Neurocomputing 149 (2015) 708–717

Contents lists available at ScienceDirect

Neurocomputing

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

Hierarchical retinal blood vessel segmentation based on feature and ensemble learning

Shuangling Wang [a], Yilong Yin [a,*], Guibao Cao [a], Benzheng Wei [b], Yuanjie Zheng [c], Gongping Yang [a]

[a] School of Computer Science and Technology, Shandong University, Jinan 250101, PR China
[b] College of Science and Technology, Shandong University of Traditional Chinese Medicine, Jinan 250355, PR China
[c] Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 5 April 2014
Received in revised form
25 June 2014
Accepted 28 July 2014
Communicated by: Shiguang Shan
Available online 12 August 2014

Keywords:
Convolutional neural network
Ensemble learning
Feature learning
Random forest
Retinal blood vessel segmentation

ABSTRACT

Segmentation of retinal blood vessels is of substantial clinical importance for diagnoses of many diseases, such as diabetic retinopathy, hypertension and cardiovascular diseases. In this paper, the supervised method is presented to tackle the problem of retinal blood vessel segmentation, which combines two superior classifiers: Convolutional Neural Network (CNN) and Random Forest (RF). In this method, CNN performs as a trainable hierarchical feature extractor and ensemble RFs work as a trainable classifier. By integrating the merits of feature learning and traditional classifier, the proposed method is able to automatically learn features from the raw images and predict the patterns. Extensive experiments have been conducted on two public retinal images databases (DRIVE and STARE), and comparisons with other major studies on the same database demonstrate the promising performance and effectiveness of the proposed method.

1. Introduction

Retinal blood vessel segmentation has been widely used in various scenarios. For example, change of the retinal blood vessel appearance is an important indicator for various ophthalmologic and cardiovascular diseases, such as diabetes, hypertension, and arteriosclerosis [1], therefore, automatic segmentation and analysis of the retinal vasculature play an extremely vital role in the implementation of screening programs for diabetic retinopathy, the evaluation of retinopathy of prematurity, foveal avascular region detection, arteriolar narrowing detection, the diagnosis of cardiovascular diseases and hypertension, and computer-assisted laser surgery [2]. Moreover, the generation of retinal maps and detection of branch points have been utilized for temporal or multimodal image registration, retinal image mosaic synthesis, optic disc identification, fovea localization and biometric identification [2].

Both manual delineation and automatic algorithms have been used in retinal vessel segmentation. However, they have not gained wide acceptance due to several challenges. Manual delineation is skill demanding, tedious, time-consuming, and infeasible if given a large volume of fundus image databases. Accuracy of the automatic segmentation algorithms (to be reviewed in Section 2) is limited due to low blood vessel contrast, irregular shaped bright and dark lesions (in the form of hemorrhages, exudates, drusen and the optic disc boundary), intricate vessel topology (including vessel crossing and branching, as well as variation of vessel diameter and vessel grey levels) and nonuniform illumination of images as well as image deformation of scaling, skewing and other distortions.

In this paper, we present the hybrid method based upon convolutional neural network (CNN) [3] and ensemble random forests (RFs) [4] for automatic retinal blood vessel segmentation. We first employed a set of preprocessing steps to correct the nonuniform illumination of retinal images and to improve vessel contrast. We then used CNN to extract a set of hierarchical features which are not only invariant to image translation, scaling, skewing and other distortions, but also contain image based multi-scale information of the geometric structure of retina. We finally trained ensemble RFs to obtain a vessel classifier. The whole pipeline of the proposed method is trainable and automatic. Moreover, our method can effectively deal with the challenges of retinal vessel segmentation, as shown by our evaluations conducted using two publicly available databases (the DRIVE [5] and STARE [1]) and comparisons with state-of-the-art. Experimental results show that our approach is competitive with state-of-the-art by achieving sensitivity/specificity/accuracy/AUC values of 0.8173/0.9733/0.9767/0.9475 for the DRIVE database and of 0.8104/0.9791/0.9813/0.9751 for the STARE

* Corresponding author. Tel./fax: +86 531 8839 1367.
E-mail address: ylyin@sdu.edu.cn (Y. Yin).

http://dx.doi.org/10.1016/j.neucom.2014.07.059
0925-2312/© 2014 Elsevier B.V. All rights reserved.

database. In contrast, previous methods of Fraz [6] produced values of 0.7406/0.9807/0.9480/0.9747 for the DRIVE database and of 0.7548/0.9763/0.9534/0.9768 for the STARE database.

The rest of this paper is organized as follows. Section 2 gives a review of related work reported in the latest literature. Section 3 provides the description of CNN and RF. Section 4 gives a detailed description of the proposed method. Section 5 evaluates the proposed method. Further, the concluding remarks are included in Section 6.

2. Related work

There are a number of methods available in the literature for retinal blood vessel segmentation, as reviewed by Fraz et al. [2]. These methods can be broadly divided into two categories: unsupervised and supervised [2,3].

2.1. Unsupervised methods

The unsupervised methods can be further classified into five main subcategories: matched filtering, morphological processing, vessel tracking, multiscale analysis, and model-based algorithms. The matched filter based methodology convolves a kernel based on the Gaussian or its derivatives with the retinal image to enhance the vessel features [1,7]. In order to design the matched filter kernel, this methodology exploits the piece-wise linear approximation, the decrease in vessel diameter along vascular length, and the Gaussian-like intensity profile of retinal blood vessels. Mathematical morphology coupling curvature evaluation [8] and matched filtering for centerline detection [9,10] is also deployed for retinal vessel segmentation. Vessel tracking based methodology [11] segments a vessel using local information and works at the level of a single vessel rather than the entire vasculature. The multiscale approaches are based upon scale-space analysis. The multiscale second-order local structure of an image (Hessian) is examined, and a vesselness measure is obtained on the basis of eigenvalue analysis of the Hessian [12]. The model based approaches include the vessel profile models [13–16], active contour models [17], and level sets based geometric models [18].

2.2. Supervised methods

Supervised methods utilizing ground truth data usually consist of two stages: (1) feature extraction, (2) classification. During feature extraction stage, the features can be obtained by means of two ways: hand-crafted features and learning based features.

The first approach is based on hand-designed feature extraction, where each pixel is projected into a set of features predefined with prior knowledge before classification. Neimeijer et al. [19] applied the Gaussian and its derivatives at multiple scales to form a feature vector for each pixel, augmented with the green channel of the RGB image. Stall et al. [5] used ridge profiles to compute a feature vector for each pixel, then performed feature selection. In [20], a feature vector is composed of the pixel intensity and two-dimensional Gabor wavelet transform responses taken at multiple scales. Ricci and Perfetti [21] deployed line operators to form a feature vector. Lupascu et al. [22] used a feature vector at different spatial scales for each pixel. In [23], a feature vector is obtained by combination of moment-invariant and gray-level features. In [6], the method utilized a feature vector based on the orientation analysis of gradient vector field, morphological transformation, line strength measures, and Gabor filter responses.

The second approach deploys Artificial neural network (ANN) or its variants to segment retinal blood vessels, where feature extraction and classification are integrated into one pipeline: feature extraction itself is directly learned from raw image and not enforced by designers, then followed in the ANN process is a generation of predictions. Nekovei and Sun [24] presented a method using a back-propagation network for the detection of blood vessels. In this method, neural network is directly applied to the image pixels without any prior feature extraction. For each pixel being classified, raw pixel values in a square window centered on it are fed to the network as input.

Although these supervised methods have achieved satisfactory segmentation results in some scenarios, there are still some issues. In hand-designed feature extraction approaches, the features must be very carefully predefined before classification, making feature detection very time-consuming and tedious. Most importantly, hand-designed features have to be redesigned for datasets with different characteristics. In contrast, the approaches based upon feature learning can extract features automatically from the raw images. Moreover, CNN is supervised feature learner able to learn complex invariances such as scale and rotational invariance. However, the classification mechanism in ANN or its variants is fairly simple: usually nonlinear activation functions are employed in the last output layer to predict patterns, which results in the low performance of ANN. Considering all of above, we propose to combine two well-known classifiers: Convolutional Neural Network (CNN) and Random Forest (RF). In this paper, CNN works as a trainable hierarchical feature extractor, and then ensemble RFs perform as the trainable classifier supplied with hierarchical features learned from CNN.

3. Methodology

The proposed method leverages trainable hierarchical features extracted using the CNN algorithm [3] and performs classification

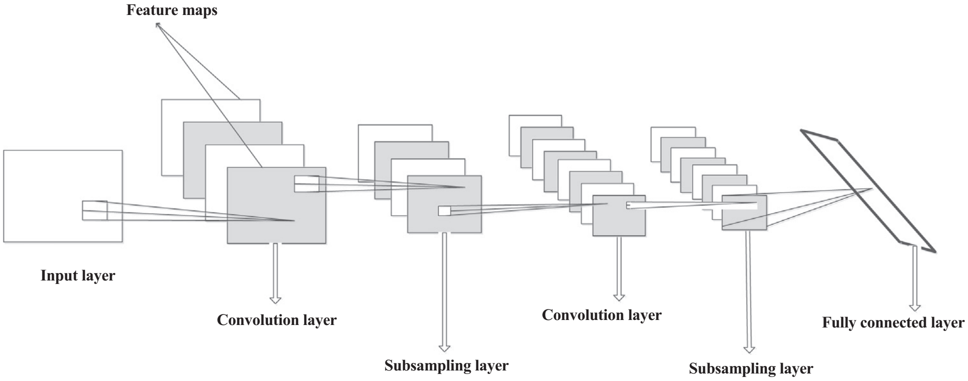


Fig. 1. Schematic diagram of convolutional neural network.

based on the random forest (RF) technique [4]. In this section, we will briefly review the CNN and RF techniques.

3.1. Trainable feature extractor: convolutional neural network

CNN was neurobiologically motivated by the findings of locally sensitive and orientation-selective nerve cells in the visual cortex, as introduced by LeCun and Bengio in 1995 [3]. The authors of CNN also designed a classic network structure that implicitly extracts relevant features from raw image, named LeNet-5 [25]. Convolutional Neural Network is a special kind of multi-layer neural networks, with the following characteristics: CNN is a feed-forward network that can extract topological properties from an image. Like almost every other neural network, it is trained with a version of the back-propagation algorithm. CNN is designed to recognize visual patterns directly from pixel images with minimal preprocessing. They can recognize patterns with extreme variability, such as handwritten characters.

CNN typically consists of a convolution layer, a subsampling layer, and a fully connected layer. In CNN, successive layers of convolution and subsampling are typically alternated. Schematic diagram of CNN is shown in Fig. 1.

3.1.1. Convolution layer (C)

Convolution layer works as a feature extraction layer. All neurons in a feature map share the same set of weights and the same bias. In this way, all neurons in a feature map detect the same feature at different positions on the input. The other feature maps in this layer use different sets of weights and biases, so different types of local features are extracted, and weight sharing reduces the number of free parameters.

3.1.2. Subsampling layer (S)

Subsampling layer works as a feature selection layer. The subsampling layers perform a local averaging and a subsampling to reduce the spatial resolution of each feature map. A certain degree of shift and distortion invariance can be achieved by reducing the spatial resolution of the feature map. In addition, the weight sharing is applied in subsampling layers.

3.1.3. Fully connected layer (FC)

This is the standard layer of a multi-layer network. It performs a linear multiplication of the input vector by a weight matrix.

3.2. Trainable traditional classifier: random forest

Random forest (RF) is a widely used “ensemble learning” method for classification and regression, which was proposed by Breiman [4]. This method has the following desirable characteristics:

1) It is simple, easily parallelized and user-friendly (only two parameters).
2) It can work efficiently on large data sets.
3) It is relatively robust even in the presence of many noisy features, Random Forest performs well, making it unnecessary to perform feature selection procedure.
4) Its accuracy is as good as Adaboost and sometimes even better.
5) It is faster than bagging or boosting.

Random forest consists of tree predictors, where each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [4]. Each tree is grown as follows: First, the training set used to grow each tree is a bootstrap sample of the observations, i.e., a sample of size N drawn with replacement from the original sample of N observations. Some observations are represented multiple times, while others are left out. The left-out observations are called “out-of-bag” and these out-of-bag samples can be used to calculate an unbiased error rate and variable importance, which eliminates the need for a test set or cross-validation. Second, the best split at each node is selected from among a random subset of the predictor variables. Different variables are used at each split in different trees. Third, each tree is grown to its large extent possible, and there is no pruning. The random forests final prediction is the

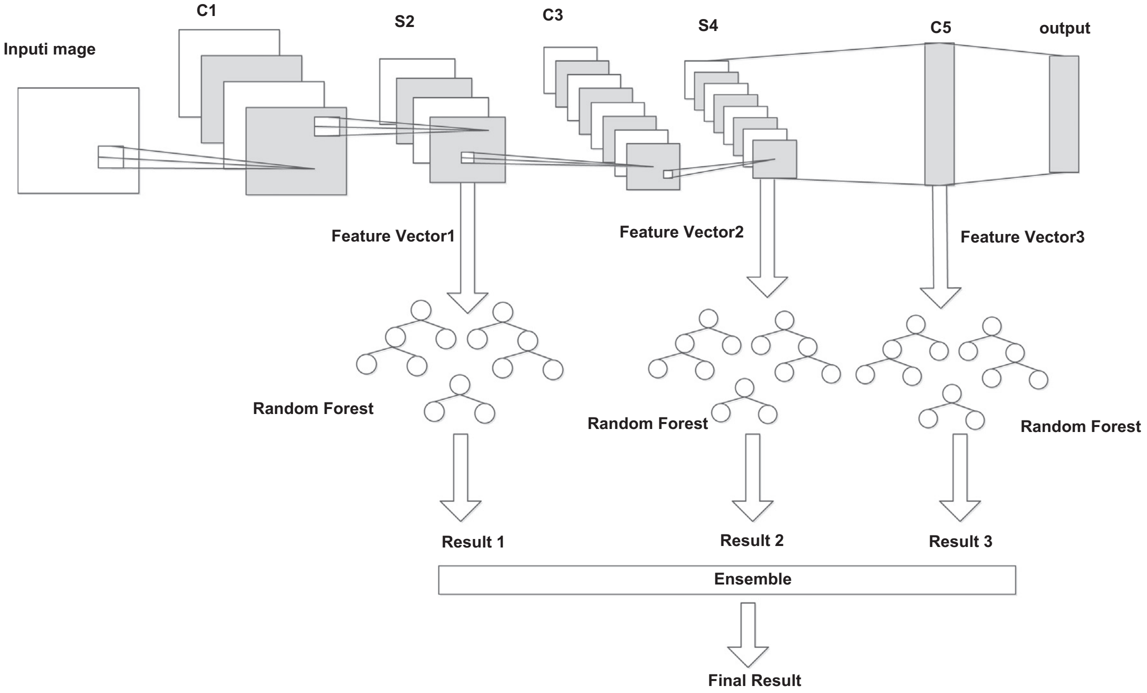


Fig. 2. Structure of the proposed method. Convolutional layer feature maps are indexed by C1, C3, and C5. Subsampling layer feature maps are indexed by S2 and S4.

mean prediction (regression) or class with maximum votes (classification). Breiman [4] showed that the prediction error converges to a limiting value as the number of trees tends to infinity.

4. Proposed method

In the proposed method, CNN performs as a trainable hierarchical feature extractor and ensemble RFs work as a classifier. It is noteworthy that features utilized in the proposed method learned from not only the last layer output of CNN but also the intermediate layers outputs. Features learned from the same layer of CNN are fed into one RF classifier. Finally, winner-takes-all is employed to ensemble the predictions of each RF. For training procedure, raw pixel values from a square sub-window centered on representative pixel sample are fed into CNN, when CNN is well trained, several RFs are trained with learned hierarchical features extracted from CNN. Finally, ensemble method of winner-takes-all is employed to predict whether the center pixel of a sub-window belongs to a blood vessel or not. For testing procedure, sliding sub-windows sampling from the testing image are directly fed into CNN to extract the learned hierarchical features, then winner classifier supplied with these features is used to predict the result. As shown in Fig. 2, our method consists of three basic steps of preprocessing, hierarchical feature extraction and ensemble classification. In preprocessing, image enhancement and superpixel based sample selection are carried out.

4.1. Preprocessing

For colored images in RGB space, the green channel shows the best vessel-background contrast, while the red and blue channels show low contrast and are very noisy [20]. Hence, we just employ the green channel of the RGB colored retinal image in the proposed method.

4.1.1. Image enhancement

First of all, images were preprocessed with histogram equalization and Gaussian filtering, for the purpose of reducing noise, making the intensity more uniform and improving the contrast of the vessels.

4.1.2. Superpixel based sample selection method

For large volumes of digital fundus retinal images, training pixel classifier is large time consuming and big storage space demanding, therefore, we introduce an automatic superpixel based sample selection method, which is part of our previous work [26]. Superpixel, generally obtained by image over-segmentation, is local nearby pixel grouping. Advantages of using superpixels lie in the preserve natural image boundaries while

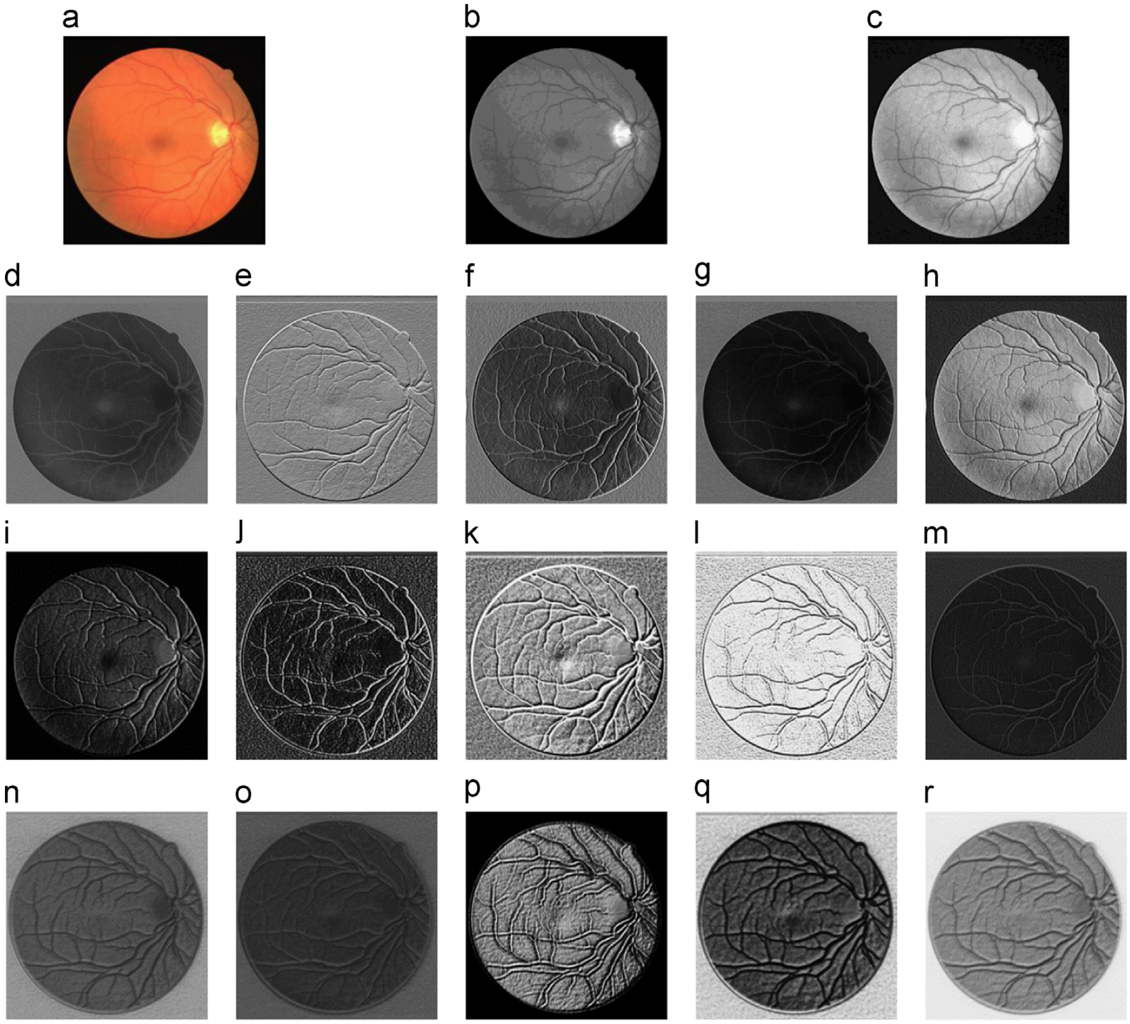


Fig. 3. Visualization of learned hierarchical level features. (a) Original image. (b) Green channel of RGB colored image. (c) Enhanced green channel of RGB colored image. (d)-(h) Learned features from S2. (i)-(m) Learned features from S4. (n)-(r) Learned features from C5.

capturing redundancy in the data as well as enforcing local consistency [27]. In this paper, we apply Simple Linear Iterative Clustering (SLIC) [28] for superpixel generation, which is an adaptation of the *k*-means algorithm. In order to reduce redundancy among pixel samples and pick out representative samples, pixels inside the same superpixel can be regarded as homogenous, then we randomly select several pixels inside a superpixel to represent all the pixels inside this superpixel. For simplicity and efficiency, only one pixel is randomly selected for one superpixel in this method.

4.2. Trainable hierarchical feature extraction

In this method, CNN is utilized as a trainable hierarchical feature extractor. For each target pixel, raw pixel values from a square subwindow centered on it are directly fed as input to the CNN. With local receptive fields, neurons in CNN can detect elementary visual features such as oriented edges, end-points, corners. These features are then combined by the successive layers in order to capture higher-order features [25]. We take full advantage of these hierarchical level features learned from different layers of CNN to train the ensemble classifier. In the proposed method, convolutional layer feature maps are indexed by C1, C3, and C5 and subsampling layer feature maps are indexed by S2 and S4. In order to learn more about the characteristics of features learned from different layers, the learned hierarchical features from S2, S4 and C5 are partly selected to be visualized using Matlab in Fig. 3. We can see from the figure that previous layer in CNN extracts elementary features with a high resolution while successive layers capture higher-order statistics with a low resolution.

As to the inputs and the size of CNN, two trade-offs should be considered. First, when the dimensionality of the input space increases, the data becomes increasingly sparse, making ANN difficult to learn a decision boundary accurately [29]. Therefore inputs to CNN should be small while keeping large enough to describe the data. Considering the average width and the average area of bright and dark lesions, a square window with size of 25 × 25 is adopted as input to the CNN. Second, the training time tends to scale with the amount of training data and size of the network, and therefore it is generally preferable to train smaller networks [29]. For the deep network, usually five or six layers are enough. We adopted the classic LeNet-5 network, with 4 × 4 convolution matrix, subsampling by factor 2, as shown in Table 1. Bold fonts indicate layers that are used to train ensemble classifier. Different maps are for detecting different features while same map detects features from different locations. Learned features in the same feature map can be seen as contextual information. Since there is no need to perform feature selection in random forest, we utilize all learned features from different maps as well as same map in one layer. Take the layer S2 for example, 11 × 11 neurons in one map can learn 121 features which characterize the contextual information. At the same time, there are 12 maps for 12 different features. Therefore, the learned features' dimension is $11 \times 11 \times 12 = 1452$, with one neuron corresponding to one dimensional feature.

4.3. Ensemble classification

Ensemble classification is a process in which multiple classifiers are strategically generated and combined to solve the same problem [30]. Ensemble learning utilizes multiple classifiers to get better predictive performance by combining the decisions from multiple weak learners into one powerful ensemble classifier. In this paper, we apply random forest as the base classifier and winner-takes-all is deployed as the ensemble model.

4.3.1. Base classifier: random forest

Learned features are extracted from layer S2, S4, C5 respectively, and then features obtained from the same layer are fed into a RF classifier individually, see Fig. 2. Thereby, three layers correspond to three RF classifiers, donated as RF-1, RF-2, RF-3, respectively. The independent classifiers work in parallel. Giver an instance, they all give their decisions which are then combined by winner-takes-all to make the final decision.

Table 1
6-layer architecture for network (not counting the input), window = 25 × 25).

Layer	Type	Maps and neurons	Kernel size	Learned feature dimension
0	input	1 map of 25 × 25 neurons		
1	convolution	12 maps of 22 × 22 neurons	4 × 4	5808
2	**subsampling**	**12 maps of 11 × 11 neurons**	**2 × 2**	**1452**
3	convolution	12 maps of 8 × 8 neurons	4 × 4	768
4	**subsampling**	**12 maps of 4 × 4 neurons**	**2 × 2**	**192**
5	**convolution**	**100 neurons**	**4 × 4**	**100**
6	output	1 neuron		

Table 2
Performance metrics for retinal vessel segmentation.

Measure	*Description*
TPR	TP/vessel pixel count
FPR	FP/non-vessel pixel count
Sensitivity(SN)	TPR or TP/(TP + FN)
Specificity(SP)	1-FPR or TN/(TN + FP)
Accuracy(Acc)	(TP + TN)/FOV pixel count
AUC	Area under the ROC curve

Table 3
Performance results on DRIVE Database images.

Image	*SN*	*SP*	*Acc*	*AUC*
1	0.8787	0.9818	0.9842	0.9637
2	0.8808	0.9793	0.9824	0.9651
3	0.8206	0.9694	0.9739	0.9352
4	0.8495	0.9769	0.9800	0.9505
5	0.7962	0.9681	0.9724	0.9347
6	0.7449	0.9581	0.9640	0.9430
7	0.8168	0.9721	0.9758	0.9382
8	0.8000	0.9715	0.9751	0.9415
9	**0.6674**	**0.9557**	**0.9609**	**0.9385**
10	0.8213	0.9758	0.9787	0.9434
11	0.8467	0.9772	0.9801	0.9367
12	0.8493	0.9784	0.9811	0.9601
13	0.7556	0.9597	0.9654	0.9432
14	0.8835	0.9846	0.9864	0.9626
15	0.8503	0.9827	0.9845	0.9301
16	0.7747	0.9661	0.9705	0.9489
17	0.7129	0.9596	0.9646	0.9276
18	0.8116	0.9755	0.9784	0.9545
19	**0.9032**	**0.9868**	**0.9884**	**0.9605**
20	0.8821	0.9859	0.9874	0.9709
Average	**0.8173**	**0.9733**	**0.9767**	**0.9475**

[a] Bold fonts denote worst case accuracy, best case accuracy, and average result, respectively.

4.3.2. *Ensemble method: winner-takes-all*

The Winner-takes-all classifier is simple but effective. Its final class output is determined by the base classifier which wins the best classification performance among all the classifiers fused. We incorporate it in our method due to two reasons:

1) In winner-takes-all, base classifiers compete with each other, thus making it easy to select the best classifier while majority voting or averaging ensemble method can avoid selection of the worst classifier by obtaining a compromise between fused classifiers. Moreover, experiment is conducted to verify winner-takes-all is the best choice, and experimental details are shown in Section 5.
2) Intuitively, higher-order features learned from successive layers of CNN are more abstract and therefore more suitable for object or image classification rather than pixel level classification, while elementary features implicitly extracted from the previous layer of CNN are suitable for pixel level classification.

In out pipeline, we choose base RF classifier trained with features from S2 (RF-1) as the winner.

5. Experiments

5.1. *Materials*

Similar to most of the retinal vessel segmentation methods, the proposed method is evaluated on two well established public databases: the DRIVE database and the STARE database.

5.1.1. *DRIVE database*

The photographs for the DRIVE (Digital Retinal Images for Vessel Extraction) database [5] were obtained from a diabetic retinopathy screening program in the Netherlands. The screening population consisted of 400 diabetic subjects between 25 and 90 years of age. Forty photographs have been randomly selected, 33 do not show any sign of diabetic retinopathy and 7 show signs of mild early diabetic retinopathy. Each image has been JPEG compressed.

Table 4
Performance results on STARE Database images.

Image	*SN*	*SP*	*Acc*	*AUC*
1	0.7985	0.9754	0.9781	0.9695
2	0.7515	0.9749	0.9772	0.9338
3	0.8502	0.9867	0.9878	0.9687
4	**0.5483**	**0.9481**	**0.9535**	**0.9623**
5	0.8246	0.9751	0.9782	0.9611
6	0.8833	0.9887	0.9897	0.9839
7	0.9416	0.9937	0.9944	0.9896
8	**0.9447**	**0.9943**	**0.9949**	**0.9884**
9	0.8955	0.9879	0.9892	0.9872
10	0.8879	0.9866	0.9880	0.9713
11	0.8918	0.9884	0.9895	0.9762
12	0.9107	0.9896	0.9907	0.9874
13	0.8789	0.9841	0.9860	0.9862
14	0.8592	0.9802	0.9827	0.9849
15	0.7851	0.9718	0.9751	0.9753
16	0.7387	0.9573	0.9633	0.9766
17	0.8889	0.9847	0.9866	0.9799
18	0.6451	0.9730	0.9749	0.9645
19	0.6311	0.9770	0.9783	0.9780
20	0.6523	0.9645	0.9678	0.9765
Average	**0.8104**	**0.9791**	**0.9813**	**0.9751**

[a] Bold fonts denote worst case accuracy, best case accuracy, and average result, respectively.

The images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45° field of view (FOV). Each image was captured using 8 bits per color plane at 768 by 584 pixels. The FOV of each image is circular with a diameter of approximately 540 pixels. For this database, the images have been cropped around the FOV. For each image, a mask image is provided that delineates the FOV.

The set of 40 images has been divided into a training set and a test set evenly and each set contains 20 images. For the training images, a single manual segmentation of the vasculature is

Table 5
Performance measures on DRIVE and STARE.

Database	*Segmentation*	*SN*	*SP*	*Acc*	*AUC*
DRIVE	2nd human observer	0.7796	0.9717	0.9464	–
	Proposed Method	**0.8173**	**0.9733**	**0.9767**	**0.9475**
STARE	2nd human observer	0.8955	0.9382	0.9347	–
	Proposed Method	**0.8104**	**0.9791**	**0.9813**	**0.9751**

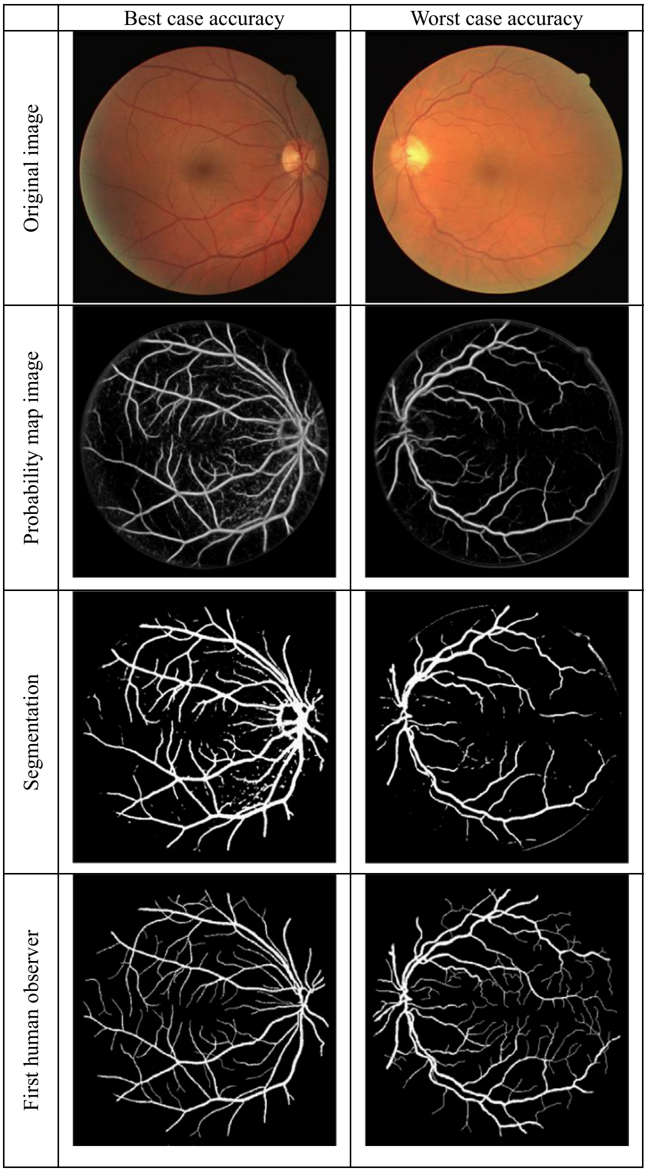


Fig. 4. Segmentation results for the DRIVE database.

available. For the test cases, two manual segmentations are available; one is used as the gold standard, the other one can be used to compare computer generated segmentations with those of independent human observer. All human observers were instructed and trained by an ophthalmologist on how to segment the vasculature manually. They were asked to mark all vessel pixels with a confidence of at least 70%.

5.1.2. STARE database

The STARE (Structured Analysis of the Retina) database [1] contains 20 images for blood vessel segmentation, out of which ten images contain pathology. The digitized images are obtained by a Topcon TRV-50 fundus camera at 35° field of view. The images were digitized to 605 × 700 pixels, 8 bits per color channel. The diameter of the FOV is approximately 650 × 700 pixels. All the images are manually segmented by two observers. The second observer pays many more attention to the thinner vessels than the first one and segmentation of the first observer is usually treated as the ground truth.

5.2. Performance measurements

In the retinal vessel segmentation process, every pixel is classified either as vessel or surrounding tissue. Consequently, there are four possible results for each segmentation: two classifications and two misclassifications. The true positive (TP) means a pixel is classified as vessel in both in the ground truth and segmented image, and the true negative (TN) means a pixel is recognized as a non-vessel in both the ground truth and segmented image. As to the two misclassifications, the false negative (FN) means a pixel is identified as non-vessel in the segmented image but as a vessel pixel in the ground truth image, and the false positive (FP) means a pixel is classified as vessel in the segmented image but non-vessel in the ground truth image.

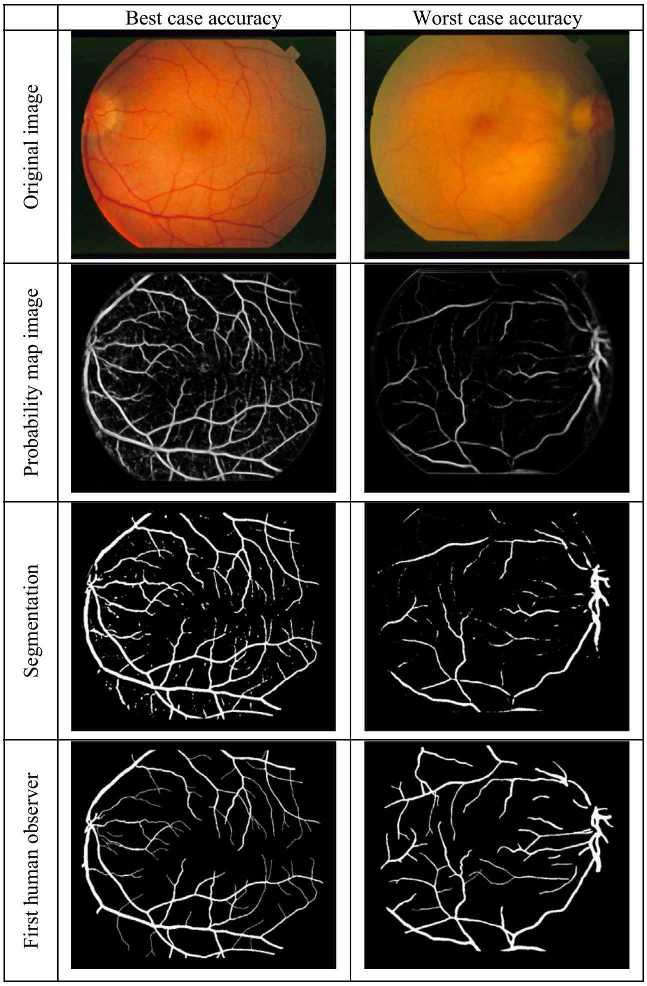


Fig. 5. Segmentation results for the STARE database.

The true positive rate (TPR) is the fraction of pixels correctly detected as vessel pixels. The false positive rate (FPR) represents the fraction of pixels incorrectly detected as vessel pixels. Sensitivity (SN) is the ability of an algorithm to identify the vessel pixels, and it is equal to TPR. Specificity (SP) reflects the ability to detect non-vessel pixels, also denoted as (1-FPR). The accuracy (Acc) represents the ratio of the total number of correctly classified pixels to the number of pixels in the image FOV.

A receive operating characteristic (ROC) curve is a plot of TPR versus FPR by varying the threshold on the probability map image. The performance metrics also include the area under ROC curve (AUC). In this paper, all the performance metrics are computed considering only pixels inside the FOV over all test images. These metrics is summarized in Table 2.

5.3. Experiments setting

In this study, all the experiments are implemented in Matlab and conducted on a high performance computer with four Intel Xeon CPUs at 2.00 GHz and 256 GB RAM. In all the experiments, networks are trained 100 epochs and each of Random Forest is trained by default parameters. For DRIVE and STARE databases, the training of both feature extractor and predictor is conducted on 20 trained image using 193,325 samples and 192,419 samples, respectively. About 10,000 pixels per image are selected by the proposed superpixel based on sample selection method. The STARE database does not provide available separate test and training sets. In this study, the training set is relatively small

Table 6
Comparison of segmentation results for individual and ensemble classifier on DRIVE images.

Method	*SN*	*SP*	*Acc*	*AUC*
CNN	0.6425	0.9478	0.9545	0.9378
RF-1	0.8173	0.9733	0.9767	0.9475
RF-2	0.7899	0.9692	0.9731	0.9476
RF-3	0.7576	0.9644	0.9690	0.9490
Average ensemble	0.7787	0.9676	0.9718	0.9521
Weighted ensemble one	0.7983	0.9704	0.9742	0.9514
Weighted ensemble two	0.8074	0.9717	0.9754	0.9502
Median ensemble	0.7959	0.9700	0.9739	0.9506
Winner-takes-all	**0.8173**	**0.9733**	**0.9767**	**0.9475**

Table 7
Comparison of segmentation results for individual and ensemble classifier on STARE images.

Method	*SN*	*SP*	*Acc*	*AUC*
CNN	0.6648	0.9626	0.9666	0.9364
RF-1	0.8104	0.9791	0.9813	0.9751
RF-2	0.7986	0.9778	0.9801	0.9768
RF-3	0.7714	0.9746	0.9773	0.9706
Average ensemble	0.7974	0.9777	0.9800	0.9777
Weighted ensemble one	0.8036	0.9783	0.9806	0.9771
Weighted ensemble two	0.8076	0.9788	0.9810	0.9777
Median ensemble	0.8009	0.9775	0.9804	0.9781
Winner-takes-all	**0.8104**	**0.9791**	**0.9813**	**0.9751**

(about 2% of the entire database), therefore, the performance is evaluated on the whole set of 20 images.

5.4. Performance of the proposed method

The proposed method was evaluated on DRIVE and STARE database images with available ground truth images. Since FOV masks are not provided for STARE images, they were automatically generated using source code downloaded from site by Hoover [1] (available at http://www.parl.clemson.edu/~ahoover/stare/). Performance results are shown in Tables 3 and 4.

Vessel segmentation results: The binary vessel segmentation image is obtained from the probability map image based on OSTU threshold [31]. The performance metrics are calculated by taking the first human observer as the ground truth. The best case accuracy, sensitivity, specificity, AUC for the DRIVE database are 0.988368, 0.903219, 0.986776, 0.960497, respectively, and the worst case measures are 0.960937, 0.667402, 0.955739 and 0.938519, respectively. The best case vessel segmentation result for the STARE database has an accuracy of 0.994903; sensitivity, specificity, AUC are 0.94471, 0.99432, and 0.988388, respectively. The worst case accuracy is 0.953499; sensitivity, specificity, and AUC are 0.548304,0.948144, and 0.962269, respectively.

5.5. Comparison with the second human observer

The average performance measures for the DRIVE, STARE databases are tabulated in Table 5. From Table 5, we can see that the average accuracy rates and average specificity rates obtained by the algorithm are higher than the second human observers for the DRIVE and STARE databases.

The segmented images with the best and worst accuracies from the DRIVE and STARE databases are illustrated in Figs. 4 and 5, respectively.

From Figs. 4 and 5, we can see that the proposed method fail to capture some fine vessels. However, fine vessels are not important for clinical importance [20,23]. It is also worth mentioning that many clinical investigations take measurements only on major ones in a limited region around the optic disc, without using fine vessels [32].

5.6. Comparison of individual classifier and ensemble classifier

In this experiment, we compare the average segmentation performance of individual classifier and ensemble classifier in terms of SN, SP, Acc, and AUC. Ensemble is performed on probability map image digitalized by continuous output of RF. For each target pixel, each individual classifier gives its probability values,

Table 8
Performance comparison of vessel segmentation methods on DRIVE images.

No	*Type*	*Methods*	*Year*	*SN*	*SP*	*Acc*	*AUC*
1.		2nd human observer	–	0.7796	0.9717	0.9470	N.A
2.	Unsupervised methods	Zana [8]	2001	0.6971	N.A	0.9377	0.8984
3.		Jiang [33]	2003	N.A	N.A	0.9212	0.9114
4.		Mendonca [9]	2006	0.7344	0.9764	0.9452	N.A
5.		Al-Diri [17]	2009	0.7282	0.9551	N.A	N.A
6.		Lam [15]	2010	N.A	N.A	0.9472	0.9614
7.		Miri [34]	2011	0.7352	0.9795	0.9458	N.A
8.		Fraz [10]	2011	0.7152	0.9759	0.9430	N.A
9.		You [35]	2011	0.7410	0.9751	0.9434	N.A
10.	Supervised methods	Niemeijer [19]	2004	N.A	N.A	0.9416	0.9294
11.		Soares [20]	2006	0.7332	0.9782	0.9461	0.9614
12.		Staal [5]	2004	N.A	N.A	0.9441	0.9520
13.		Ricci [21]	2007	N.A	N.A	0.9595	0.9558
14.		Lupascu [22]	2010	0.7200	N.A	0.9597	0.9561
15.		Marin [23]	2011	0.7067	0.9801	0.9452	0.9588
16.		Fraz [6]	2012	0.7406	0.9807	0.9480	0.9747
17.		**Proposed Method**	**2013**	**0.8173**	**0.9733**	**0.9767**	**0.9475**

[a] N.A = Not Available.

Table 9
Performance comparison of vessel segmentation methods on stare images.

No	*Type*	*Methods*	*Year*	*SN*	*SP*	*Acc*	*AUC*
1.		2nd human observer	–	0.8951	0.9384	0.9348	–
2.	Unsupervised methods	Hoover [1]	2000	0.6747	0.9565	0.9264	N.A
3.		Jiang [33]	2003	N.A	N.A	0.9009	N.A
4.		Mendonca [9]	2006	0.6996	0.9730	0.9440	N.A
5.		Lam [14]	2008	N.A	N.A	0.9474	0.9392
6.		Al-Diri [17]	2009	0.7521	0.9681	N.A	N.A
7.		Lam [15]	2010	N.A	N.A	0.9567	0.9739
8.		Fraz [10]	2011	0.7311	0.9680	0.9442	N.A
9.		You [35]	2011	0.7260	0.9756	0.9497	N.A
10.	Supervised methods	Staal [5]	2004	N.A	N.A	0.9516	0.9614
11.		Soares[20]	2006	0.7207	0.9747	0.9479	0.9671
12.		Ricci [21]	2007	N.A	N.A	0.9584	0.9602
13.		Marin [23]	2011	0.6944	0.9819	0.9526	0.9769
14.		Fraz [6]	2012	0.7548	0.9763	0.9534	0.9768
15.		**Proposed Method**	**2013**	**0.8104**	**0.9791**	**0.9813**	**0.9751**

[a] N.A = Not Available.

and then these probability values are combined by the certain ensemble method. For weighted ensemble, weight vector for weighted ensemble one is 0.5, 0.3, 0.2, and for weighted ensemble two 0.7, 0.2, 0.1 respectively. Experimental results are shown in Tables 6 and 7.

From Tables 6 and 7, following conclusions can be drawn:

1) Each individual RF performs better than pure CNN, which verified that the proposed hybrid method combine the merits of feature learning and traditional classifier.
2) RF-1 obtains the better performance than RF-2 and RF-3. This is because learned features fed into RF-1 are more elementary while learned features fed into RF-2 and RF-3 are more abstract with higher-order statistics.
3) For ensemble method, using a winner-takes-all method results in the best overall performance. Winner-takes-all can get the best classifier while average or weighted ensemble allows avoiding the selection of the worst classifier, but fail to outperform the best individual classifier of the ensemble.
4) Results on STARE database images demonstrate the flexibility of the proposed method on different type of images.

5.7. Comparison with published methods

In Tables 8 and 9, we compare our approach with the most recent methods in terms of sensitivity, specificity, accuracy, and AUC on the DRIVE and STARE databases, respectively.

Tables 8 and 9 show us that the proposed method achieves state of the art SN and Acc on both DRIVE and STARE database images.

Following conclusions can be drawn from Tables 8 and 9:

1) Square windows directly sample on raw image and raw image intensities are utilized directly as inputs.
2) Learned features are general, not problem specific. Therefore, an important practical advantage of square windows is not necessary to design features with a priori knowledge of the problem.
3) Individual classifier can be easily parallel.
4) By winner-takes-all ensemble method, ensemble can maintain advantages while get the best classifier.

5.8. Cross-training

In order to test the proposed method is robust for training set, the following cross-training experiment has been carried out. Specifically, both feature extractor and classifier are trained on one same database while tested on another. Cross-training results on DRIVE and STARE in terms of sensitivity, specificity, accuracy, and AUC are shown in Table 10. Performance comparison results with cross-training in terms of average accuracy are shown in Table 11. Results in Tables 10 and 11 further prove that the proposed method is independent of training set.

Table 10
Cross-training results on drive and stare.

	SN	*SP*	*Acc*	*AUC*
Drive (trained on STARE)	0.8431	0.9774	0.9803	0.9271
STARE(trained on DRIVE)	0.7116	0.9675	0.9710	0.9391

Table 11
Performance comparison of results with cross training in terms of average accuracy.

Method	*Drive (trained on STARE)*	*STARE (trained on DRIVE)*
Soares [20]	0.9397	0.9327
Ricci [21]	0.9266	0.9464
Marin [23]	0.9448	0.9528
Fraz [6]	0.9456	0.9493
Proposed Method	**0.9803**	**0.9710**

6. Conclusion

In this paper, we proposed a hierarchical retinal blood vessel segmentation method based on feature and ensemble learning. The proposed method has several unique characteristics. First, our features are extracted using not only the last layer output but also the intermediate output of CNN and therefore contain multiscale information of the geometric structure of the retina. Second, we are the first to introduce random forest into retinal blood vessel segmentation, and employ winner-takes-all as the classifier ensemble method. Third, the whole pipeline of the proposed method is automatic and trainable, which is accomplished by a combination of feature learning and ensemble learning. Fourth, our method was validated using two publically available databases and shown to outperform state-of-the-art. Finally, our method was shown to better handle the challenges in retinal vessel segmentation. This is because CNN is able to extract scale and rotational invariant features and RF is well-known for strong generalization capability.

Training time for one epoch is approximately 120 min for CNN. Since CNN is trained for 100 epochs, the total training time is about 8 days. Our future work would first focus on improving the computational time for training of CNN. To achieve this goal, we plan to deploy Genetic Algorithm (GA) for the parametric optimization of CNN. Moreover, we are interested in trying the max-pooling CNN [36] as the hierarchical feature extractor and comparing with CNN. We also have a plan to apply the proposed method to other biomedical images, e.g. membrane segmentation.

Acknowledgments

The authors would particularly like to thank the anonymous reviewers for their helpful suggestions. This work is supported by NSFC Joint Fund with Guangdong under Key Project U1201258, Program for New Century Excellent Talents in University of Ministry of Education of China (NCET-11–0315) and Shandong Natural Science Funds for Distinguished Young Scholar under Grant no. JQ201316.

References

[1] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, IEEE Trans. Med. Imaging 19 (2000) 203–210.
[2] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, S.A. Barman, Blood vessel segmentation methodologies in retinal images–A survey, Comput. Methods Progr. Biomed. 108 (2012) 407–433.
[3] Y. LeCun, Y. Bengio, Convolutional networks for images, speech, and time series, The Handbook of Brain Theory and Neural Networks 3361 (1995).
[4] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[5] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (2004) 501–509.
[6] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, S.A. Barman, An ensemble classification-based approach applied to retinal blood vessel segmentation, IEEE Trans. Biomed. Eng. 59 (2012) 2538–2548.
[7] L. Gang, O. Chutatape, S.M. Krishnan, Detection and measurement of retinal vessels in fundus images using amplitude modified second-order Gaussian filter, IEEE Trans. Biomed. Eng. 49 (2002) 168–172.

[8] F. Zana, J.-C. Klein, Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation, IEEE Trans. Image Process. 10 (2001) 1010–1019.

[9] A.M. Mendonca, A. Campilho, Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction, IEEE Trans. Med. Imaging 25 (2006) 1200–1213.

[10] M.M. Fraz, S. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A. R. Rudnicka, C.G. Owen, An approach to localize the retinal blood vessels using bit planes and centerline detection, Comput. Methods Progr. Biomed. 108 (2012) 600–616.

[11] F.K. Quek, C. Kirbas, Vessel extraction in medical images by wave-propagation and traceback, IEEE Trans. Med. Imaging 20 (2001) 117–131.

[12] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering,, Medical Image Computing and Computer-Assisted Intervention—MICCAI'98, Springer (1998) 130–137.

[13] L. Wang, A. Bhalerao, R. Wilson, Analysis of retinal vasculature using a multiresolution hermite model, IEEE Trans. Med. Imaging 26 (2007) 137–152.

[14] B.S. Lam, H. Yan, A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields, IEEE Trans. Med. Imaging 27 (2008) 237–246.

[15] B.S. Lam, Y. Gao, A.-C. Liew, General retinal vessel segmentation using regularization-based multiconcavity modeling, IEEE Trans. Med. Imaging 29 (2010) 1369–1381.

[16] H. Narasimha-Iyer, J.M. Beach, B. Khoobehi, B. Roysam, Automatic identification of retinal arteries and veins from dual-wavelength images using structural and functional features, IEEE Trans. Biomed. Eng. 54 (2007) 1427–1435.

[17] B. Al-Diri, A. Hunter, D. Steel, An active contour model for segmenting and measuring retinal vessels, IEEE Trans. Med. Imaging 28 (2009) 1488–1497.

[18] K. Sum, P.Y. Cheung, Vessel extraction under non-uniform illumination: a level set approach, IEEE Trans. Biomed. Eng. 55 (2008) 358–360.

[19] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff, Comparative study of retinal vessel segmentation methods on a new publicly available database, Medical Imaging, International Society for Optics and Photonics (2004) 648–656.

[20] J.V. Soares, J.J. Leandro, R.M. Cesar, H.F. Jelinek, M.J. Cree, Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification, IEEE Trans. Med. Imaging 25 (2006) 1214–1222.

[21] E. Ricci, R. Perfetti, Retinal blood vessel segmentation using line operators and support vector classification, IEEE Trans. Med. Imaging 26 (2007) 1357–1365.

[22] C.A. Lupascu, D. Tegolo, E. Trucco, FABC: retinal vessel segmentation using AdaBoost, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 1267–1274.

[23] D. Marín, A. Aquino, M.E. Gegúndez-Arias, J.M. Bravo, A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features, IEEE Trans. Med. Imaging 30 (2011) 146–158.

[24] R. Nekovei, Y. Sun, Back-propagation network and its configuration for blood vessel detection in angiograms, IEEE Trans. Neural Netw. 6 (1995) 64–72.

[25] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324.

[26] S. Wang, G. Cao, B. Wei, Y. Yin, G. Yang, C. Li, Hierarchical level features based trainable segmentation for electron microscopy images, Biomed. Eng. Online 12 (2013) 59.

[27] A. Lucchi, K. Smith, R. Achanta, V. Lepetit, P. Fua, A fully automated approach to segmentation of irregularly shaped cellular structures in EM images, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010, Springer (2010) 463–471.

[28] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 2274–2282.

[29] E. Jurrus, A.R. Paiva, S. Watanabe, J.R. Anderson, B.W. Jones, R.T. Whitaker, E. M. Jorgensen, R.E. Marc, T. Tasdizen, Detection of neuron membranes in electron microscopy images using a serial neural network architecture, Med. Image Anal. 14 (2010) 770–783.

[30] R. Polikar, Ensemble based systems in decision making, IEEE Circuits Syst. Mag. 6 (2006) 21–45.

[31] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (1975) 23–27.

[32] Y. Wang, G. Ji, P. Lin, E. Trucco, Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition, Pattern Recognition 46 (2013) 2117–2133.

[33] X. Jiang, D. Mojon, Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 131–137.

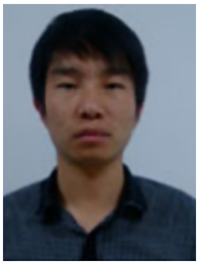
[34] M.S. Miri, A. Mahloojifar, Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction, IEEE Trans. Biomed. Eng. 58 (2011) 1183–1192.

[35] X. You, Q. Peng, Y. Yuan, Y.-m. Cheung, J. Lei, Segmentation of retinal blood vessels using the radial projection and semi-supervised approach, Pattern Recognit. 44 (2011) 2314–2324.

[36] J. Nagi, F. Ducatelle, G.A. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, L.M. Gambardella, Max-pooling convolutional neural networks for vision-based hand gesture recognition, in: Proceedings of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (IEEE2011), 2011, pp. 342–347.

Shuangling Wang received her BS in 2011 from the Shandong University, Weihai. Now she is studying at the Shandong University for a Master degree in computer application technology. Her main research interests are machine learning and biomedical image processing.

Yilong Yin is the Director of MLA Lab and a Professor of the Shandong University. He received his Ph.D. degree in 2000 from the Jilin University. From 2000 to 2002, he worked as a post-doctoral fellow in the Department of Electronic Science and Engineering, Nanjing University. His research interests include machine learning, data mining, and biometrics.



Guibao Cao received her BSEE in 2011 from the Harbin University of Commerce. Now he is studying at the Shandong University for a Master degree in computer application technology. His main research interests are pattern recognition, machine learning, and image processing.

Benzheng Wei received B.Sc., M.Sc. degrees in Computer Science and Technology from Shandong Institute of Light Industry and Shandong University, China in 2000 and 2007, respectively. Form 2010, he has been an Associate Professor in the College of Science and Technology, Shandong University of Traditional Chinese Medicine. His research interests focus on Medical image processing, Pattern analysis and Medical information processing.

Yuanjie Zheng is a senior research investigator in the Perelman School of Medicine at the University of Pennsylvania and the primary contact of Image Analysis Core at the Penn Vision Research Center. He received his Ph.D. degree in Pattern Recognition and Intelligent Systems from Shanghai Jiao Tong University of China in 2006. His current research interests are in computer vision, medical image analysis and clinical study.

Gongping Yang received his Ph.D. degree in computer software and theory from Shandong University, China, in 2007. Now he is a professor in School of Computer Science and Technology, Shandong University. His research interests are biometrics, medical image processing, and so forth.

Update

Neurocomputing

Volume 226, Issue , 22 February 2017, Page 270–272

DOI: https://doi.org/10.1016/j.neucom.2016.08.031

Neurocomputing 226 (2017) 270–272

Contents lists available at ScienceDirect

Neurocomputing

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

Corrigendum

Corrigendum to “Hierarchical retinal blood vessel segmentation based on feature and ensemble learning” [Neurocomputing 149 (2015) 708–717]

Shuangling Wang[a], Yilong Yin[a,*], Guibao Cao[a], Benzheng Wei[b], Yuanjie Zheng[c], Gongping Yang[a]

[a] *School of Computer Science and Technology, Shandong University, Jinan 250101, PR China*
[b] *College of Science and Technology, Shandong University of Traditional Chinese Medicine, Jinan 250355, PR China*
[c] *Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA*

ARTICLE INFO

Available online 14 September 2016

The authors regret “a programming error exists in their experiments which invalidates the conclusions of their paper. After modification, the performance of the proposed method is the third best on the DRIVE database in terms of accuracy, and remains the best on the STARE database. The detail of the programming error is as follows:

Table 3
Performance results on DRIVE Database images.

Image	*SN*	*SP*	*Acc*	*AUC*
1	0.8787	0.9818	0.9683	0.9637
2	0.8808	0.9793	0.9645	0.9651
3	0.8206	0.9694	0.9477	0.9352
4	0.8495	0.9769	0.9599	0.9505
5	0.7962	0.9681	0.9477	0.9347
6	0.7449	0.9581	0.9280	0.9430
7	0.8168	0.9721	0.9515	0.9382
8	0.8000	0.9715	0.9500	0.9415
9	**0.6674**	**0.9557**	**0.9219**	**0.9385**
10	0.8213	0.9758	0.9573	0.9434
11	0.8467	0.9772	0.9603	0.9367
12	0.8493	0.9784	0.9623	0.9601
13	0.7556	0.9597	0.9307	0.9432
14	0.8835	0.9846	0.9727	0.9626
15	0.8503	0.9827	0.9684	0.9301
16	0.7747	0.9661	0.9410	0.9489
17	0.7129	0.9596	0.9292	0.9276
18	0.8116	0.9755	0.9567	0.9545
19	**0.9032**	**0.9868**	**0.9764**	**0.9605**
20	0.8821	0.9859	0.9749	0.9709
Average	**0.8173**	**0.9733**	**0.9533**	**0.9475**

DOI of original article: http://dx.doi.org/10.1016/j.neucom.2014.07.059

[*] Corresponding author.
E-mail address: ylyin@sdu.edu.cn (Y. Yin).

http://dx.doi.org/10.1016/j.neucom.2016.08.031

Table 4
Performance results on STARE Database images.

Image	***SN***	***SP***	***Acc***	***AUC***
1	0.7985	0.9754	0.9558	0.9695
2	0.7515	0.9749	0.9541	0.9338
3	0.8502	0.9867	0.9754	0.9687
4	**0.5483**	**0.9481**	**0.9064**	**0.9623**
5	0.8246	0.9751	0.9563	0.9611
6	0.8833	0.9887	0.9793	0.9839
7	0.9416	0.9937	0.9879	0.9896
8	**0.9447**	**0.9943**	**0.9891**	**0.9884**
9	0.8955	0.9879	0.9778	0.9872
10	0.8879	0.9866	0.9755	0.9713
11	0.8918	0.9884	0.9788	0.9762
12	0.9107	0.9896	0.9811	0.9874
13	0.8789	0.9841	0.9710	0.9862
14	0.8592	0.9802	0.9649	0.9849
15	0.7851	0.9718	0.9494	0.9753
16	0.7387	0.9573	0.9260	0.9766
17	0.8889	0.9847	0.9727	0.9799
18	0.6451	0.9730	0.9498	0.9645
19	0.6311	0.9770	0.9562	0.9780
20	0.6523	0.9645	0.9355	0.9765
Average	**0.8104**	**0.9791**	**0.9621**	**0.9751**

Table 6
Comparison of segmentation results for individual and ensemble classifier on DRIVE images.

Method	***SN***	***SP***	***Acc***	***AUC***
CNN	0.6425	0.9478	0.9089	0.9378
RF-1	0.8173	0.9733	0.9533	0.9475
RF-2	0.7899	0.9692	0.9464	0.9476
RF-3	0.7576	0.9644	0.9381	0.9490
Average ensemble	0.7787	0.9676	0.9436	0.9521
Weighted ensemble one	0.7983	0.9704	0.9485	0.9514
Weighted ensemble two	0.8074	0.9717	0.9508	0.9502
Median ensemble	0.7959	0.9700	0.9478	0.9506
Winner-takes-all	**0.8173**	**0.9733**	**0.9533**	**0.9475**

Table 7
Comparison of segmentation results for individual and ensemble classifier on STARE images.

Method	***SN***	***SP***	***Acc***	***AUC***
CNN	0.6648	0.9626	0.9319	0.9364
RF-1	0.8104	0.9791	0.9621	0.9751
RF-2	0.7986	0.9778	0.9597	0.9768
RF-3	0.7714	0.9746	0.9546	0.9706
Average ensemble	0.7974	0.9777	0.9595	0.9777
Weighted ensemble one	0.8036	0.9783	0.9606	0.9771
Weighted ensemble two	0.8076	0.9788	0.9615	0.9777
Median ensemble	0.8009	0.9775	0.9596	0.9781
Winner-takes-all	**0.8104**	**0.9791**	**0.9621**	**0.9751**

Table 10
Cross-training results on drive and stare.

	SN	***SP***	***Acc***	***AUC***
Drive (trained on STARE)	0.8431	0.9574	0.9428	0.9271
STARE(trained on DRIVE)	0.7116	0.9675	0.9413	0.9391

Table 5
Performance measures on DRIVE and STARE.

Database	*Segmentation*	*SN*	*SP*	*Acc*	*AUC*
DRIVE	2nd human observer	0.7796	0.9717	0.9464	–
	Proposed Method	**0.8173**	**0.9733**	**0.9533**	**0.9475**
STARE	2nd human observer	0.8955	0.9382	0.9347	–
	Proposed Method	**0.8104**	**0.9791**	**0.9621**	**0.9751**

Table 8
Performance comparison of vessel segmentation methods on DRIVE images.

No	*Type*	*Methods*	*Year*	*SN*	*SP*	*Acc*	*AUC*
17.		**Proposed Method**	**2013**	**0.8173**	**0.9733**	**0.9533**	**0.9475**

Table 9
Performance comparison of vessel segmentation methods on stare images.

No	*Type*	*Methods*	*Year*	*SN*	*SP*	*Acc*	*AUC*
15.		**Proposed Method**	**2013**	**0.8104**	**0.9791**	**0.9621**	**0.9751**

Table 11
Performance comparison of results with cross training in terms of average accuracy.

Method	*Drive (trained on STARE)*	*STARE (trained on DRIVE)*
Proposed Method	**0.9428**	**0.9413**

When computing Accuracy, they tried to assess whether their result OR and the ground truth GT were the same by this equation: count=sum(sum((OR-GT)= =0)), they failed to notice that OR and GT that they used were unsigned integers, the negative values of (OR-GT) was automatically reassigned to 0, so all the accuracies in their paper are inaccurate, below are the list of changes (Tables 3, 4, 6,7):

They also declare a typo error of SP of Drive (trained on STARE) in Table 10, the value is changed to 0.9574 from 0.9774.

The main text and Tables 5, 8, 9, 11, in which the accuracy values referred to the above mentioned Tables must also be updated:

Section 1, paragraph 3: Experimental results show that our approach is competitive with state-of-the-art by achieving sensitivity/specificity/accuracy/AUC values of 0.8173/0.9733/0.9533/0.9475 for the DRIVE database and of 0.8104/0.9791/0.9621/0.9751 for the STARE database.

Section 5.4, paragraph 2: The best case accuracy, sensitivity, specificity, AUC for the DRIVE database are 0.976359, 0.903219, 0.986776, 0.960497, respectively, and the worst case measures are 0.921903, 0.667402, 0.955739 and 0.938519, respectively. The best case vessel segmentation result for the STARE database has an accuracy of 0.989061; sensitivity, specificity, AUC are 0.94471, 0.99432, and 0.988388, respectively. The worst case accuracy is 0.906358; sensitivity, specificity, and AUC are 0.548304,0.948144, and 0.962269, respectively".

The authors would like to apologise for any inconvenience caused.

Yilong Yin is the Director of MLA Lab and a Professor of the Shandong University. He received his pH.D. degree in 2000 from the Jilin University. From 2000 to 2002, he worked as a post-doctoral fellow in the Department of Electronic Science and Engineering, Nanjing University. His research interests include machine learning, data mining, and biometrics.