*, 2006.

[16] T. A. Syed and J. Sivaswamy, "Latent factor model based classification for detecting abnormalities in retinal images," in *2015 Asian Conference on Pattern Recognition (ACPR)*, Nov 2015, pp. 411–415.