Ground truth free retinal vessel segmentation by learning from simple pixels

Beiji Zou\(^1\) | Hongpu Fu\(^1,2\) | Zailiang Chen\(^1\) | Qing Liu\(^1\)

\(^1\) School of Computer Science, Central South University, Changsha, China
\(^2\) Hunan First Normal University, Changsha, China

Correspondence
Qing Liu, School of Computer Science, Central South University, Changsha, China.
qing.liu.411@gmail.com

Abstract
Retinal vessel segmentation is fundamental for the automatic retinal image analysis and ocular disease screening. This paper aims to learn a ground truth free feature aggregation strategy for the vessel segmentation. Five vesselness maps modelling the vessels’ profile, appearance, and shape are first generated. Together, the histogram of the local binary pattern and the green colour are extracted. In each vesselness map, the pixels with large vesselness values are regarded as simple positive samples. The pixels with small vesselness values are regarded as simple negative samples, and the pixels with mediocre values are treated as difficult pixels. The simple positive samples and simple negative samples near the difficult pixels consist of the training dataset while the rest vesselness maps together with the local binary pattern histogram, and green colour channel are used as the features to learn a strong classifier. Then, without leveraging any ground truth, multiple kernel boosting is used to combine four support vector machine kernels to learn a strong vessel model for each image. Applying this learnt model to the pixels with mediocre values in the single vesselness map, their label will be determined. Totally, five strong vessel models are learnt. Finally, pixels with the majority supports from the strong vessel models are labelled as vessel pixels. The proposed method achieves accuracy of 94.83%, sensitivity of 72.59%, and specificity of 98.11% on DRIVE dataset, and accuracy of 95.51%, sensitivity of 78.09%, and specificity of 97.56% on STARE. It outperforms the state-of-the-art unsupervised methods and achieves comparable performances to the supervised methods.

1 | INTRODUCTION
Retinal fundus images provide a window to inspect the fundus of the eye, and they are widely used for the diagnosis of various pathologies, such as age-related macular degeneration, diabetic retinopathy and glaucoma. Manual analysis of the retinal images is time consuming and tedious for ophthalmologists. Moreover, it is impossible to accurately quantify the fine structures in the fundus. Therefore, the automation of the analysis becomes highly important.

As one of the basic procedures in automatic analysis, vessel segmentation is still a challenging task. On one hand, the width of the vessels has large variability. For example, vessels at the end of each branch are always only several pixel wide or even only one pixel wide in the images. On the other hand, vessel segmentation task is adversely affected by other structures with high contrast to their surrounding pixels such as exudates and haemorrhagic spots.

Recently, several methods have been proposed to segment the vessels from the retinal fundus images, which can be classified into two classes, that is, unsupervised methods and supervised methods.

Unsupervised Methods: The unsupervised methods directly model the shape prior, appearance prior or the profile prior of the vessels, and use mathematical morphology operators or manually design matched filters to enhance the vessels, yielding a vesselness map ranging from [0,1]. In the vesselness map, pixels with high values are highly possible to be a vessel pixel. So they can segment the vessels from the vesselness map by thresholding. For example, Zana and Klein [1] propose to use mathematical morphology operators with linear structuring elements to model the vessels’ morphological properties, such
as linearity and connectivity. Mendonca and Campilho [2] use a top-hat transform with circular structuring elements to enhance the vessels first, and then reconstruction operators are used to segment the vessels. In [3], revised top-bottom-hat transformation is proposed for segmentation. Dash and Bhoi [4] generate a vessel-enhanced image and then use adaptive threshold to extract the vessels. Inspired by the shape and profile prior of the vessel, matched filters are proposed. For example, assuming that the vessels are bar-shaped structures, Azzopardi et al. use the trainable bar-selective combination of shifted filter responses (B-COSFIRE) to measure the vesselness [5]. Nguyen et al. assume that the vessels are line-like structures, and propose to use the multi-scale line filter responses to measure the vesselness [6]. Inspired by [6], improved line detectors are proposed for vessel segmentation in [7, 8]. Based on the observation that the profile of the vessels are Gaussian-shaped and symmetric with respect to the peak position, Zhang et al. propose to use the responses of the first-order of derivative of Gaussian to measure the vesselness [9]. Gou et al. also propose to use the responses of multi-scale Gaussian filtering to measure the vesselness [10] while Shukla et al. propose to use the responses of fractional filtering to measure the vesselness [11].

**Supervised Methods:** The supervised methods formulate the vessel segmentation as a classification problem, and they use the ground truth to learn aggregation strategies to aggregate various features for the classification. For example, Ricci and Perfetti [12] learn a support vector machine (SVM) to detect the vessel points from line filter responses. Lupascu et al. [13] learn an Adaboost classifier which takes multi-scale local intensity structure, spatial properties and geometry features as input, while Fraz et al. [14] learn a LogitBoost classifier. In [15], grey-level and moment invariants based features are extracted to learn a neural network scheme for vessel segmentation. In [16], information theory and machine learning techniques are used to automatically select a subset of B-COSFIRE filters to further improve the performances of [5]. Li et al. [17] train a deep neural network to learn a cross-modality data transform from retinal image to vessel map. Fu et al. [18] generate a vessel probability map by a fully convolutional neural networks, and then further refine the probability map by a fully connected conditional random fields. Thereafter, deep learning technologies are used to segment the vessels in [19–23]. Supervised methods are more invariable to vessel deformations and brightness variation since they combine different priors about the vessels. However, on one hand, in such methods, the ground truth is necessary. It needs professional skills to label the pixel-level ground truth. Also it is tedious and time consuming. Specifically, for deep learning–based vessel segmentation methods, Graphics Process Units (GPUs) are required to train the vessel segmentation model. On the other hand, supervised methods are dataset-dependent. Their performances usually drop sharply when the images are from different datasets. To learn a feature aggregation strategy without relying on the ground truth, we propose a ground truth free learning framework to segment the vessels. We observe that the vesselness maps provide some vessel pixels and background pixels, which can be used to learn a feature aggregation strategy and the vesselness maps can be generated by unsupervised methods. Specifically, pixels with high vesselness values are highly possible to be vessel pixels while pixels with small vesselness values are highly possible to be background. We call these pixels simple pixels. Pixels with mediocre vesselness values are uncertain whether to be vessel pixels or background pixels. We call these pixels difficult pixels. Thereby, it is natural for us to learn the feature aggregation strategy from the simple pixels to determine the labels of the difficult pixels.

There are two stages to learn vessel model from simple pixels. The first stage is the vesselness maps generation. In this stage, five vesselness maps are generated where the first two are proposed in this work: one is the centreline-boundary contrast map, modelling the appearances of the vessels; the other is the difference map of diffusion which uses the difference between the diffused image and the original one to measure the vesselness. The rest three are generated using existing unsupervised vessel segmentation methods, that is, the response map of B-COSFIRE filter [5], the line detector response map [6] and the green channel without the local mean. Those vesselness maps are used to generate the simple samples for the learning of the strong vessel model. The second stage is learning the strong vessel model from the simple samples. For each vesselness map of the retinal image, we use the simple pixels as the training set. For each sample, three feature descriptors are exploited, including the rest vesselness values, the histogram of local binary pattern (LBP) [24, 25] and the green colour channels to characterise the sample. Then we learn a strong vessel model based on the multiple kernel boosting (MKB) [26] to classify the difficult pixels. To exploit rich feature representations, we use four kernels including the linear function, polynomial function, radial basis function (RBF), and the sigmoid function. Finally, the strong vessel model is utilised to determine the labels of the difficult pixels. Totally, five strong vessel models are learnt for each retinal image and pixels those win three or more votes are labelled as vessel pixels finally.

Overall, the main contributions of this paper are twofold:

1. We propose two vesselness maps. One is the centreline-boundary contrast map and the other is the difference map of diffusion. They provide training samples for the strong vessel model.
2. We propose a vessel segmentation framework which learns a strong vessel model from the simple pixels for each retinal image. For each image, the strong model is customised and ground truth free.

Preliminary work has been published in MICCAI workshop [27]. We extend our previous work [27] by making the following improvements: (1) we use the LBP histogram to describe the texture information. (2) To cope with the various type of the features, we propose to use the MKB [26] to learn the feature aggregation strategy, while the conference version [27] simply aggregate the vesselness maps and
the green colour channel via a single SVM with RBF kernel function. (3) The experimental results on two widely used datasets DRIVE and STARE show the effectiveness of LBP and MKB for vessel segmentation. Specifically, comparing to unsupervised methods, our method achieves better performances. Comparing to supervised methods, without using any ground truth information, our method achieves comparable performances.

In the remainder of this paper, we first introduce the proposed method in Section 2. Then we report our experimental results in Section 3. We discuss our methods and the comparisons to state-of-the-arts in Section 4. Finally, we conclude our paper in Section 5.

2 | METHODOLOGY

The proposed unsupervised method obtains training samples from vesselness maps and learns strong vessel model from these samples; it includes two stages. The first stage is the generation of the vesselness maps and the second stage is the learning of the strong vessel model. The final results are obtained by voting.

2.1 | Vesseness maps generation

We extract vesselness maps from the green channel of the retinal images since the contrast between vessels and background in green channel is higher than that in the red and blue ones [1, 5, 12, 28]. To take the advantages of various vesselness maps, five vesselness maps are extracted. The first one is the multi-scale centreline-boundary contrast map \( f_{C} \), which takes the appearance and the linearity into consideration. The second one is the difference of the diffusion map \( f_{D} \), which considers the appearance difference between the vessels and the background from the view of diffusion system. The third one is the response of B-COSFIRE filters \( f_{B} \) [5], which assumes that the vessels are bar-shaped while the fourth one is the response of the multi-scale line filters \( f_{L} \) [6], which supposes the prior that the vessels are line-like. The last one is the green channel without the local mean \( f_{M} \), which models the appearance of the vessels. In the following, we describe the first two maps in detail since \( f_{B} \) and \( f_{L} \) have been described in [5] and [6], respectively, and the last one is easy to obtain.

### Multi-scale centreline-boundary contrast map

According to the properties that the intensities of the vessels are lower than the background pixels and the vessels seem to be linear locally, we present a centreline-boundary contrast filter with line structure to enhance the vessels. Figure 1(a) shows a basic centreline-surround contrast filter with size \( 5 \times 7 \). Generally, we define a basic centreline-boundary contrast filter size of \((2r_{1}+1) \times (2r_{2}+1)\) with direction \(0^\circ\) by

\[
G(x, y; r_{1}, r_{2}, 0^\circ) = \begin{cases} 
\frac{1}{4r_{2}+2}, & \text{if } x = 1 \text{ or } (2r_{1} + 1) \\
-\frac{1}{2r_{2}+1}, & \text{if } x = r_{1} + 1 \\
0, & \text{otherwise}
\end{cases}, \quad (1)
\]

where \(x\) is the row index ranging from 1 to \((2r_{1} + 1)\) and \(y\) is the column index ranging from 1 to \((2r_{2} + 1)\). Note that \((2r_{1} + 1)\) is the height of the filter and \((2r_{2} + 1)\) is the width of the filter.

Vessels stretch over the whole retinal image and converge into the optic disk from different directions. Moreover, the width of vessels is variable. Usually, the width of the trunk vessels is larger than their branch vessels. To handle the vessels in different directions, following [6] [12], we rotate \(G(\cdot; r_{1}, r_{2}, 0^\circ)\) every \(15^\circ\) from \(0^\circ\) to \(180^\circ\), and generate 12 filters. To enhance the vessels with variable width, the common strategy is multi-scale technology. For example, Nguyen et al. [6] adopt eight scales from 1 to 15 with a step 2 for the line detectors. Since we compute the contrast between the vessels and the background, we adopt a larger scale ranging from 3 to 19. So we generalise the centreline-boundary contrast filter by varying the half height \(r_{1}\) and half width \(r_{2}\) from 1 to 9 with step 1. Correspondingly, given an image \(I\), the centreline-boundary contrast vesselness
map is computed by
\[ f_C = \sum_{r_1, r_2} \max_{\theta} f \circ G(\cdot; r_1, r_2, \theta), \]

where \( \circ \) is a convolution operator. Thus, vessel-like pixels are enhanced due to their high contrast to background while the background pixels are suppressed for their smoothness. Figure 1(c) shows an example of a multi-scale centreline-boundary contrast map.

**Difference map of diffusion**

Regarding the image as a partial differential equation (PDE) based diffusion system, the intensity value can be regarded as concentration at a location [29]. Thus, when the PDE is applied on the retinal image, the intensities will defuse from high values to low values. As for retinal images, the intensities of the vessel pixels are lower than those of the background pixels. So the intensities of the vessel pixels increase after diffusion. Moreover, the background pixels are almost homogeneous. Consequently, the intensities of the background pixels change slightly. Therefore, we propose to use the difference map between the diffused image and the original image to measure the vesselness.

Mathematically, we formulate the image diffusion as an evolutionary PDE controlled by a function \( g(x, y, t) \):
\[ \frac{\partial I}{\partial t} = \text{DIV}(g(x, y, t) \nabla I), \]

where \( \text{DIV} \) is a divergence operator, \( g(x, y, t) \) is a non-linear diffusion function of and \( \nabla = \left[ \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right] \) is the spatial gradient. We adopt tri-diagonal matrix algorithm to solve the diffusion equation [30]. Then the difference map between the diffused image and the original image is obtained by
\[ f_D = I_t - I, \]

where \( I_t \) is the diffused image at time \( t \). Usually, a large \( t \) yields a difference map with clear background while the small vessels are eliminated. On the contrast, a small \( t \) yields a difference map with abundant details about vessels as well as noises. To make a balance between the less noise and more details about the vessels, we set \( t = 10 \). Figure 1(d) shows an example of the difference map of diffusion.

**2.2 Unsupervised vessel segmentation**

Until now, we obtain five unsupervised vesselness maps \( f_C, f_D, f_B, f_L, f_M \). We normalise each vesselness map into \([0,1]\), and the vesselness value indicates the probability of the pixel to be a vessel pixel. In each vesselness map, the pixels with extremely large vesselness values are highly possible to be vessel pixels. Similarly, pixels with extremely small vesselness values are highly possible to be background pixels. Thus these simple pixels can be directly labelled as vessel pixels or background pixels. However, the labels of pixels with mediocre vesselness values are controversial only according to one single vesselness map. To determine the labels of these difficult pixels, we learn a strong classifier on simple pixels sampled from one single vesselness map. Taking the \( f_C \) as example, Figure 2 illustrates the flow chart of our method using the simple pixels to determine the labels of the difficult pixels in \( f_C \).

In detail, for each vesselness map \( f \in \{f_C, f_D, f_B, f_L, f_M\} \), we first decompose it into two sets: the vessel candidate set \( S_C \) and the background set \( S_B \). \( S_C \) collects the top \( T_1 \) pixels and \( S_B \) collects the rest pixels. Pixels in \( S_B \) are directly regarded as background pixels while the labels of pixels in \( S_C \) are given by a strong classifier. To learn the strong classifier, we extract the training samples from \( f \). We take the top \( T_2 \) pixels in \( f \) as the positive samples, denoted as \( S_p \). Pixels near the \( S_C \) are regarded as the negative samples, denoted by \( S_n \). For each sample in the training set, we extract three types of features. The first one is the rest vesselness maps \( \{f_C, f_D, f_B, f_L, f_M\} \). Usually, the vessel pixels close to the vessel centreline have smaller values in green channel than those vessel pixels close to the vessel wall. Vessel pixels close to the vessel wall have smaller values than the background pixels. To describe such a structure, we compute the LBP histogram [24, 25] of the patch centred on the sample from the green channel as the second feature. Due to the high contrast between the vessels and background in green channel, similar to [28], we also use the green channel as the third feature. To directly learn an SVM for the retinal image, it is required to determine the appropriate kernel beforehand. However, it is still not clear how these vessel values, texture information as well as colour information can be well combined. To handle this problem, we propose to use multiple kernel boosting [26, 31] to combine several SVMs with different kernels.

Formally, for a sample \( s \), the linear combination of \( L \) kernels \( \{k_l\}_{l=1}^L \) is defined as
\[ k(s, s_i) = \sum_{l=1}^L \beta_l k_l(s, s_i), \text{ s.t. } \sum_{l=1}^L \beta_l = 1, \beta_l \in \mathbb{R}^+, \]

where \( s, s_i \in S \), and \( \beta_l \) is the \( l \)th kernel weight. Then the decision function is defined as
\[ Y(s) = \sum_{i=1}^N \alpha_i y_i \sum_{l=1}^L \beta_l k_l(s, s_i) + b \]

where \( N \) is the number of the samples in \( S \), \( \{\alpha_i\} \) is the Lagrange multiplier and \( b \) is the bias in the standard SVM, and \( y_i = 1 \) if \( s_i \) is from \( S_p \) otherwise \( y_i = -1 \). Let \( b_i \) denote the response from a single SVM:
\[ b_i(s) = \sum_{i=1}^N \alpha_i y_i k_i(s, s_i) + b_i \]
where \( \sum_{j=1}^{f_c} b_j = b \), then we can rewrite the decision function Equation (6) as

\[
Y(z) = \sum_{j=1}^{f_c} \beta_j b_j(z).
\] (8)

Treating each single SVM \( b_j(z) \) as a weak classifier and the strong vessel model as the linear combination of all the SVMs, it is natural for us to formulate the problem with Adaboost algorithm in the form:

\[
Y(z) = \sum_{j=1}^{f} \beta_j b_j(z),
\] (9)

where \( f \) is the number of the boosting iterations. Following [26], we first separately train the single SVM classifier \( b_j(z) \) to construct an SVM pool. In our case, we use four kernel functions, that is, the linear function, the polynomial function, the RBF and the sigmoid function and extract three kinds of features for each sample. Thereby, the SVM pool consists of 12 SVMs. Then we iteratively select one SVM with smallest weighted classification error \( \epsilon_j \) into the decision function, where \( \epsilon_j \) is computed by

\[
\epsilon_j = \frac{\sum_{i=1}^N w_j(i) [b_j(s_i)] [1 + \text{sgn}(\gamma_j b_j(s_i))] / 2 \sum_{i=1}^N w_j(i) [b_j(s_i)]}{\sum_{i=1}^N w_j(i) [b_j(s_i)]}. \] (10)

In Equation (10), \( \text{sgn}(\cdot) \) is a function that equals 1 when the input is larger than 0. Otherwise, it outputs 0. \( w_j(i) \) is the weight of the sample \( s_i \) at iteration \( j \). When \( j = 1 \), \( w_j(i) = 1/N \). When \( j > 1 \), \( w_j(i) = \frac{w_{j-1}(i) e^{-\beta_j y_j(s_i)}}{Z_{j-1}} \) and \( Z_{j-1} \) is the normalisation factor. When the weight of the chosen classifier computed by \( \tilde{\beta}_j = 0.5 \log \frac{\epsilon_j}{\epsilon_j} \) is less than 0, we stop the iteration.

For each vesselness map, we learn a strong vessel model. Totally we learn five strong vessel models. We apply these strong models to the vessel candidate set, and label the pixels winning the majority voting as vessel pixels, obtaining the segmentation.

As there are small isolated holes and noises in the segmentation, we fill the holes whose area sizes are less than 20 pixels. Meanwhile, we compute the equivalent diameter \( R_E \) of the circle with the same area to each connected component.
in the segmentation and the major axis $R_M$ of the ellipse that has the same normalised second central moments as the connected component. The larger the ratio $R_E/R_M$, the closer the connected components looks like a circle. Thus, to remove the noises that look like circles, we remove the small connected components satisfying $R_E/R_M \leq T_3$, where $T_3$ is a threshold.

3 | EXPERIMENTAL RESULTS

We use two widely used public datasets to evaluate the proposed method. One is DRIVE [32] which includes 20 retinal images for training and 20 images for testing. These images are captured by a Canon CR5 non-mydriatic 3CCD camera at 45° field of view. For each image in the testing set, two manually segmented maps by two observers and a mask image are provided. The first observer marked 577,649 pixels as vessel and 3,960,494 pixels as background (12.7% vessel) while the second observer labelled 14.9% pixels as vessel. Following [6, 32], we use the segmentations of the first observer as ground truth for evaluation and the performance of the second observer as a benchmark for comparison.

In our experiments, since the vessel pixels account for 12.7% in DRIVE and 10.4% in STARE, to ensure that almost all the vessel pixels are included by the candidate set, we set $T_1$ to 14% for DRIVE [32] and 13% for STARE [33], respectively. To ensure that almost all the pixels in the simple positive set are vessel pixels, we choose the top 7% pixels as simple positive pixels. So we set $T_2 = 7\%$ for both two datasets. We vary $T_3$ from 0.1 to 0.8 with a step 0.1 on the training set of DRIVE [32]. We find that the performances are best when $T_3 = 0.4$. So we set $T_3 = 0.4$ on both datasets.

Following the previous methods [5, 14, 17], we adopt accuracy ($\text{Acc}$), sensitivity ($\text{Se}$) and specificity ($\text{Sp}$) to compare the performance of our method with the state-of-the-art methods. The accuracy measures the proportion of pixels that are correctly detected. The sensitivity measures the proportion of vessel pixels that are correctly detected. The specificity is the proportion of background pixels that are correctly detected. Conventionally, only the pixels in the field of the view are taken into consideration since pixels out the field of the view can be easily labelled as background via thresholding. The performance results for each image in the two datasets are reported in Table 1. The last row of Table 1 is the average value of sensitivity, specificity and accuracy.

| TABLE 1 Performance results on DRIVE [32] and STARE [33]. Se, Sp and Acc represent sensitivity, specificity and accuracy, respectively. |
|-----------------|-------|-------|-------|
| DRIVE           |     |     |     |
| Image name      | Se   | Sp   | Acc  |
| 01_test         | 0.7832 | 0.9754 | 0.9803 |
| 02_test         | 0.7402 | 0.9862 | 0.9496 |
| 03_test         | 0.6745 | 0.9816 | 0.9369 |
| 04_test         | 0.7244 | 0.9851 | 0.9504 |
| 05_test         | 0.6979 | 0.985 | 0.9461 |
| 06_test         | 0.6637 | 0.9848 | 0.9395 |
| 07_test         | 0.682 | 0.9808 | 0.9413 |
| 08_test         | 0.6654 | 0.9746 | 0.9359 |
| 09_test         | 0.7276 | 0.9801 | 0.9504 |
| 10_test         | 0.7253 | 0.9804 | 0.9499 |
| 11_test         | 0.7062 | 0.9795 | 0.9441 |
| 12_test         | 0.7397 | 0.9787 | 0.9488 |
| 13_test         | 0.6897 | 0.9827 | 0.9412 |
| 14_test         | 0.7562 | 0.972 | 0.9467 |
| 15_test         | 0.7773 | 0.9641 | 0.9447 |
| 16_test         | 0.7424 | 0.9821 | 0.9508 |
| 17_test         | 0.7066 | 0.9803 | 0.9466 |
| 18_test         | 0.792 | 0.9754 | 0.9543 |
| 19_test         | 0.8381 | 0.9801 | 0.963 |
| 20_test         | 0.8123 | 0.9718 | 0.9548 |
| average         | 0.7259 | 0.9811 | 0.9483 |

| STARE           |     |     |     |
| Image name      | Se   | Sp   | Acc  |
| im0001          | 0.6726 | 0.9704 | 0.9381 |
| im0002          | 0.6371 | 0.9736 | 0.9429 |
| im0003          | 0.7707 | 0.9719 | 0.9556 |
| im0004          | 0.7205 | 0.9737 | 0.9478 |
| im0005          | 0.6677 | 0.9756 | 0.9379 |
| im00044         | 0.8071 | 0.981 | 0.9659 |
| im00077         | 0.8518 | 0.974 | 0.9607 |
| im00081         | 0.8641 | 0.9753 | 0.9639 |
| im00082         | 0.8589 | 0.9718 | 0.9597 |
| im0139          | 0.7443 | 0.9823 | 0.9561 |
| im0162          | 0.8556 | 0.9712 | 0.96 |
| im0163          | 0.8915 | 0.975 | 0.9662 |
| im0235          | 0.7936 | 0.9779 | 0.9554 |
| im0236          | 0.7926 | 0.9806 | 0.9572 |
| im0239          | 0.7835 | 0.9776 | 0.9549 |
| im0240          | 0.671 | 0.9809 | 0.9389 |
| im0255          | 0.7937 | 0.9834 | 0.9602 |
| im0291          | 0.835 | 0.9797 | 0.9698 |
| im0319          | 0.8487 | 0.9694 | 0.9623 |
| im0324          | 0.757 | 0.9676 | 0.9551 |
| average         | 0.7809 | 0.9756 | 0.9551 |

4 | DISCUSSION

In this section, we first discuss the effectiveness of each component in the proposed method. Then to show that the proposed method reaches the state-of-the-art performance, we compare the proposed method with the state-of-the-arts.

4.1 | Component analysis

To demonstrate the effectiveness of the proposed unsupervised learning framework, we first compare our method with each vesseldness maps. As is shown in Table 2, when we use an RBF SVM to learn the classifier [27], the specificity on DRIVE [32] is 97.07% and the specificity of the best component, that is, [5], is 97.04%. The accuracy of our method is superior to each vesselness map. When we combine the vesselness maps together with histogram of LBP [24] [25] and green colour features, the specificity is further improved. Although the sensitivity decreases, the final accuracy is still improved on DIRVE. Similarly, on STARE [33], our method with RBF SVM achieves better accuracy than single vesselness map. When we adopt MKB to combine the features, the accuracy is further improved. It is worth noting that Table 2 also demonstrates the effectiveness of the two proposed vesselness maps $f_C$ and $f_D$. 
TABLE 2 Single vesselness map versus Aggregation on DRIVE [32] and STARE [33]. Se, Sp and Acc represent sensitivity, specificity and accuracy respectively.

| Single vesselness map versus Aggregation | DRIVE | STARE |
|-----------------------------------------|-------|-------|
|                                          | Se    | Sp    | Acc  | Se    | Sp    | Acc  |
| Centreline-boundary contrast f_C        | 0.7659| 0.9604| 0.9354| 0.7037| 0.9766| 0.9493|
| Difference of diffusion f_D             | 0.7388| 0.9561| 0.9283| 0.6217| 0.9843| 0.9470|
| B-COSFIRE f_B [5]                       | 0.7655| 0.9704| 0.9442| 0.7716| 0.9563| 0.9497|
| Line detector f_L [6]                   | 0.7426| 0.9700| 0.9407| 0.8014| 0.9481| 0.9325|
| Green channel without local mean f_M    | 0.6498| 0.9721| 0.9312| 0.6930| 0.9557| 0.9298|
| Ours with RBF SVM [27]                  | 0.7718| 0.9707| 0.9451| 0.7822| 0.9745| 0.9541|
| Ours with MKB                           | 0.7259| 0.9811| 0.9483| 0.7809| 0.9756| 0.9551|

**FIGURE 3** Visual comparisons between our method with MKB and with single RBF SVM. Pixels in yellow are the true positive pixels. Pixels in red are the false negative pixels, and pixels in green are the false positives pixels. Error detection examples are marked in magenta ellipses. Missing detection examples are marked in white ellipses.

Figure 3 shows two examples to illustrate the effectiveness of using MKB-based SVM. When combining the LBP histogram [24, 25] and green channel using the SVM based on MKB, less background pixels with high contrast to their surrounding pixels are segmented out. We also validate the effectiveness of the LBP [24, 25] and green channel in vessel segmentation and the post-processing. Table 3 shows that the accuracies on both two datasets after combining with LBP [24, 25] are slightly improved compared to those only aggregating the vesselness maps and green channel. After filling the holes and removing the small circle-like structures, the accuracies are further improved.

### 4.2 Comparisons with the state-of-the-arts

We compare our method with seven supervised methods and nine unsupervised methods. Among the supervised methods, Fu et al. [18] and DUNet [21] are deep learning based. The results of these compared methods are from their original paper and they are compared under the same experimental protocol. The results of our method and the compared methods on DRIVE [32] and STARE [33] are reported in Tables 4 and 5, respectively. Generally, our method outperforms the first five unsupervised methods in terms of sensitivity, specificity and accuracy on both two datasets. Compared to the last four unsupervised methods, our method achieves comparable performances on both two datasets. This indicates that aggregating multiple vesselness maps by learning from simple pixels is able to further advance the segmentation performances. Comparing to the supervised methods, the proposed method still achieves better sensitivity. Although the performances of [17] and [34] are superior to our method on DRIVE [32] when the models are trained on the training images from DRIVE [32], their accuracies decrease when their models are
TABLE 3 Validation of the features and post-processing on DRIVE [32] and STARE [33]. Se, Sp and Acc represent sensitivity, specificity and accuracy, respectively.

| Feature type | DRIVE | | | STARE | | |
|--------------|-------|---|---|-------|---|---|
|              | Se     | Sp | Acc | Se     | Sp | Acc |
| Vesselness maps + G | 0.7371 | 0.9777 | 0.9467 | 0.7890 | 0.9697 | 0.9505 |
| Vesselness maps + G + LBP | 0.7322 | 0.9790 | 0.9473 | 0.7879 | 0.9703 | 0.9510 |
| Vesselness maps + G + LBP with post precessing | 0.7259 | 0.9811 | 0.9483 | 0.7809 | 0.9756 | 0.9551 |

TABLE 4 Comparisons with the state-of-the-art methods on the DRIVE [32]. Se, Sp and Acc represent sensitivity, specificity and accuracy respectively.

| Type | Methods | Year | Se | Sp | Acc |
|------|---------|------|----|----|-----|
| Second observer | | | 0.7757 | 0.9725 | 0.9473 |
| Supervised methods | Staal et al. [32] | 2004 | NA | NA | 0.9441 |
| | Soares et al. [28] | 2006 | 0.7332 | 0.9782 | 0.9461 |
| | Fu et al. [18] | 2016 | 0.7294 | NA | 0.9470 |
| | Fraz et al. [14] trained on DRIVE | 2012 | 0.7406 | 0.9807 | 0.9480 |
| | Fraz et al. [14] trained on STARE | 2012 | 0.7242 | 0.9792 | 0.9456 |
| | Li et al. [17] trained on DRIVE | 2016 | 0.7569 | 0.9816 | 0.9527 |
| | Li et al. [17] trained on STARE | 2016 | 0.7273 | 0.9810 | 0.9486 |
| | Wang et al. [34] trained on DRIVE | 2015 | 0.8173 | 0.9733 | 0.9533 |
| | Wang et al. [34] trained on STARE | 2015 | 0.8431 | 0.9574 | 0.9428 |
| | DUNet [21] trained on DRIVE | 2019 | 0.7963 | 0.9800 | 0.9566 |
| | DUNet [21] trained on STARE | 2019 | 0.6505 | 0.9914 | 0.9481 |
| Unsupervised methods | Al-Diri et al. [35] | 2009 | 0.7282 | 0.9551 | NA |
| | Fraz et al. [36] | 2012 | 0.7125 | 0.9759 | 0.9430 |
| | Nguyen et al. [6] | 2013 | 0.7426 | 0.9700 | 0.9407 |
| | Singh et al. [37] | 2016 | 0.7594 | 0.9708 | 0.9522 |
| | Azzopardi et al. [5] | 2015 | 0.7655 | 0.9704 | 0.9442 |
| | Zhou et al. [8] | 2020 | 0.7262 | 0.9803 | 0.9475 |
| | Shukla et al. [11] | 2020 | 0.7015 | 0.9836 | 0.9476 |
| | Wang et al. [3] | 2019 | 0.7287 | 0.9775 | 0.9446 |
| | Yin et al. [38] | 2015 | 0.7246 | 0.9790 | 0.9403 |
| | Ours | 2020 | 0.7259 | 0.9811 | 0.9483 |

Their accuracies decrease to 0.9486 and 0.9428, respectively, while ours is 0.9483. Similarly, [17] and [34] achieve better accuracy on STARE [33] when training their models on STARE [33]. But their accuracies decrease to 0.9545 and 0.9413 while ours is 0.9551. The performance drop comes from an assumption. Commonly, supervised methods assumes that testing images and training images follow the same distribution. Once the testing image does not follow the distribution of the training images, supervised methods tend to make wrong decisions. It is worth to mention that our method achieves similar performances on both DRIVE and STARE to deep learning-based method [18]. The possible reason is that the spatial scale of vessels is extremely small and preserving the spatial resolution when feature extraction contributes to the segmentation performance. However, the standard convolution neural network used in Fu et al. [18] for feature extraction is unable to preserve the resolution of feature maps at high-level due to the maxpooling layer with downsampling. For the deep learning-based methods DUNet [21], it is obvious that it achieves higher sensitivity and accuracy than ours, and similar specificity to ours on DRIVE when it is trained on DRIVE. However, when it is trained on STARE, the performances of DUNet [21] drop sharply and are inferior to ours in terms of sensitivity and accuracy. This indicates that the performances of DUNet [21] is dataset-dependent. Similarly, DUNet [21] achieves superior to ours in terms of specificity and accuracy on STARE when it is trained on STARE. But its performances drop sharply when it is trained on DRIVE and are inferior to ours on sensitivity, specificity and accuracy consistently. Additionally, the training of DUNet
TABLE 5 Comparisons with the state-of-the-art methods on the STARE [33]. Se, Sp and Acc represent sensitivity, specificity and accuracy, respectively.

| Type            | Methods                          | Year | Se  | Sp  | Acc  |
|-----------------|----------------------------------|------|-----|-----|------|
| Supervised      | Second observer                  |      | 0.8951 | 0.9386 | 0.9350 |
|                 | Staal et al. [32]                | 2004 | NA  | NA  | 0.9516 |
|                 | Soares et al. [28]               | 2006 | 0.7207 | 0.9747 | 0.9479 |
|                 | Fu et al. [18]                   | 2016 | 0.7140 | NA  | 0.9545 |
|                 | Fraz et al. [14] trained on STARE| 2012 | 0.7548 | 0.9763 | 0.9534 |
|                 | Fraz et al. [14] trained on DRIVE| 2012 | 0.7010 | 0.9770 | 0.9495 |
|                 | Li et al. [17] trained on STARE  | 2016 | 0.7726 | 0.9844 | 0.9628 |
|                 | Li et al. [17] trained on DRIVE  | 2016 | 0.7027 | 0.9828 | 0.9545 |
|                 | Wang et al. [34] trained on STARE| 2015 | 0.8104 | 0.9791 | 0.9621 |
|                 | Wang et al. [34] trained on DRIVE| 2015 | 0.7116 | 0.9675 | 0.9413 |
|                 | DUNet [21] trained on STARE      | 2019 | 0.7595 | 0.9878 | 0.9641 |
|                 | DUNet [21] trained on DRIVE      | 2019 | 0.7000 | 0.9759 | 0.9474 |
| Unsupervised    | Al-Diri et al. [35]              | 2009 | 0.7521 | 0.9681 | NA   |
|                 | Fraz et al. [36]                 | 2012 | 0.7311 | 0.968  | 0.9442 |
|                 | Nguyen et al. [6]                | 2013 | 0.8014 | 0.9481 | 0.9325 |
|                 | Singh et al. [37]                | 2016 | 0.7939 | 0.9376 | 0.9270 |
|                 | Azzopardi et al. [5]             | 2015 | 0.7716 | 0.9563 | 0.9497 |
|                 | Zhou et al. [8]                  | 2020 | 0.7865 | 0.9730 | 0.9535 |
|                 | Shukla et al. [11]               | 2020 | 0.7023 | 0.9863 | 0.9573 |
|                 | Wang et al. [3]                  | 2019 | 0.7526 | 0.9733 | 0.9503 |
|                 | Yin et al. [38]                  | 2015 | 0.7023 | 0.9863 | 0.9573 |
|                 | Ours                             | 2020 | 0.7809 | 0.9756 | 0.9551 |

**FIGURE 4** Visual comparisons on DRIVE [32] between B-COSFIRE [5], multi-scale line detector [6] and our method. Pixels in yellow are the true positive pixels. Pixels in red are the false negative pixels, and pixels in green are the false positives pixels. Error detection examples are marked in magenta ellipses. Missing detection examples are marked in white ellipses.

We make visual comparisons between [5], [6] and our method. Figure 4 shows two examples from DRIVE, in which pixels in yellow are the true positive pixels. Pixels in red are the false negative pixels, and pixels in green are the false positives.

model [21] needs pixel-wise annotation and costly GPUs while ours does not need. For the rest methods in Tables 4 and 5, our method achieves superior accuracies on both two datasets.
FIGURE 5 Visual comparisons on STARE [33] between B-COSFIRE [5], multi-scale line detector [6] and our method. Pixels in yellow are the true positive pixels. Pixels in red are the false negative pixels and pixels in green are the false positives pixels. Error detection examples are marked in magenta ellipses. Missing detection examples are marked in white ellipses.

FIGURE 6 Examples for fail segmentation. (a) From DRIVE and (b) the segmentation result by ours. (c) From STARE and (d) the segmentation result by ours. Pixels in yellow are the true positive pixels. Pixels in red are the false negative pixels, and pixels in green are the false positives pixels. We can see that our method fails to segment the tiny vessels in the first example and wrongly classifies the red lesions to vessel pixels in the second example.

pixels. From the first example, we can see that B-COSFIRE [5] and multi-scale line detector [6] fail to segment the vessel pixels in white ellipses due to the low contrast to background. From the second example, we can see that B-COSFIRE [5] and multi-scale line detector [6] are confused by the noise background and optic disc. On the contrary, our method works well. The possible reason is that our method is able to aggregate multiple vesselness maps by learning to make decisions on difficult pixels more accurately. Figure 5 shows two examples from STARE. Similarly, our method works better when the background is messy with lesions than B-COSFIRE [5] and multi-scale line detector [6].

Although our method is able to aggregate the vesselness maps to achieve advanced performances, it still fails to segment tiny vessels as is shown in Figure 6(b). The possible reason is that the features we exploit are not discriminative enough to distinguish the tiny vessels from background. Additionally, since red lesions have similar appearances to vessels, our method tends to treat them as vessel pixels wrongly, as is shown in Figure 6(d).

5 | CONCLUSION

We present two vesselness maps and a new framework for vessel segmentation in retinal images. Our framework learns strong vessel models from the simple pixels whose labels can be determined easily to determine the labels of the difficult pixels that are controversial in a single feature map. To combine various types of features, we propose to use MKB to choose the kernels automatically. For each retinal image, the strong vessel models are customised which only use the information from the image. Our method shows its advantages in two aspects: (1) comparing to most of other unsupervised methods, it avoids the sensitive threshold selection. Moreover, our method achieves better performances on the two public datasets DRIVE [32] and STARE.
[33] than the existing unsupervised methods. (2) Comparing to supervised methods, our method does not rely on the expensive ground truth while it achieves competitive performances to them.

REFERENCES

1. Zana, F., Klein, J.: Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. IEEE Trans. Image Process. 10(7), 1010–1019 (2001)
2. Mendonca, A.M., Campilho, A.: Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans. Med. Imaging 25(9), 1200–1213 (2006)
3. Wang, W., et al.: Segmenting retinal vessels with revised top-bottom-hat transformation and flattening of minimum circumscribed ellipse. Med. Biol. Eng. Comput. 57(7), 1481–1496 (2019)
4. Dash, J., Bhoi, N.: An unsupervised approach for extraction of blood vessels from fundus images. J. Digit. Imaging 31(6), 857–868 (2018)
5. Azzopardi, G., et al.: Trainable COSFIRE filters for vessel delineation with application to retinal images. Med. Image Anal. 19(1), 46–57 (2015)
6. Nguyen, U.T.V., et al.: An effective retinal blood vessel segmentation method using multi-scale line detection. Pattern Recognit. 46(3), 703–715 (2013)
7. Yue, K., et al.: Improved multi-scale line detection method for retinal blood vessel segmentation. IET Image Process. 12(8), 1450–1457 (2018)
8. Zhou, C., et al.: A new robust method for blood vessel segmentation in retinal fundus images based on weighted line detector and hidden Markov model. Comput. Methods Programs Biomed. 187, 105231 (2020)
9. Zhang, B., et al.: Retinal vessel extraction by matched filter with first-order derivative of Gaussian. Comput. Biol. Med. 40(4), 438–445 (2010)
10. Gou, D., et al.: Retinal vessel extraction using dynamic multi-scale matched filtering and dynamic threshold processing based on histogram fitting. Mach. Vis. Appl. 29(4), 655–666 (2018)
11. Shukla, A.K., et al.: A fractional filter based efficient algorithm for retinal blood vessel segmentation. Biomed. Signal Process. Control 59, 101883 (2020)
12. Recci, E., Perfetti, R., Retinal blood vessel segmentation using line operators and support vector classification. IEEE Trans. Med. Imaging 26(10), 1357–1365 (2007)
13. Lupascu, C.A., et al.: FABC: Retinal vessel segmentation using AdaBoost. IEEE Trans. Inf. Technol. Biomed. 14(5), 1267–1274 (2010)
14. Fraz, M.M., et al.: An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans. Biomed. Eng. 59(9), 2538–2548 (2012)
15. Marin, D., et al.: 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(1), 146–158 (2011)
16. Strisciuglio, N., et al.: Supervised vessel delineation in retinal fundus images with the automatic selection of B-COSFIRE filters. Mach. Vis. Appl. 27(8), 1137–1149 (2016)
17. Li, Q., et al.: A cross-modality learning approach for vessel segmentation in retinal images. IEEE Trans. Med. Imaging 35(1), 109–118 (2016)
18. Huazhu, F., et al.: Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. 2016 IEEE 13th International Symposium on Biomedical Imaging, April 2016, Prague, Czech Republic, pp. 698–701. IEEE, Piscataway, NJ (2016)
19. Yan, Z., et al.: Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. IEEE Trans. Biomed. Eng. 65, 1912–192 (2018)
20. He, Q., et al.: Multi-label classification scheme based on local regression for retinal vessel segmentation. 25th IEEE International Conference on Image Processing. IEEE, Piscataway, NJ (2018)
21. Jin, Q., et al.: DUNet: A deformable network for retinal vessel segmentation. Knowledge-Based Systems. 178, 149–162 (2019)
22. Zhang, S., et al.: Attention guided network for retinal image segmentation. In: Shen, D. (ed.) International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 797–805. Springer, Cham (2019)
23. Zhang, S., et al.: Retinal image segmentation with a structure-texture demixing network. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 765–774. Springer, Cham (2020)
24. Ojala, T., et al.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24(7), 971–987 (2002)
25. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29(6), 915–928 (2007)
26. Yang, F., et al.: Human tracking by multiple kernel boosting with locality affinity constraints. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) Asian Conference on Computer Vision, pp. 39–50. Springer, Berlin (2010)
27. Liu, Q., et al.: Retinal vessel segmentation from simple to difficult. In: Chen, X., Garvin, M.K., Liu, J., Trucco, E., Xu, Y. (eds.) Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop, OMA 2016, Athens, Greece, pp. 57–64. Springer, Berlin (2016)
28. Soares, J.V.B., et al.: Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification. IEEE Trans. Med. Imaging 25(9), 1214–1222 (2006)
29. Salinas, H.M., Fernandez, D.C., Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography. IEEE Trans. Med. Imaging 26(6), 761–771 (2007)
30. Ralli, J.: PDE based image diffusion and AOS, April 27, (2014), [Online]. http://www.jarnoralli.fi/joomla/images/pdf/non_linear_image_diffusion_and_aos_ralli_2014.pdf
31. Yang, F., et al.: Robust visual tracking via multiple kernel boosting with affinity constraints. IEEE Trans. Circuits Syst. Video Technol. 24(2), 242–254 (2014)
32. Staal, J., et al.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging 23(4), 501–509 (2004)
33. Hoover, A.W., et al.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging 19(3), 203–210 (2000)
34. Wang, S., et al.: Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. Neurocomputing 149, 708–717 (2015)
35. Aldiri, B., et al.: An active contour model for segmenting and measuring retinal vessels. IEEE Trans. Med. Imaging 28(9), 1488–1497 (2009)
36. Mozamm Fraz, M., et al.: An approach to localize the retinal blood vessels using bit planes and centerline detection. Comput. Methods Programs Biomed. 108(2), 600–616 (2012)
37. Singh, N.P., Srivastava, R.: Retinal blood vessels segmentation by using Gumbel probability distribution function based matched filter. Comput. Methods Programs Biomed. 129, 40–50 (2016)
38. Yin, B., et al.: Vessel extraction from non-fluorescein fundus images using orientation-aware detector. Med. Image Anal. 26(1), 232–242 (2015)

How to cite this article: Zou B, Fu H, Chen Z, Liu Q. Ground truth free retinal vessel segmentation by learning from simple pixels. IET Image Process. 2021;15:1210–1220. https://doi.org/10.1049/ipr.12098