This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



# A Cross-modality Learning Approach for Vessel Segmentation in Retinal Images

Qiaoliang Li, *Member, IEEE*, Bowei Feng, LinPei Xie, Ping Liang*, Huisheng Zhang and Tianfu Wang

[1]

**Abstract—This paper presents a new supervised method for vessel segmentation in retinal images. This method remolds the task of segmentation as a problem of cross-modality data transformation from retinal image to vessel map. A wide and deep neural network with strong induction ability is proposed to model the transformation, and an efficient training strategy is presented. Instead of a single label of the center pixel, the network can output the label map of all pixels for a given image patch. Our approach outperforms reported state-of-the-art methods in terms of sensitivity, specificity and accuracy. The result of cross-training evaluation indicates its robustness to the training set. The approach needs no artificially designed feature and no preprocessing step, reducing the impact of subjective factors. The proposed method has the potential for application in image diagnosis of ophthalmologic diseases, and it may provide a new, general, high-performance computing framework for image segmentation.**

***Index Terms*—Cross-modality learning, retinal image, vessel segmentation, deep learning.**

## I. Introduction

Retinal fundus images have been widely used for diagnosis, screening and treatment of cardiovascular and ophthalmologic diseases [1], including age-related macular degeneration (AMD), diabetic retinopathy (DR), glaucoma, hypertension, arteriosclerosis and choroidal neovascularization, among which AMD and DR have been considered the two leading causes of blindness [2].

Vessel segmentation is a basic step required for the quantitative analysis of retinal fundus images [3]. The segmented vascular tree can be used to extract the morphological attributes of blood vessels, such as length, width, branching and angles. Moreover, the vascular tree has been adopted in multimodal retinal image registration [4] and retinal mosaic [5] as the most stable feature in the images. In [6], the vascular tree is also used for biometric identification due to its uniqueness.

Manual segmentation of the vascular tree in retinal images is a tedious task that requires experience and skill [3, 7]. In the development of a computer-assisted diagnostic system for ophthalmic disorders, automatic segmentation of retinal vessels has been accepted as a vital and challenging step [7]. The size, shape and intensity level of retinal vessels can vary hugely in different local areas. The width of a vessel often ranges from 1 to 20 pixels, depending on both the anatomical width of the vessel and the image resolution. The existence of vessel crossing, branching and centerline reflex makes it difficult [1] [8, 9] to segment the vessels accurately using artificially designed features. Pathologies in the form of lesions and exudates can further complicate the automatic segmentation.

In the past decades, several methods have been proposed for the segmentation of vessels in retinal images, and they can be divided into two categories: unsupervised and supervised methods.

*1) Unsupervised Methods*

The unsupervised methods use filter responses or other model-based techniques to extract vessels. According to the image processing methodologies, these methods can be classified into three sub-categories: matched filtering, vessel tracking and model-based approaches.

A matched filtering technique convolves a 2-D kernel with the retinal image, and the matched filter response indicates the presence of the vessel. The latest proposed method, presented by Azzopardi et al. [3], uses a combination of shifted filter responses (COSFIRE) to detect vessels, and the average accuracy, sensitivity and specificity are 0.9442, 0.7655 and 0.9704, respectively, on the DRIVE database and 0.9563, 0.7716 and 0.9701 for the STARE database. The matched filtering methodology works well for healthy images but is prone to increased false positive rates with pathological images [1]. Vessel tracking algorithms [10] use local information to segment a vessel between two points. The center of the longitudinal cross-section of a vessel is determined using gray-level intensity and tortuosity. This type of method can provide highly accurate vessel widths, but they often are unable to detect vessel segments that have no seed points. The model-based approaches apply the explicit vessel models, such as vessel profile models [11] and deformable models [12], to extract the retinal vessels. In profile models, the intensity of the vessel cross-section is considered as a Gaussian shape. Other profiles, such as the second-order derivative Gaussian [13], are also exploited to increase the segmentation accuracy with low image quality. The profile model opts to be compromised by

This work was supported by the Project of the National Science Foundation of China under Grants 61401285, 81401750, Guangdong Science Plan 2014A010103035，and Shenzhen Science Plan CYJ20140418182819179，JSGG20130411160504896, JCYJ20130408173226864.

Qiaoliang Li, Bowei Feng, Linpei Xie, Ping Liang, Huisheng Zhang and Tianfu Wang are with Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Department of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China. (Email: lql@szu.edu.cn, feng.bowei@163.com,xielinpei@email.szu.edu.cn, liangping@szu.edu.cn, isaac_zhs@126.com, tfwang@szu.edu.cn. *Corresponding author: Ping Liang* )

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



vessel crossing and branching. More complicated deformable models, such as parametric deformable models [12] and geometric deformable models [14], are also reported for vessel segmentation.

*2) Supervised Methods*

Supervised methods use extracted feature vectors to train a classifier to discriminate between vessel and non-vessel pixels. The algorithm can learn a set of rules of vessel extraction on the basis of a training set. The performance of supervised methods is usually better than that of unsupervised ones and can produce good results for healthy retinal images [1].

Most supervised methods adopt support vector machines (SVM) [15] or artificial neural networks (ANN) [16] as the classifier. Compared to SVM, a multilayer neural network can model more complicated relations between the input and output. In [16], Marin et al. presented a neural network-based supervised methodology. They simultaneously use gray-level and moment invariant-based features to build a 7-D feature vector and utilize a multilayer feed forward neural network for training and classification. The reported accuracy, sensitivity and specificity of their method are 0.9452, 0.7067 and 0.9801, respectively, on the DRIVE database and 0.9526, 0.6944 and 0.9819 for the STARE database. The decision tree has also been used in vessel classification. Fraz et al. [8] present a supervised method using an ensemble system of bagged and boosted decision trees. They utilize four techniques to extract the feature vector, including gradient vector field, morphological transformation, line feature and Gabor responses. The average accuracy, sensitivity and specificity are improved to 0.9480, 0.7406 and 0.9807 for DRIVE and 0.9534, 0.7548 and 0.9763 for STARE. Cheng et al. [17] use a random forest to fuse the rich information encoded in the hybrid context-aware features and achieve comparable performance to state-of-the-art methods [8, 16]. Most published supervised methods use artificially designed features to model the retinal vessel. However, the manual feature design is a heuristic and laborious procedure that is heavily dependent on experience and skills. Moreover, to address pathology, image noise and other complicated cases, the parameters used in the algorithm usually need to be careful adjusted.

In this paper, a new supervised method for vessel segmentation is presented. The segmentation task is remolded as a cross-modality data transformation problem. The first modality is the color retinal image, while the second modality is the vessel map. A wide and deep neural network is proposed to model the relations between the retinal image and the vessel map. The main contributions of this study include: (1) a novel neural network (that can output the label map of all pixels for the input image patch instead of just the single label of the center pixel) and the training algorithm; and (2) a synthesis strategy to construct the vessel probability map of the retinal image.

Our approach outperforms reported state-of-the-art methods in terms of sensitivity, specificity and accuracy. The result of cross-training evaluation indicates its robustness to the training set. Moreover, the method has two advantages: (1) no artificially designed features and no preprocessing steps are needed in the algorithm, reducing the impact of subjective factors, and (2) thanks to the powerful induction ability of the network, each pixel can be supported by multiple local neighborhoods, reducing the uncertainty caused by image noise and pathology. The method proposed in this paper benefits from the recent development of deep learning technology [18, 19]. Compared to traditional supervised methods, deep learning can achieve unsupervised learning of the optimal feature for a given task [20], improving the reliability of classification.

To the best of our knowledge, we are the first to regard the task of vessel segmentation as a cross-modality data transformation problem, and we have presented an efficient training technique to solve the problem. Theoretically, in addition to retina vessels, the proposed method should also work in the segmentation of other types of objects, such as layer segmentation in 3D OCT images, part segmentation in CT/MR volume, etc. As long as the input images and the corresponding label maps are provided, the network can automatically learn the underlying mapping function. Therefore, the proposed network may provide a new, general computing framework for image segmentation.

The paper is organized as follows. The proposed method is presented in Section II. In Section III, the performance metrics and the robustness of the algorithm are assessed by experiments. Finally, a discussion and conclusion are given in Section IV.

## II. PROPOSED METHOD

### A. Method Overview

In vessel segmentation, the traditional supervised method usually predicts the class label of each pixel in the image using the image intensity or artificially designed features. In such algorithms, the segmentation task is considered a problem of classification. Unlike the conventional supervised method [8, 16], we remold the task of segmentation as a cross-modality data transformation problem. The first modality is the color retinal image, while the second modality is the vessel map, the size of which is same as the retinal image. To be specific, given a retinal vessel image $x$, we attempt to use a mapping function $f(x; w)$ to obtain the corresponding vessel map $y$, where $w$ is the parameter set. Therefore, to complete the task of segmentation, we need to learn the underlying mapping function between the retinal images and the vessel maps.

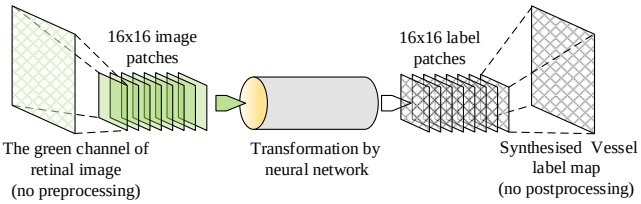



Fig. 1. The diagram of the proposed method.

The goal of cross-modality learning is to minimize

$$L = \sum_{i=1}^{N} l\left(y^i, f(x^i; w)\right) + \Phi(w), \quad (1)$$

where $x^i$ and $y^i$ are the i-th image and i-th label map in the training set, $l(\cdot)$ is the loss function, and $\Phi(w)$ is the regularization term, which penalizes the complexity of weights. In this study, we use a specifically designed neural network to model the mapping function between the retinal images and the vessel maps. With this neural network, the small patches

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



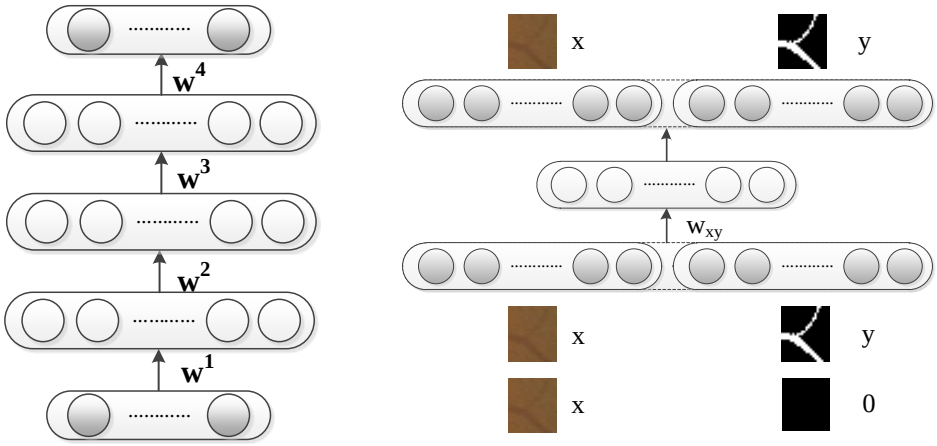



Fig. 2. The architecture of cross-modality learning for retinal image segmentation. (a) The five-layer neural network. (b) The denoising autoencoder model for pre-training.

extracted from the green channel of a retinal image can be directly transformed to vessel label patches with no preprocessing (Fig. 1). Subsequently, these vessel label patches can be synthesized into a complete vessel map with no post-processing.

### *B. Network Structure*

Here we attempt to predict all class labels of the pixels within an image patch at the same time, which demands that the neural network have a much stronger induction capability than that of traditional ones. As proved many times in the literature [18, 19], the complex relation between the input and output usually cannot be modeled by shallow networks. To extract a complete vessel map of the input retinal image, there must be enough intermediate layers of feature extraction in the network, and the multiple features can be simultaneously detected in each layer. Accordingly, in this study, we have specifically designed a deep neural network to model the mapping function. As shown in Fig. 2 (a), the proposed neural network contains a total of five layers with a sigmoid activation function. The input and output layers have 256 units, and the three hidden layers have 400 units.

### *C. Pre-training of the First Layer*

The proposed neural network has three hidden layers, and the unit number of a hidden layer is larger than that of the input layer. This type of neural network is difficult to directly train using the standard backward propagation algorithm. The traditional layer-wise training mechanism [19] is also unsuitable to train the network. The feature learnt by the autoencoder [19] or RBM [18] in layer-wise training is how to recover the input rather than how to transform the input into another space. In this study, we use a specifically designed two-stage procedure, pre-training of the first layer and overall training, to train the network efficiently.

#### *1) Theory of the Autoencoder*

The autoencoder model [21] is often used to build deep networks. An autoencoder takes an input vector $z$ and maps it to a hidden representation $h$, which is subsequently mapped back to $z'$, a 'reconstructed' version of $z$. The parameters of this model are optimized to minimize the reconstruction error:

$$\theta^* = \arg\min_\theta \frac{1}{n}\sum_{i=1}^{n} L(z, z') \quad (2)$$

where the loss function is the squared error $L(z, z) = \|z - z'\|^2$. In the denoising autoencoder, instead of $z$, a corrupted version of the input vector named $\tilde{z}$ is used as the input to train the autoencoder. The corruption is often conducted by randomly selecting a fixed number of components of $z$ and forcing their value to 0. The denoising autoencoder model has been proved to be able to reconstruct a clean input from a corrupted one.

#### *2) Concatenation of the Input and Output*

To use the characteristics of the denoising autoencoder to exploit the relations between the retinal image and label map, we have designed a training strategy based on concatenation of the input and output. Specifically, we concatenate the input image x and the vessel map y in an input vector $z^1 = [x, y]$ ($[\cdot]$ denotes concatenation), and, at the same time, we concatenate the input image x and map zero in another vector $z^2 = [x, 0]$. Because we attempt to obtain the label map given only the images, we set half of the training set to be $z^1$ and half to be $z^2$, while the output only consists of $z^1$. Thereby, the autoencoder is forced to learn the relations between the image and the label map (similar to recovering a clean input from a corrupted one). Suppose the image patch is N*N in size and the learnt weight of the first layer of the autoencoder is $w_{xy}$(Fig. 2 (b)); then, the weight of the first layer of the proposed network (Fig. 2 (a)) can be initialized by the first half of the learnt weight: $w^1 = w_{xy}(1: N^2)$.

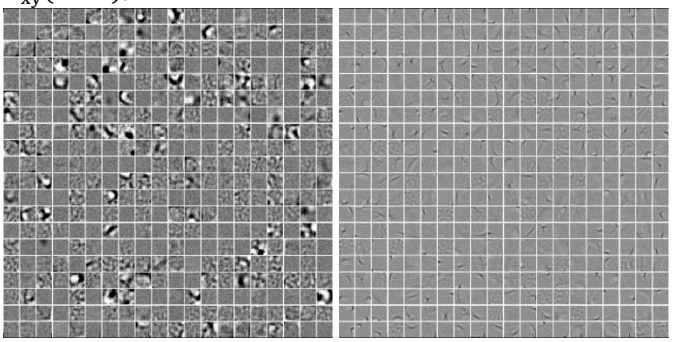


Fig. 3. The basis learnt by the encoder. (a) The parameters of the first layer learnt by pre-training. (b) The fine-tuned parameters of the first layer.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



As shown in Fig. 3(a), the denoising autoencoder has learnt a group of bases with different feature location and orientation. This pre-training procedure is important to the optimization of the network because directly learning the parameters of five layers is likely to converge in local minima. However, it must be noted that the pre-training used here is very different to unsupervised lay-wise pre-training [18] [19]. In the traditional initialization of a deep network, the unsupervised lay-wise pre-training learns a group of bases that can best reconstruct the input vector. In this study, the pre-training is designed to learn the mapping function $f(x; w)$, and it can be treated as an important step in moving forward from the original image to the vessel map.

### *D. Overall Training*

The parameters of the first layer are initialized by pre-training, and the parameters of the left layers are randomly initialized. Subsequently, all parameters of the network, $w^1 \sim w^4$, are fine-tuned using the retinal images as input, while the label maps are output. A standard backward propagation algorithm is used to train the network. The learning rate is set to 0.5, the batch size is 100, and the total number of training epochs is 30. For comparison, the tuned version of $w^1$ is shown in Fig. 3 (b).

### *E. Map Synthesis*

To complete the segmentation, we use a synthesis strategy to construct the probability map of the retinal image. In the traditional supervised method, the probability value of being vessels for each pixel is estimated using only one image patch surrounding it. In this study, our network model can output a label map in the same size as the input image patch, thereby each pixel could be 'supported' by multiple image patches in the neighborhood. Suppose that the total set of image patches composing pixel $c$ is $S$ and that the i-th patch is $b_i$; the probability value of pixel $c$ being a vessel is obtained by

$$\begin{aligned} P(c \in C_v) &= \sum_{i=1}^{N} P(c \in C_v | b_i) P(b_i) \\ &= \frac{1}{N} \sum_{i=1}^{N} P(c \in C_v | b_i), b_i \in S, \end{aligned} \quad (3)$$

where $N$ is the number of patches in space $S$. The resulting $P(c \in C_v)$ can be considered an average of the estimated probability value from all supporting patches. Alternately, we can formally regard $S$ as a space of image patches and $b_i$ as an element of $S$, and then the above equation is just a full probabilistic inference under the hypothesis that $P(c \in C_v | b_i) = 0, b_i \notin S$. In our method, $N$ is 256 because the patch size is fixed to 16. That is to say, the vessel probability of each pixel is deduced from 256 supporting patches.

## III. EXPERIMENTAL EVALUATION

### *A. Materials*

We have evaluated the method using three publicly available datasets, DRIVE, STARE and CHASE_DB1. The DRIVE dataset [22] contains a total of 40 color retinal images, in which 20 images are used as the training set and the other 20 images are used as the test set. Each image was captured at 565 $\times$ 584 pixels. For each image in DRIVE, the manually segmented vessel tree is provided together with a binary mask for the FOV area. The images in the training set are used to train the network, while the images in the test set are used for evaluation. The STARE data set [23] contains 20 retinal images, and each image was digitalized to 700 $\times$ 605 pixels. There is no separate training and test set available for this dataset. In the published literature, there are two types of techniques to conduct training on this dataset. The first is the 'random samples' technique, which builds a set of training samples randomly extracted from all 20 images. Although this 'random samples' approach has been used in most published methods [17, 24], there is overlap between the training and test samples in this technique, raising concern of overly optimistic results. The other technique is 'leave-one-out' [22] [25], in which each image is tested using a network model trained on the other 19 images. In this technique, the test and training samples have no overlap, and the network needs to be trained 20 times on different image sets. Currently, we have trained our model on STARE using a 'leave-one-out' approach. The CHASE_DB1 data set [1] contains 28 retinal images with a resolution of 999 $\times$ 960 pixels, collected from both the left and right eyes of 14 school children. We have divided this data set into 2 sets: the training set contains the first 20 images, while the test set contains the left 8 images. For all images of the training set in each dataset, we randomly extracted 30% of the image pixels as training samples to reduce the computation time of network training. Because the binary FOV mask is not present for the STARE and CHASE_DB1 datasets, we have manually created the masks for these two datasets as described in [25]. Note that, in this study, the input patches are randomly sampled over the entire image (including those outside the FOV), and we expect that the network can learn to discriminate the vessels from the FOV edges.

### *B. Evaluation Criterion*

In vessel segmentation, there are two class labels: vessel and non-vessel. By comparing the segmentation results with the manual ground truth, we obtain four measures: the vessel pixels that are predicted as vessels are denoted as true positives (TP), the vessel pixels that are predicted as non-vessels are denoted as false negatives (FN), the non-vessel pixels that are predicted as non-vessels are denoted as true negatives (TN), and the non-vessel pixels that are predicted as vessels are denoted as false positives (FP).

We use three criterions to compare the performance of the proposed method with other state-of-the-art methods: the sensitivity (Se), specificity (Sp) and accuracy (Acc). The evaluation metrics are calculated only for pixels inside the FOV. These metrics are defined as

$$Se = \frac{TP}{TP+FN},$$

$$Sp = \frac{TN}{TN+FP},$$

$$Acc = \frac{TP+TN}{TP+FN+TN+FP}.$$

Because the quantitative measures are dependent on the threshold on the probability map, a received operating

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



characteristic (ROC) curve is plotted with the true positive fraction (Se) versus the false positive fraction (1-Sp) by varying the threshold. The performance of the methods is also evaluated with the area under the ROC curve (AUC). The AUC will be equal to 1 when the classifier is perfect.

### C. Network Size Evaluation

The unit number (width) of the hidden layer has a great impact on the overall performance of the network. Here we use the DRIVE dataset to obtain an optimal configuration of the network size. As shown in Fig. 4, when the width increases from 200 to 700, the training error decreases from 3.1 to 1.82; however, the overall performance measure, AUC on the test set, first increases from 0.957 to 0.975 and then decreases to 0.961. To obtain an accurate model with good generalization ability, we set the unit number of the hidden layer to be 400 in the experiments.

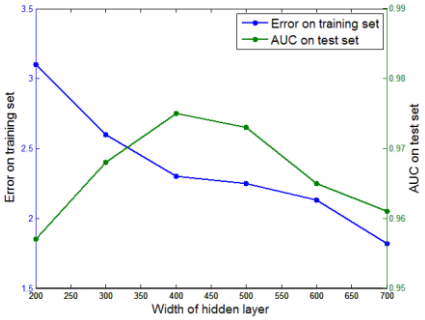



Fig. 4. Width evaluation of the hidden layers.

### D. Experimental Results

To evaluate the proposed method, we calculate the Se, Sp, Acc and AUC for the DRIVE, STARE and CHASE_DB1 datasets. Furthermore, we use a cross-training strategy to test the feasibility of the proposed method for application to a retinal image without a specific training set, proving its potential for use under realistic conditions. Finally, we show the results for ten pathological images.

*1) Vessel Segmentation*

For each retinal image in the test set, we obtain a probability map using the network model trained on the corresponding training set. The binary vessel segmentation is conducted using an optimized threshold on the probability map, and the evaluation metrics are calculated by taking the manual results from the first observer as ground truth. The optimized threshold is set as the one at which the maximum Acc on the training set is achieved. The average results of the evaluation metrics for the DRIVE, STARE and CHASE_DB1 datasets are shown in Table I.

TABLE I.
PERFORMANCE MEASURES FOR DRIVE, STARE AND CHASE_DB1.

| Dataset | Method | Se | Sp | Acc | AUC |
|---|---|---|---|---|---|
| DRIVE | Proposed method | 0.7569 | 0.9816 | 0.9527 | 0.9738 |
| | 2nd Human Observer | 0.7760 | 0.9724 | 0.9472 | - |
| STARE | Proposed method | 0.7726 | 0.9844 | 0.9628 | 0.9879 |
| | 2nd Human Observer | 0.8952 | 0.9384 | 0.9349 | - |
| CHASE_DB1 | Proposed method | 0.7507 | 0.9793 | 0.9581 | 0.9716 |
| | 2nd Human Observer | 0.8105 | 0.9711 | 0.9545 | - |

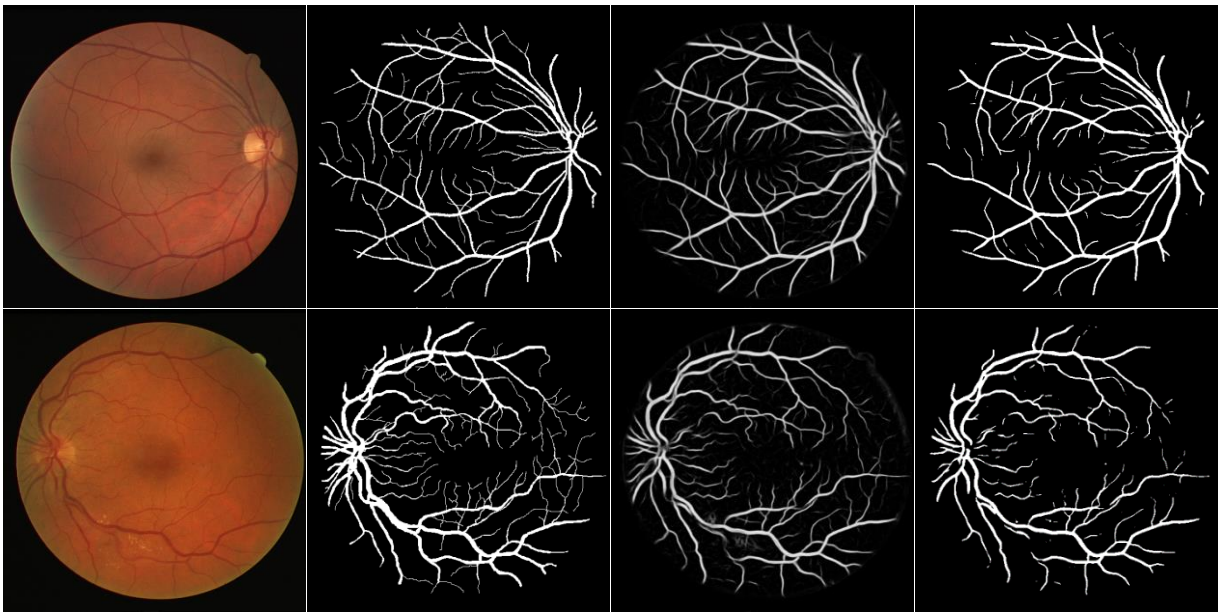


Fig. 5. Segmentation results for the DRIVE dataset. The first through fourth columns are the retinal image, manual segmentation, probability map and final segmentation result. The first row is the best accuracy, and the second row is the worst accuracy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



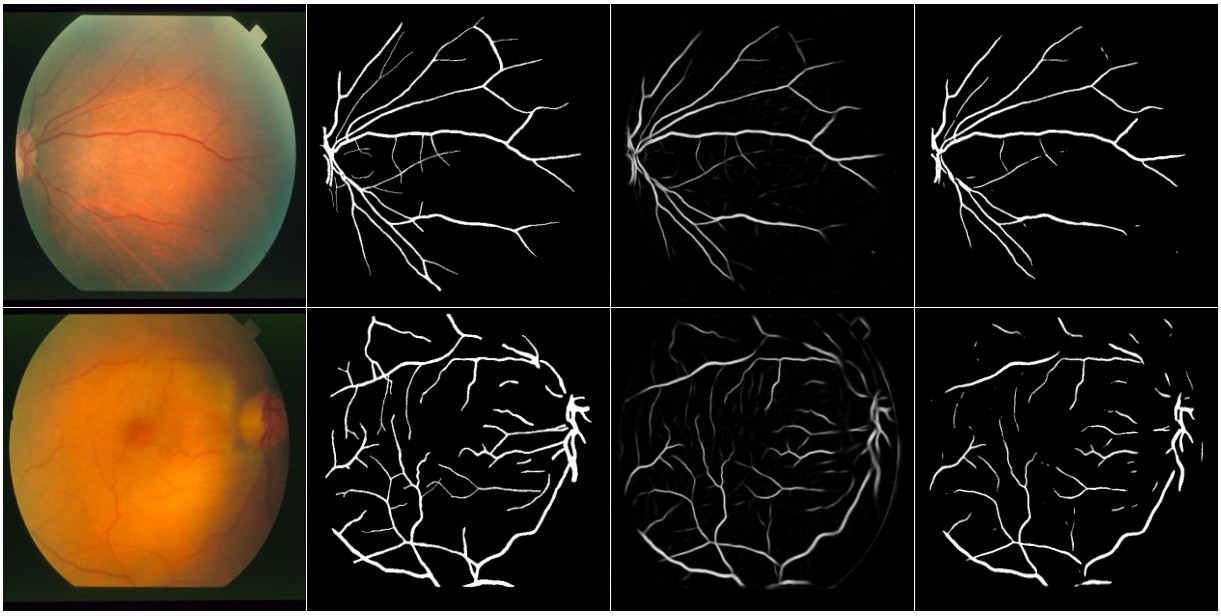


Fig. 6. Segmentation results for STARE dataset.

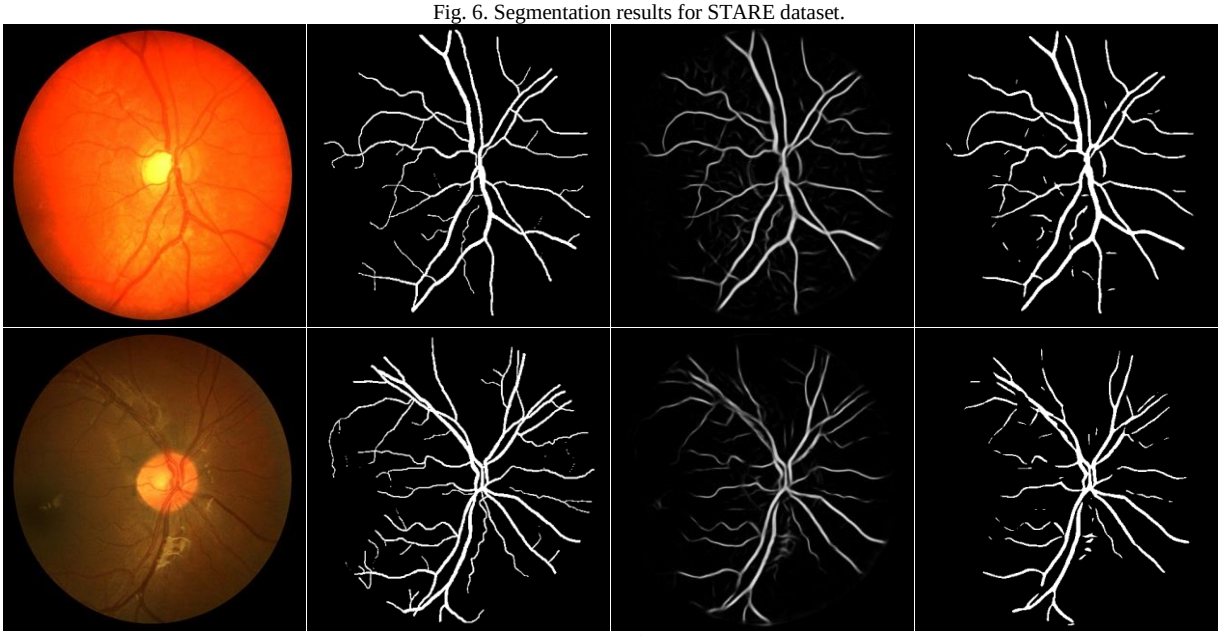


Fig. 7. Segmentation results for the CHASE_DB1 dataset.

TABLE II.

COMPARISON WITH STATE-OF-THE-ART METHODS ON THE DRIVE DATASET

| No | Type | Methods | Year | Se | Sp | Acc | AUC |
|---|---|---|---|---|---|---|---|
| 1 | Unsupervised methods | Zana [26] | 2001 | 0.6971 | N.A | 0.9377 | 0.8984 |
| 2 | | Jiang [27] | 2003 | N.A | N.A | 0.9212 | 0.9114 |
| 3 | | Mendonca [28] | 2006 | 0.7344 | 0.9764 | 0.9452 | N.A |
| 4 | | Al-Diri [29] | 2009 | 0.7282 | 0.9551 | N.A | N.A |
| 5 | | Lam [30] | 2010 | N.A | N.A | 0.9472 | 0.9614 |
| 6 | | Miri [31] | 2011 | 0.7352 | 0.9795 | 0.9458 | N.A |
| 7 | | Fraz [9] | 2011 | 0.7152 | 0.9759 | 0.9430 | N.A |
| 8 | | You[32] | 2011 | 0.7410 | 0.9751 | 0.9434 | N.A |
| 9 | | Azzopardi [3] | 2015 | 0.7655 | 0.9704 | 0.9442 | 0.9614 |
| 10 | Supervised methods | Niemeijer [33] | 2004 | N.A | N.A | 0.9416 | 0.9294 |
| 11 | | Soares[25] | 2006 | 0.7332 | 0.9782 | 0.9461 | 0.9614 |
| 12 | | Staal [22] | 2004 | N.A | N.A | 0.9441 | 0.9520 |
| 13 | | Ricci [24] | 2007 | N.A | N.A | 0.9595 | 0.9558 |
| 14 | | Lupascu [34] | 2010 | 0.7200 | N.A | 0.9597 | 0.9561 |
| 15 | | Marin [16] | 2011 | 0.7067 | 0.9801 | 0.9452 | 0.9588 |
| 16 | | Fraz [8] | 2012 | 0.7406 | 0.9807 | 0.9480 | 0.9747 |
| 17 | | Cheng [17] | 2014 | 0.7252 | 0.9798 | 0.9474 | 0.9648 |
| 18 | | **Proposed method** | **2015** | **0.7569** | **0.9816** | **0.9527** | **0.9738** |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



TABLE III.
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE STARE DATASET

| No | Type | Methods | Year | Se | Sp | Acc | AUC |
|---|---|---|---|---|---|---|---|
| 1 | Unsupervised methods | Hoover | 2000 | 0.6747 | 0.9565 | 0.9264 | N.A |
| 2 | | Jiang[27] | 2003 | N.A | N.A | 0.9009 | N.A |
| 3 | | Mendonca[28] | 2006 | 0.6996 | 0.9730 | 0.9440 | N.A |
| 4 | | Al-Diri[29] | 2009 | 0.7521 | 0.9681 | N.A | N.A |
| 5 | | Lam[30] | 2010 | N.A | N.A | 0.9567 | 0.9739 |
| 7 | | Fraz[9] | 2011 | 0.7311 | 0.9680 | 0.9442 | N.A |
| 8 | | You[32] | 2011 | 0.7260 | 0.9756 | 0.9497 | N.A |
| 9 | | Azzopardi [3] | 2015 | 0.7716 | 0.9701 | 0.9563 | 0.9497 |
| 10 | Supervised methods | Staal[22] | 2004 | N.A | N.A | 0.9516 | 0.9614 |
| 11 | | Soares[25] | 2006 | 0.7207 | 0.9747 | 0.9479 | 0.9671 |
| 12 | | Ricci[24] | 2007 | N.A | N.A | 0.9584 | 0.9602 |
| 13 | | Fraz [8] | 2012 | 0.7548 | 0.9763 | 0.9534 | 0.9768 |
| 14 | | **Proposed method** | **2015** | **0.7726** | **0.9844** | **0.9628** | **0.9879** |

TABLE IV.
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE CHASE_DB1 DATASET

| No | Type | Methods | Year | Se | Sp | Acc | AUC |
|---|---|---|---|---|---|---|---|
| 1 | Unsupervised methods | Azzopardi [3] | 2015 | 0.7585 | 0.9587 | 0.9387 | 0.9487 |
| 2 | Supervised methods | Fraz [8] | 2012 | 0.7224 | 0.9711 | 0.9469 | 0.9712 |
| 3 | | **Proposed method** | **2015** | **0.7507** | **0.9793** | **0.9581** | **0.9716** |

The AUC produced by the proposed method are greater than 0.97 for all three datasets (Fig. 8). The sensitivity, specificity and accuracy of the best case for the DRIVE dataset are 0.8737, 0.9754 and 0.9631, respectively, and the values of the worst case are 0.6734, 0.9872 and 0.9415 (Fig. 5). For the STARE dataset, the values of the best case are 0.7389, 0.9950 and 0.9771, and the values of the worst case are 0.5759, 0.9959 and 0.9529 (Fig. 6). For the CHASE_DB1 dataset, the values of the best case are 0.8548, 0.9793 and 0.9677, and the values of the worst case are 0.6298, 0.9879 and 0.9484 (Fig. 7).

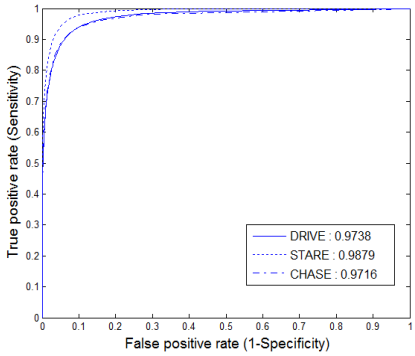



Fig. 8. ROC curves for the DRIVE, STARE and CHASE_DB1datasets.

*2) Comparison with the State-of-the-art*

The results of the proposed method are compared to those of the state-of-the-art methods in Tables II, III and IV for the DRIVE, STARE and CHASE_DB1 datasets, respectively. In general, the proposed method achieves better performance than the state-of-the-art methods regarding the sensitivity, specificity and accuracy jointly. The AUC achieved by the proposed method for the three data sets are 0.9738, 0.9879 and 0.9716, respectively. The AUC values are higher than those of published methods on the STARE and CHASE_DB1datasets and slightly lower than [8] on the DRIVE dataset. For the STARE dataset, the often-used 'random samples' approach [17, 24] will raise concerns of overly optimistic results. In this study, we have used a 'leave-one-out' approach [22] [25] to evaluate the method on this dataset, and our method obtains satisfactory results compared to those of published methods. A comparative analysis of the results is presented in Section IV.

*3) Cross-Training*

The network is cross-trained to test the capability of the proposed method in realistic applications. Specifically, the network is tested for one dataset after being trained on the other two datasets. The results of cross-training for the three datasets are shown in Table V. The Acc results are decreased from 0.9527, 0.9628 and 0.9581 to 0.9485, 0.9535 and 0.9417 for the DRIVE, STARE and CHASE_DB1 datasets when they are trained on the STARE, CHASE_DB1 and DRIVE datasets, respectively. The average decrease value of Acc is 0.009, which is close to that reported in [8]. As shown in Table V, we have achieved better performance in cross-training compared to other published methods [16] [8] [25] [24].

Table V.
PERFORMANCE MEASURE OF CROSS-TRAINING

| Dataset | Segmentation | Se | Sp | Acc | AUC |
|---|---|---|---|---|---|
| DRIVE | Trained on STARE [25] | - | - | 0.9397 | - |
| | Trained on STARE [24] | - | - | 0.9266 | - |
| | Trained on STARE [16] | - | - | 0.9448 | - |
| | Trained on STARE [8] | 0.7242 | 0.9792 | 0.9456 | 0.9697 |
| | **Trained on STARE (Proposed method)** | **0.7273** | **0.9810** | **0.9486** | **0.9677** |
| | **Trained on CHASE_DB1 (Proposed method)** | **0.7307** | **0.9811** | **0.9484** | **0.9605** |
| | | | | | |
| STARE | Trained on DRIVE [25] | - | - | 0.9327 | - |
| | Trained on DRIVE [24] | - | - | 0.9464 | - |
| | Trained on DRIVE [16] | - | - | 0.9528 | - |
| | Trained on DRIVE [8] | 0.7010 | 0.9770 | 0.9495 | 0.9660 |
| | **Trained on DRIVE (Proposed method)** | **0.7027** | **0.9828** | **0.9545** | **0.9671** |
| | **Trained on CHASE_DB1 (Proposed method)** | **0.6944** | **0.9831** | **0.9536** | **0.9620** |
| | | | | | |
| CHASE _DB1 | Trained on STARE [8] | 0.7103 | 0.9665 | 0.9415 | 0.9565 |
| | **Trained on STARE (Proposed method)** | **0.7240** | **0.9768** | **0.9417** | **0.9553** |
| | **Trained on DRIVE (Proposed method)** | **0.7118** | **0.9791** | **0.9429** | **0.9628** |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



*4) Pathological Cases*

To compare the performance of the proposed method in pathological cases, we specifically compute the performance measure of ten pathological images from the STARE dataset (Table VI).

TABLE VI.
PERFORMANCE MEASURE OF PATHOLOGICAL RETINAL IMAGES.

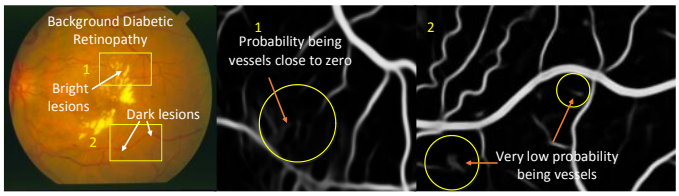
| Segmentation | Se | Sp | Acc |
|---|---|---|---|
| Hoover [23] | 0.6587 | 0.9565 | 0.9258 |
| Soares [25] | 0.7181 | 0.9765 | 0.9500 |
| Fraz [8] | 0.7262 | 0.9764 | 0.9511 |
| **Proposed** | **0.7800** | **0.9805** | **0.9672** |

## IV. DISCUSSION AND CONCLUSION

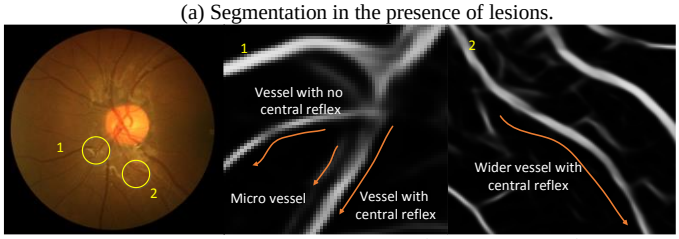
In this paper, we have presented a new supervised approach for vessel segmentation in retinal images. The segmentation task is remolded as a cross-modality data transformation problem. The proposed network can automatically learn the vascular feature in the training procedure. Thereby, no artificially designed features are needed in the segmentation, reducing the impact of subjective factors. Furthermore, the proposed network is wide and deep and has more a powerful ability for induction than general neural networks have. The network can directly output the label map of all pixels in a given image. In the prediction process, each pixel can be supported by multiple neighborhoods, reducing the uncertainty brought by image noise and pathology compared to one-time prediction.

When jointly regarding the performance measures of sensitivity, specificity and accuracy, our approach outperforms state-of-the-art methods on the DRIVE, STARE and CHASE_DB1 datasets. If we consider only the accuracy measure Acc, the values achieved by this method on the STARE and CHASE_DB1 data sets are 0.9628 and 0.9581, higher than those of published algorithms. For the DRIVE dataset, the Acc is 0.9527, slightly lower than Ricci's 0.9595. It has been demonstrated that the Ricci [24] method is very sensitive to the training set. In cross-training, its accuracy decreased from 0.9595 to 0.9266, while the Acc decreased from 0.9527 to 0.9485 for the proposed method. It appears that this method is more robust to the training set. This characteristic is very important for a segmentation algorithm in clinical applications because it makes it possible to apply the trained model to new images without re-training. In the cross-training, the performance measure of this method is significantly improved compared to Ricci's [24] and slightly improved compared to the latest methods publish by Fraz [8] and Marin [16].

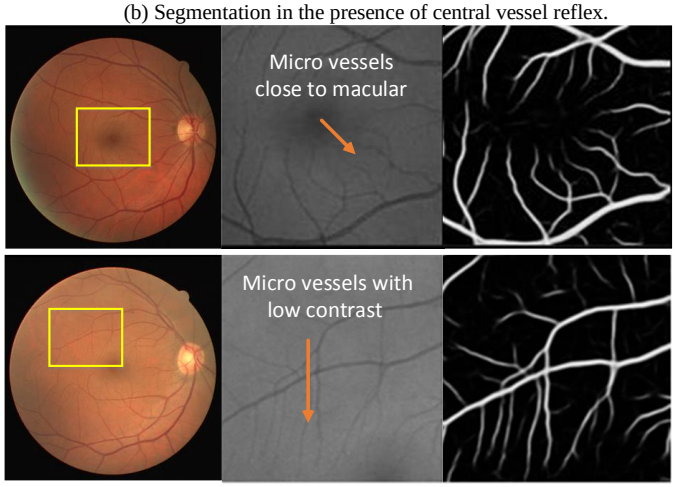
The AUC of ROC curves is also used to evaluate the performance of the method. In retinal images, the number of vessel pixels is much fewer than that of the background. The AUC of the ROC curve has been accepted as an appropriate accuracy measure in the classification of objects with an unbalanced distribution [1]. Considering the measure AUC, the values reached by this method for the three data sets are 0.9738, 0.9879 and 0.9716, respectively. The AUC of this method is higher than published methods on the STARE and CHASE_DB1 datasets and slightly lower than the result of 0.9747 on DRIVE that was produced by [8]. Note that the Se and Acc of this method on DRIVE are 0.7569 and 0.9527, higher than the 0.7406 and 0.9480 reported in [8]. Therefore, this method tends to obtain more accurate segmentation results on DRIVE compared to [8].




(a) Segmentation in the presence of lesions.




(b) Segmentation in the presence of central vessel reflex.




(c) Segmentation of micro vessels with low contrast.

Fig. 9. Segmentation in the presence of lesions, central vessel reflex and micro vessels.

As numerous methodologies are available for retinal vessel segmentation, the active problems are: 1) segmentation in the presence of lesions; 2) segmentation in the presence of central vessel reflex; and 3) segmentation of micro vessels with low contrast. In Fig. 9 (a), we have shown an image with diabetic retinopathy, where bright and dark lesions are both present. Compared to the true vessels, the probability value of the lesions produced by the proposed method is significantly lower for both bright and dark lesions. Therefore, the segmentation results could be expected to be less influenced by the presence of lesions. To show the effectiveness of our method in the presence of central vessel reflex, we have selected one image from the CHASE dataset (Fig. 9 (b)). The vessels in area '1' and area '2' have central reflex. In area '1,' the micro vessels, vessels with no central reflex and vessels with central reflex are present at the same time, while there are only vessels with central reflex in area '2.' The vessels in area '1' and area '2' are all correctly recognized by the proposed algorithm. In Fig. 9 (c), we have selected micro vessels surrounding the macular and micro vessels with very low

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2015.2457891, IEEE Transactions on Medical Imaging



contrast from two images in the DRIVE dataset. The macular has low image intensity and is close to the intensity of vessels in the green channel. Nevertheless, the micro vessels surrounding the macular are correctly classified, while the entire macular is recognized as a non-vessel. Furthermore, the micro vessels with very low contrast are also classified with high accuracy, indicating the ability of the method for vessel segmentation in the presence of low-contrast micro vessels.

In previous work [8][23][25], the performance of the segmentation method has been objectively evaluated for pathological images, and our algorithm also has better performance in handling pathological cases. From Table VI, it is observed that the accuracy of the method is 0.9672, higher than the best results of 0.9511 reported by Fraz [8]. Limited by the currently available datasets, the evaluation is conducted on ten pathological images in this study. In the future, we will build a new dataset that contains retinal images from patients of diabetic retinopathy, glaucoma and other ophthalmic diseases, and the new dataset can be used to evaluate the capability of the algorithms in handling pathological images.

Compared to other supervised methods [8, 16, 17, 24], our method can achieve unsupervised learning of the optimal feature for vessel detection. Because the learning process is forced to exploit the intrinsic characteristics of vessels in all images contained in the training set, the learnt feature could be expected to be more reliable to pathology, noise and different imaging conditions. The manual design of a feature is a heuristic and laborious procedure that requires experience and skills. The parameters, and occasionally the flowchart of the algorithm, need to be carefully adjusted to address complicated cases. For our approach, the model can be updated by re-training based on the original one to address new cases, while the algorithm remains unchanged.

The number of samples used in training is 30% of the total number of pixels in the training set. The percent value is close to that of the Soares [25] method. However, in the Lupascu [34] and Fraz [8] methods, less than 10% of the total pixels are used as training samples. Because we use a deep and wide network, the freedom of which is very large, there must be sufficient training samples to obtain a correct convergence. Fortunately, the convergence is related to the absolute number of training samples, and a smaller percent of samples could be expected to be enough if a large number of images are provided for training.

In theory, in addition to blood vessels, the proposed neural network can also be used in the segmentation of other types of object. For a given segmentation task, we only need to input image samples to the network while outputting the results of the label map. An appropriate model can be trained using the presented pre-training and fine-tuned approach. Moreover, the input of the network is in one dimension, and any 2D image or 3D voxel can be rearranged in 1D to use this model. Therefore, this method may provide a new, general, high-performance framework for image segmentation. We will explore the use of this network model in 3D vessel segmentation, specific tissue localization and other tasks in the near future.

The average time required to segment one retinal image is approximately 70 seconds with an AMD Athlon II X4 645 CPU running at 3.10 GHz with 4 GB of RAM. This method is currently implemented using MATLAB 2014a, and the computational performance could be further improved. According to Table VII, the processing time of this method is close to recently proposed supervised methods. The average time required to train the network for one dataset is approximately 7 hours. In comparison, the GMM classifier used by Soares [25] is trained in approximately 8 hours. Some feature-based supervised methods, such as those of Fraz [8] and Marin [16], need only several minutes to train the classifier because the freedom of the parameter space has been largely reduced in the feature extraction. Nevertheless, the training procedure can be performed offline and can be accelerated by GPU programming.

TABLE VII.
AVERAGE TIME FOR PROCESSING ONE IMAGE.

| Type | Method | Year | Processing time |
|---|---|---|---|
| Unsupervised methods | Jiang [27] | 2003 | 10 s |
| | Mendonca [28] | 2006 | 2.5 min |
| | Al-Diri[29] | 2009 | 11 min |
| | Azzopardi [3] | 2015 | 10 s |
| Supervised methods | Staal [22] | 2004 | 15 min |
| | Soares [25] | 2006 | 3 min |
| | Marin [16] | 2011 | 1.5 min |
| | Fraz [8] | 2012 | 2 min |
| | **Proposed method** | **2015** | **1.2 min** |

The resolution of the datasets used in our work is less than 1000 pixels, both in width and height. For high-resolution datasets, such as HRF [35], the vessel width can be 30 pixels or larger. An image patch with a size of 16x16 may fall into the flat area inside the vessel, and it also may fall into the background, but the labels of these two areas are completely contrary: full 1 or full 0. To guarantee correct convergence, a larger patch size is suggested to be used for high-resolution datasets. Therefore, more storage space and a longer training time are required when a high-resolution dataset is introduced because our model is fully connected. This aspect could be considered as a potential limitation of the method. However, we know that, theoretically, both the ANN and the convolutional neural network (CNN) can be used to learn the relation between the retinal image and the label map. The CNN has shared weights, and it may reduce the number of weights compared to the ANN for the same task. How to train a special CNN that can simultaneously predict all labels of the pixels in one retinal image patch is still unknown. We will explore this aspect in future work, and we also plan to establish a theoretical framework for efficient training of this type of neural network, for both ANNs and CNNs, benefiting its use in medical image analysis.

In conclusion, we have presented a novel supervised vessel segmentation approach for retinal images. The method has the potential for application in retinal image diagnosis systems thanks to its high accuracy and robustness.

## V. ACKNOWLEDGMENTS

The authors would like to thank J. J. Staal, A. Hoover, M. M. Fraz and their colleagues for making their databases publicly available. The authors would also like to thank Doctor

Fang from Nanshan Hospital of Shenzhen for providing guidance on retinal images.

## REFERENCES


[1] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, *et al.*, "Blood vessel segmentation methodologies in retinal images - A survey," *Computer Methods and Programs in Biomedicine,* vol. 108, pp. 407-433, Oct 2012.

[2] M. D. Abramoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, *et al.*, "Automated Analysis of Retinal Images for Detection of Referable Diabetic Retinopathy," *Jama Ophthalmology,* vol. 131, pp. 351-357, Mar 2013.

[3] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COSFIRE filters for vessel delineation with application to retinal images," *Medical Image Analysis,* vol. 19, pp. 46-57, Feb 2015.

[4] F. Zana and J. C. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform," *IEEE Transactions on Medical Imaging,* vol. 18, pp. 419-428, May 1999.

[5] C. Ruggiero, S. Benvenuti, and M. Giacomini, "Mathematical modeling of retinal mosaic formation by mechanical interactions and dendritic overlap," *IEEE Transactions on Nanobioscience,* vol. 6, pp. 180-185, Jun 2007.

[6] C. Marino, M. G. Penedo, M. Penas, M. J. Carreira, and F. Gonzalez, "Personal authentication using digital retinal images," *Pattern Analysis and Applications,* vol. 9, pp. 21-33, May 2006.

[7] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *Acm Computing Surveys,* vol. 36, pp. 81-121, Jun 2004.

[8] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, *et al.*, "An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation," *IEEE Transactions on Biomedical Engineering,* vol. 59, pp. 2538-2548, Sep 2012.

[9] M. M. Fraz, S. A. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, *et al.*, "An approach to localize the retinal blood vessels using bit planes and centerline detection," *Computer Methods and Programs in Biomedicine,* vol. 108, pp. 600-616, Nov 2012.

[10] I. C. Liu and Y. Sun, "Recursive Tracking of Vascular Networks in Angiograms Based on the Detection Deletion Scheme," *IEEE Transactions on Medical Imaging,* vol. 12, pp. 334-341, Jun 1993.

[11] B. S. Y. Lam and H. Yan, "A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields," *IEEE Transactions on Medical Imaging,* vol. 27, pp. 237-246, Feb 2008.

[12] L. Espona, M. J. Carreira, M. G. Penedo, and M. Ortega, "Retinal Vessel Tree Segmentation using a Deformable Contour Model," *19th International Conference on Pattern Recognition, Vols 1-6,* pp. 2128-2131, 2008.

[13] L. Gang, O. Chutatape, and S. M. Krishnan, "Detection and measurement of retinal vessels, in fundus images using amplitude modified second-order Gaussian filter," *IEEE Transactions on Biomedical Engineering,* vol. 49, pp. 168-172, Feb 2002.

[14] K. W. Sum and P. Y. S. Cheung, "Vessel extraction under non-uniform illumination: A level set approach," *IEEE Transactions on Biomedical Engineering,* vol. 55, pp. 358-360, Jan 2008.

[15] A. Osareh and B. Shadgar, "Automatic Blood Vessel Segmentation in Color Images of Retina," *Iranian Journal of Science and Technology Transaction B-Engineering,* vol. 33, pp. 191-206, Apr 2009.

[16] D. Marin, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo, "A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features," *IEEE Transactions on Medical Imaging,* vol. 30, pp. 146-158, Jan 2011.

[17] E. K. Cheng, L. Du, Y. Wu, Y. J. Zhu, V. Megalooikonomou, and H. B. Ling, "Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features," *Machine Vision and Applications,* vol. 25, pp. 1779-1792, Oct 2014.

[18] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science,* vol. 313, pp. 504-507, Jul 28 2006.

[19] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, pp. 1798-1828, Aug 2013.

[20] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *Journal of Machine Learning Research,* vol. 11, pp. 3371-3408, Dec 2010.

[21] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, *et al.*, "Higher Order Contractive Auto-Encoder," *Machine Learning and Knowledge Discovery in Databases, Pt Ii,* vol. 6912, pp. 645-660, 2011.

[22] 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 Transactions on Medical Imaging,* vol. 23, pp. 501-509, Apr 2004.

[23] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging,* vol. 19, pp. 203-210, Mar 2000.

[24] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Transactions on Medical Imaging,* vol. 26, pp. 1357-1365, Oct 2007.

[25] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "ri*Imaging*, vol. 25, pp. 1214-1222, Sep 2006.

[26] F. Zana and J. C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Transactions on Image Processing*, vol. 10, pp. 1010-1019, Jul 2001.

[27] X. Y. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 131-137, Jan 2003.

[28] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1200-1213, Sep 2006.

[29] B. Al-Diri, A. Hunter, and D. Steel, "An Active Contour Model for Segmenting and Measuring Retinal Vessels," *IEEE Transactions on Medical Imaging*, vol. 28, pp. 1488-1497, Sep 2009.

[30] B. S. Y. Lam, Y. S. Gao, and A. W. C. Liew, "General Retinal Vessel Segmentation Using Regularization-Based Multiconcavity Modeling," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 1369-1381, Jul 2010.

[31] M. S. Miri and A. Mahloojifar, "Retinal Image Analysis Using Curvelet Transform and Multistructure Elements Morphology by Reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 1183-1192, May 2011.

[32] X. G. You, Q. M. Peng, Y. Yuan, Y. M. Cheung, and J. J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," *Pattern Recognition*, vol. 44, pp. 2314-2324, Oct-Nov 2011.

[33] 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," *Medical Imaging 2004: Image Processing, Pts 1-3*, vol. 5370, pp. 648-656, 2004.

[34] C. A. Lupascu, D. Tegolo, and E. Trucco, "FABC: Retinal Vessel Segmentation Using AdaBoost," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp. 1267-1274, Sep 2010.

[35] J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, *et al.*, "Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database," *IET Image Processing*, vol. 7, pp. 373-383, Jun 2013.