Pattern Recognition 98 (2020) 107068

Contents lists available at ScienceDirect

Pattern Recognition

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

Improving retinal vessel segmentation with joint local loss by matting

He Zhao[a], Huiqi Li[a,*], Li Cheng[b,c,*]

[a] Beijing Institute of Technology, China
[b] Department of Electrical and Computer Engineering, University of Alberta, Canada
[c] Bioinformatics Institute, A*STAR, Singapore

ARTICLE INFO

Article history:
Received 29 January 2019
Revised 6 August 2019
Accepted 25 September 2019
Available online 27 September 2019

Keywords:
Vessel segmentation
Retinal images
Deep learning
Local matting loss

ABSTRACT

Besides the binary segmentation, many retinal image segmentation methods also produce a score map, where a nonnegative score is assigned for each pixel to indicate the likelihood of being a vessel. This observation inspires us to propose a new approach as a post-processing step to improve existing methods by formulating segmentation as a matting problem. A trimap is obtained via a bi-level thresholding of the score map from existing methods, which is instrumental in focusing the attention to pixels of these unknown areas. A dedicated end-to-end matting algorithm is further developed to retrieve those vessel pixels in the unknown areas, and to produce the final vessel segmentation by minimizing global pixel loss and local matting loss. Our approach is shown to be particularly effective in rescuing thin and tiny vessels that may lead to disconnections of vessel fragments. Moreover, it is observed that our approach is capable of improving the overall segmentation performance across a broad range of existing methods.

1. Introduction

Automated analysis of retinal fundus images is gaining popularity in a broader range of clinical specialties including diabetic retinopathy, cardiovascular disease, and hypertension [1–3]. In many clinical applications, it is crucial to extract detailed retinal vessel morphology for follow-up vasculature analysis. On the other hand, numerous retinal image segmentation methods have been developed over the years, including both supervised and unsupervised methods, which have resulted in noticeable progresses. Unfortunately, it is still a challenging problem to extract small vessels which are usually thin and blurry, thus difficult to be separated from the textural background.

In this paper, we propose a matting-based approach to extract these small retinal vessels with global pixel loss and local matting loss. We observe that many existing methods produce a score map as a by-product, where for each pixel, a nonnegative score is presented to quantify its affinity to the vascular foreground. This inspires us to propose an end-to-end approach to boost the performance of an existing baseline segmentation method. Given an input retinal image, a score map is obtained by applying an existing segmentation method of interest. A bi-level threshold of the score map gives rise to a trimap representation focusing on these unknown area pixels that are yet to be considered as part of the vessels, while the rest pixels that are clearly either background or foreground. A dedicated end-to-end matting algorithm is further developed to classify the vessel pixels in the unknown areas and to complete the final vessel segmentation.

The main contributions of our approach are three-fold. First, a deep learning framework is proposed to improve the vessel segmentation performance of existing methods, where the segmentation problem is transformed into a matting task. Second, a new loss function is proposed in vessel segmentation, in which the global pixel loss and local matting loss are combined to handle the ambiguous pixels that often reside around the boundary of small vessels. Empirical evidence suggests our approach is particularly effective in segmenting small vessels. Third, it is capable of improving the results of *a wide range* of existing methods, being either supervised or unsupervised. When applying to the state-of-the-art methods such as DRIU [4] and Kernel boost [5], our approach still yields better results. Moreover, working with the unsupervised methods such as MSLD [6] that usually performs inferior to the supervised counterparts, our approach helps to significantly boost the performance to a level that is comparable with the best supervised methods. In addition to the typical fundus images, our approach also works well with images acquired by other retinal imaging instruments such as scanning laser ophthalmoscope.

2. Related work

Retinal image segmentation methods can be roughly categorized into two types: unsupervised and supervised methods.

* Corresponding authors.
E-mail addresses: huiqili@bit.edu.cn (H. Li), lcheng5@ualberta.ca (L. Cheng).

https://doi.org/10.1016/j.patcog.2019.107068
0031-3203/

Comparing with unsupervised methods, in general supervised methods deliver better segmentation results. However, it relies on a set of ground-truth training examples for constructing a dedicated model. Meanwhile, unsupervised methods tend to execute at a faster speed; they are feasible to be deployed on new datasets when annotations are not available.

More specifically, unsupervised methods are designed by encoding the domain knowledge to best capture retinal vessel characteristics. In [7,8], mathematical morphological operators are engaged in identifying the vessels. These vessels are then spatially enhanced by linear filters and curvature-based analysis. To facilitate the proper delineation of vessel boundaries, one may consider to adopt Hessian-based techniques [9,10] to incorporate the second order derivatives or eigenvalues. In the meantime, image filtering based methods have also been widely deployed in retinal vessel segmentation [11–15]. In [11], both vessel centerlines and vessel segments are extracted from an input image. The centerline is extracted by the first order derivative of a Gaussian filter, while a multidirectional morphological top-hat operator is utilized to segment the vessels. Chaudhuri et al. [12] develop a bank of 2-D matched filters with twelve directions for detecting vessels based on the Gaussian-smoothed shape of vessel cross-sectional profile. Wang et al. [13] further develop a matched filter that combines multiwavelet kernels to distinguish vessels from lesions or noise background. The final vessel segmentation is attained after a follow-up adaptive thresholding. In [14], vessel areas obtained from matched filter response are further identified by applying different threshold probes. A matched filter is applied to enhance the vessels with threshold operator to obtain the binary vascular trees in [16] and the keypoints are also detected after the reconstruction of vascular trees. Recently, the B-COSFIRE (Bar Combination Of Shifted Filter Responses) filter is proposed by [15] to detect vessels by evaluating the empirical mean of a bank of Different-of-Gaussian filters. Authors in [6] consider the line detector responses at different scales collectively to deliver the final segmentation. Vessel tracking or tracing is another way to segment retinal vessels. The approach of [17] is based on a Bayesian method with maximum a posteriori (MAP) formulation to locate vascular structure by connecting sampled edge points in a tracking fashion. A follow-up work by [18] combines MAP criterion with multiscale line detection to exploit two-dimensional vessel information. In [19], retinal vessels are further separated into arteries and veins by keypoint detector and graph search algorithm. Meanwhile, the authors in [20] consider to extract vessels by means of an orientation-aware detector to capture the locally oriented and linearly elongated structural property of vessels. An active contour model is also proposed by [21], to take advantage of the local phase enhancement map to provide a reliable vessel map, as well as the region-level information of pixel intensities to exclude possible outliers. In [22], the authors transfer the 2D image to a Lie-group space of positions and orientations. The vessels are then extracted by applying multi-scale second-order Gaussian filters.

Supervised methods, on the other hand, are data-driven in which a set of well-annotated training examples are required by default. This often results in better segmentation performance in practice. As one of the early works, Staal et al. [23] consider a two-step method by first obtaining the vessel features from a ridge detector, then utilizing a K-nearest neighbor classifier to predict vessel pixels. In [24], authors examine features from pixel intensity and Gabor wavelet responses over scales, which are subsequently fed into a Bayesian classifier for vessel segmentation. An ensemble system of bagged and boosted decision trees is proposed in [25] to produce segmentation result based on a collection of features including the filtered responses and morphological operations. Authors in [5] introduce a gradient boosting approach to learn discriminative convolutional filters. Based on segmentation results of [5], the authors of [26] propose a learning based iterative scheme to detect and connect weak vessel fragments by latent classification trees. A discriminatively trained fully connected CRF model is introduced in [27], where segmentation is formulated as inferring the maximum a posteriori assignment in a conditional random field. In [28], a set of structural contextual features are extracted and fed into gradient boosting trees for pixelwise classification. Remarkable results have been attained recently by the deep learning methods. In [29], a U-shape network structure has been studied that utilizes the short-cut connection as an expending path to produce an image-to-image segmentation. This network structure, also referred to as U-net, has since been adopted in a wide range of medical image analysis tasks. Li et al. [30] regard the segmentation task as a cross-modality transformation task and develop a neural network based prediction model. Inspired by [31,32], the authors of [33] design convolutional neural networks (CNNs) with side-output layers to learn feature representations. It also contains a conditional random field layer to take into account of global pixel correlations. DRIU [4] considers a multi-task learning approach addressing both vessel segmentation and optic disc detection in a single CNN model. The work of [34] performs supervised segmentation of un-annotated new dataset using cross-domain synthesized training images. This is achieved by adopting the generative adversarial networks to synthesize retinal images having the textural appearance of the target images while maintaining the vessel structure annotations from existing benchmark datasets.

Image matting is a problem that is closely related to segmentation. Originally developed in the film-making industry, its applications mainly focus on looking at humans and natural scenes. As described in [35], there is an alpha matte channel that linearly interpolate between the foreground and the background, focusing on the unknown regions of the trimap. Specifically, a pixel of the input image, x_i, is assumed to be a convex sum of the background value b_i and the foreground value f_i using the alpha channel $m_i \in [0, 1]$:

$$x_i = m_i f_i + (1 - m_i) b_i. \tag{1}$$

A Bayesian matting algorithm is proposed in [36], which utilizes a set of local Gaussians to learn the distributions of local foreground and background. A widely used method is proposed by [37], where a cost function is derived from local smoothness assumptions on foreground and background, and the optimal matte is attained by solving the incurred linear system of equations. KNN matting [38] applies K-nearest-neighbor matching in the feature space to approximate nonlocal neighborhoods without sophisticated assumptions and advanced sampling strategies. Deep learning models have also been developed in recent years for getting the alpha matte [39–41], in which compositional loss is utilized to solve the alpha matting problem with convolutional network structures.

3. Our approach

The main aim of our approach is to improve the segmentation results of the existing segmentation methods with a score map of the input retinal image. Fig. 1(a) illustrates our proposed pipeline. First, the image segmentation problem is transformed into a closely related matting problem to focus on segmenting the unknown regions. This requires a trimap which is obtained by bi-level thresholding of the score map. It is then passed through an end-to-end matting network to produce the final foreground matte.

3.1. Transforming to image matting

More formally, let us denote the RGB retinal image as $\boldsymbol{x} \in \mathbb{R}^{W \times H \times 3}$, the segmentation ground truth as $\boldsymbol{y} \in \{0, 1\}^{W \times H}$, the

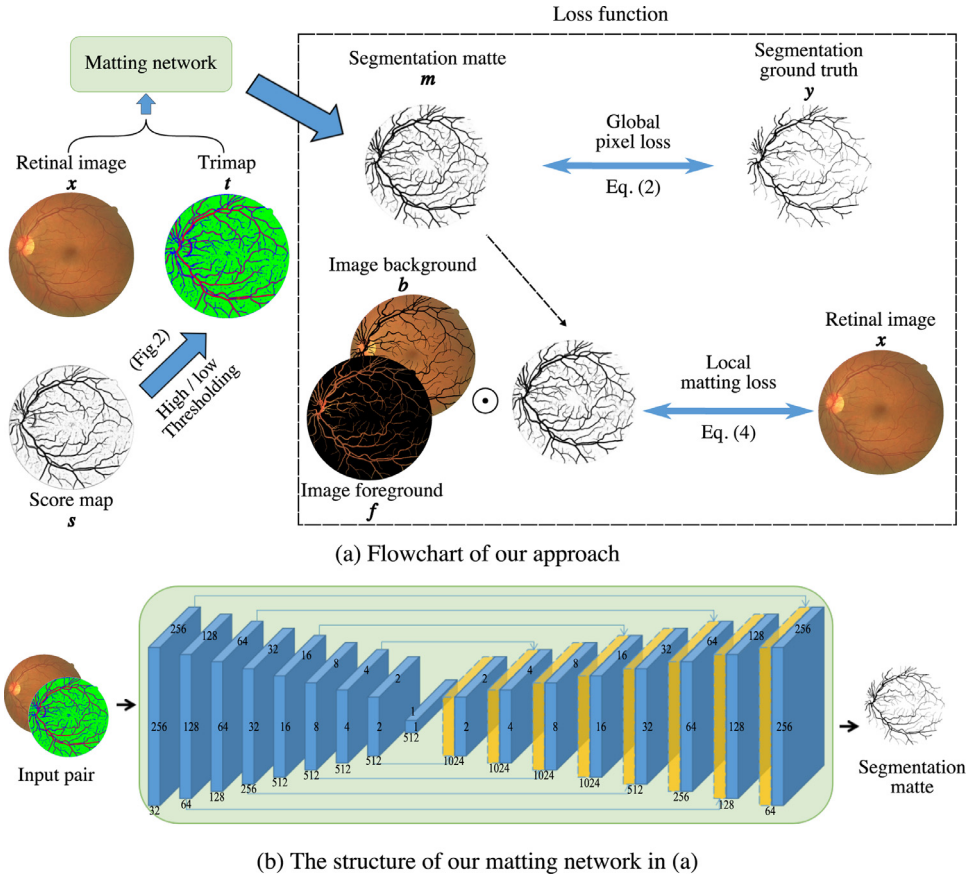


Fig. 1. (a) Flowchart of our approach. Left panel can be regarded as testing stage operations, while the boxed right panel displays two elements of our loss function that play a key role in learning the matting function $\mathcal{M}_\xi$ during the training stage. (b) Details of our encoder-decoder matting network structure. A dense U-net network structure, where each 3D block corresponds to a specific layer of feature maps; The yellow-color blocks represent the feature maps copied over from the encoder counterpart. See text for details.

score map from baseline method as $\boldsymbol{s} \in [0, 1]^{W \times H}$. Here W and H are the image width and height, respectively. The corresponding trimap is $\boldsymbol{t} \in \{0, 0.5, 1\}^{W \times H}$, which is obtained by applying bi-level thresholds γ_l and γ_h on the score map $\boldsymbol{s}$, as presented in Fig. 2. The trimap $\boldsymbol{t}$ contains three types of pixels: The 0-valued and 1-valued pixels are obtained by the bi-level thresholds $< \gamma_l$ and $> \gamma_h$, respectively; the 0.5-valued pixels are in the unknown region, which corresponds to those pixel locations indexed by i satisfying $\gamma_l \le s_i \le \gamma_h$ for score s_i of pixel i.

In preparing the trimap, we have the following assumption regarding the two thresholds, γ_l and γ_h: the two thresholds are set to be sufficiently low and high values on the score map respectively, to ensure the existence of minimum amount of false negatives or missings, if not zero, at this starting stage. In image matting terms, these pixels corresponds to the so-called *definite background* and *definite foreground*, respectively. Note it is acceptable to introduce additional false positives, which in our context are nothing more than a few extra pixels in the unknown areas. Empirically this assumption holds well for most existing segmentation methods. The big vessels including the main trunks grown out of the optic disk tend to be highly ranked in the score map, while the majority background pixels attain very low scores. Consequently they are readily thresholded as either the definite foreground with value 1 in the trimap, or the definite background with value 0. The main uncertainly lies on the small vessels or weak signals around the vessel boundaries – they usually fall under the unknown regions with value 0.5 in the trimap, and these pixels are precisely the places where our attention should be focused on.

Now, we have obtained the aforementioned trimap containing three types of areas: definite background, definite foreground, and unknown region; they are displayed in Fig. 2 as green, red, and blue pseudo-colors, respectively. The follow-up matting process [35] thus concentrates on the blue-color unknown regions containing small vessels and blurry vessel boundaries, to extract meaningful foreground vessel structure. In other words, it is to learn a ξ-parameterized function mapping of $\mathcal{M}_\xi(\boldsymbol{x}, \boldsymbol{t}) \to \boldsymbol{m} \in \mathbb{R}^{W \times H}$ that takes as input a retinal image $\boldsymbol{x}$ and its trimap $\boldsymbol{t}$, to predict a matte $\boldsymbol{m}$. The parameters ξ denotes the neural network weights to be learned at the training stage. The final segmentation result $\hat{\boldsymbol{y}}$ is thus obtained by applying a global threshold τ as $\hat{y}_i = \mathbf{1}(m_i \ge \tau)$ for each of the pixels indexed by i. Here $\mathbf{1}$ is the indicator function, m_i and $\hat{y}_i$ refer to the matte and segmentation values at pixel i, respectively. To deal with this problem, an end-to-end matting pipeline is proposed as in Fig. 1(b). The matting network takes as input the color retinal image and the trimap produced by a baseline segmentation method. In particular, as highlighted in right panel of Fig. 1(a), a local matting loss is introduced together with a global pixel loss function, which plays a key role for our approach in delivering superior segmentation performance. We will describe these aspects in the following subsections.

3.2. Model structure

As shown in Fig. 1(b), our matting model is a ξ-parameterized deep convolutional neural net function $\mathcal{M}_\xi$. It follows the U-net structure of [29] with dense concatenations from the encoder

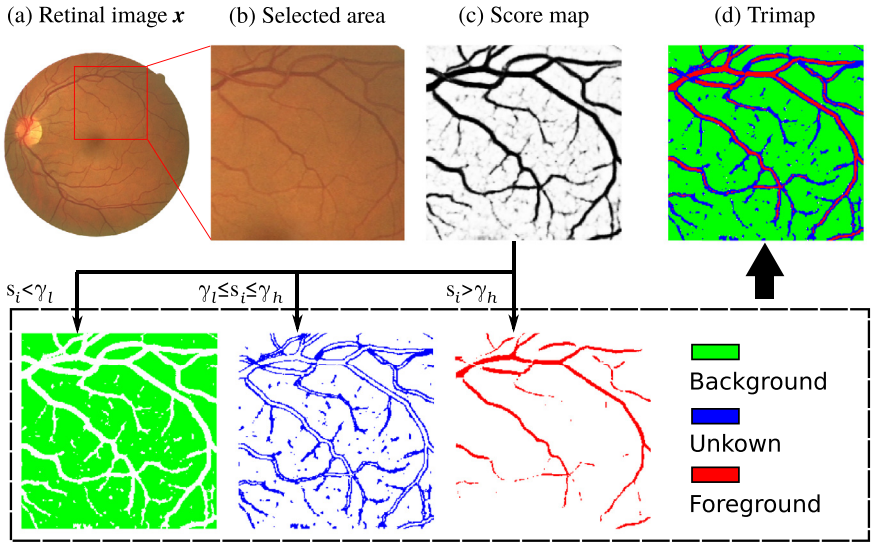


Fig. 2. A trimap illustration. (a) Original retinal image $\boldsymbol{x}$. (b) Zoomed-in view of selected area. (c) The score map from an existing segmentation method. (d) The corresponding trimap obtained by applying bi-level thresholds of γ_l and γ_h. s_i indicates the ith pixel value in the score map $\boldsymbol{s}$. γ_l and γ_h are the low and high thresholds, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

layers to the corresponding decoder layers. The input to our network is the concatenation of retinal image $\boldsymbol{x}$ and trimap $\boldsymbol{t}$, while the output is the segmentation matte $\boldsymbol{m}$. The encoder part of our model is stacked by several convolutional layers to downsample the feature maps to a vector code. The multiple convolutional layers are responsible for extracting salient features and capturing from local to global representations. The decoder part, on the other hand, employs subsequent transposed convolutional (also called deconvolutional) layers to reconstruct a sequence of feature maps from the code, and produce the segmentation matte in the end. The concatenation of feature maps between the convolutional layer and their counterpart deconvolutional layers are important in retaining the global structural information. This operation also allows the back-propagating gradients to pass directly from the decoder layers to the encoder ones, to avoid possible issues of gradient vanishing. In our network model, a kernel size of 3 is adopted for both convolutional and deconvolutional layers with a stride of 2 for downsampling and upsampling purposes, respectively. Detailed information such as number of filters is described in Fig. 1(b). A batchnorm layer and *relu* activation function are followed after each convolutional layer. At the last layer, the *sigmoid* activation function is used to squash the output values between zero and one.

3.3. Loss function

During training, the loss function plays an essential role of learning the optimal parameters ξ. For each training example $\boldsymbol{x}$, a pair of foreground and background ground-truths, $\boldsymbol{f}$ and $\boldsymbol{b}$, are constructed from the segmentation ground-truth $\boldsymbol{y}$, as $f_i = y_i \cdot x_i$ and $g_i = (1 - y_i) \cdot x_i$, respectively, for each pixel indexed by i. Now, to train the encoder-decoder network $\mathcal{M}_\xi$, our loss function consists of the following two parts, the global pixel loss and the local matting loss. The global pixel loss takes into account the pixel-wise squared difference between the ground-truth $\boldsymbol{y}$ and predicted matte $\boldsymbol{m}$:

$$l_{pix}(\boldsymbol{y}, \boldsymbol{m}) = \sum_i (y_i - m_i)^2. \tag{2}$$

This loss helps to make a rough segmentation of the vessels foreground from the background, which may lead to an overall reasonable result from the viewpoint of a whole image. It nonetheless performs less effective on small vessels. This is to be expected, as the set of big vessel pixels constitutes the majority of the foreground, thus they dominate the final loss function value. As a result, it may ignore the errors incurred from wrongly segmenting those small vessels, which are often weak signals that are more difficult to be distinguished from the background.

Inspired by the matting definition, a local matting loss is derived from (1), which focuses more on the large errors introduced locally, by taking into account the composition law formed in the image matting process. As expressed in Eq. (1), an input image is decomposed pixelwise as a convex sum of the foreground ground-truth $\boldsymbol{f}$ and background ground-truth $\boldsymbol{b}$ by the segmentation matte $\boldsymbol{m}$. Then our loss is defined as the squared difference between the real input retinal image $\boldsymbol{x}$ and the assemble counterpart of image foreground $\boldsymbol{f}$ and image background $\boldsymbol{b}$ following the composition law in a pixel-by-pixel manner:

$$l_{\text{mat}}(\boldsymbol{y}, \boldsymbol{m}) = \sum_i (f_i \cdot m_i + b_i \cdot (1 - m_i) - x_i)^2. \tag{3}$$

Moreover, from the aforementioned assumption, matting mistakes in $\boldsymbol{m}$ usually occur within the unknown regions of trimap $\boldsymbol{t}$. Thus our proposed loss module could focus attention to these unknown regions of small vessels and blurred boundaries, as:

$$l_{\text{mat}}(\boldsymbol{y}, \boldsymbol{m}) = \sum_i \mathbf{1}(s_i == 0.5) \cdot (f_i \cdot m_i + b_i \cdot (1 - m_i) - x_i)^2. \tag{4}$$

Finally, our loss function is defined as

$$L(\boldsymbol{y}, \boldsymbol{m}) = \omega l_{pix} + (1 - \omega) l_{\text{mat}}. \tag{5}$$

Here ω is a trade-off constant, which is empirically set to 0.5 in this paper. With the help of both global and local losses, our model is adaptable to properly segment both the main vessel trunks and the small vessel fragments.

4. Empirical experiments

4.1. Benchmark datasets and implementation details

Empirically we have examined on three retinal fundus image benchmark datasets. They are DRIVE [23], STARE [14], and HRF [42]. DRIVE contains 40 color retinal images with a FOV of

Table 1
Quantitative segmentation evaluation results on the benchmark datasets of DRIVE, STARE and HRF. Baseline segmentation methods include supervised learning methods: Kernel Boost [5], DRIU [4], as well as unsupervised method MSLD [6]. Other state-of-the-art methods are also considered, including FC-CRF [27], BCOSFIRE [15], LCMBoost [26], DeepVessel [33] and LAD-OS [22]. Results are reported in F1-score (%), Recall (i.e. Sensitivity) (%), Precision (%) and Specificity (%).

		Kernel Boost		DRIU		MSLD		*FC-CRF*	*BCOSFIRE*	*LCMBoost*	*DeepVessel*	*LAD-OS*
		Baseline	Ours	Baseline	Ours	Baseline	Ours					
DRIVE	F1-score	75.88	81.13	81.62	**82.29**	72.98	81.15	78.57	78.73	76.00	79.26	78.22
	Recall	77.23	80.78	82.46	**83.29**	66.11	81.30	78.97	78.67	74.58	78.65	76.50
	Precision	74.58	**81.48**	80.80	81.31	81.44	81.05	78.54	78.87	78.62	79.82	80.12
	Specificity	96.79	97.76	97.61	97.67	98.58	97.68	97.92	97.98	98.00	98.11	**98.19**
STARE	F1-score	77.30	79.16	82.56	**83.51**	77.74	80.66	78.74	78.42	78.72	79.12	80.11
	Recall	75.94	78.03	83.34	**84.33**	74.15	81.13	77.73	79.18	75.74	79.50	80.27
	Precision	78.71	80.32	81.79	**82.71**	81.70	80.24	80.45	77.78	83.02	78.79	80.04
	Specificity	98.33	98.48	98.50	98.57	98.63	98.37	98.50	98.13	**98.67**	98.24	98.37
HRF	F1-score	75.84	77.31	76.93	**78.13**	58.56	77.29	71.80	70.30	-	-	77.47
	Recall	74.34	76.45	77.11	**78.09**	53.50	76.14	72.12	76.32	-	-	76.51
	Precision	77.45	78.22	76.78	78.20	64.88	**78.53**	72.14	65.34	-	-	78.50
	Specificity	98.20	97.95	98.06	98.18	97.53	98.01	97.63	96.59	-	-	**98.26**

45° at 584×565 pixel resolution. We follow the standard partition of DRIVE and split it into the training and the testing sets, each containing 20 images. STARE includes 20 images captured at 35° FOV with a resolution of 700×605 pixels. Following the convention, the first 10 STARE images are used for training, and the rest are for testing. HRF contains 45 high resolution images with size of 3304×2336, where we adopt a train/test split of 22/23 images, respectively. A separate model is trained on different dataset but with the same model structure and training settings.

Our approach is implemented in Python, and the Tensorflow library is used for building the proposed matting neural net. The training data are augmented with rotations over every 45°, as well as horizontal and vertical flipping for larger training set. Throughout experiments, the network parameters are updated by Adam optimizer [43] with learning rate of 0.005. The kernel size of convolutional layer is set to 3 based on the parameter sensitivity study. The lower and upper thresholds are set to $\gamma_l = 0.2$ and $\gamma_h = 0.9$, respectively, which are found sufficient empirically to ensure minimum false negatives in constructing the trimap. Training time on DRIVE is around 25 min, and prediction on one image takes 0.5741 s. All empirical computation is carried out on a desktop PC with an Intel iCore 7 CPU, 16GB main memory, and a Titan-X GPU.

To demonstrate the effectiveness of our approach in boosting the performance over a range of existing methods, we have considered three state-of-the-art segmentation methods as the exemplar baselines. They include two supervised methods, Kernel Boost [5], and DRIU [4], as well as an unsupervised method, MSLD [6]. For these baselines, their original implementations are used.

4.2. Segmentation results

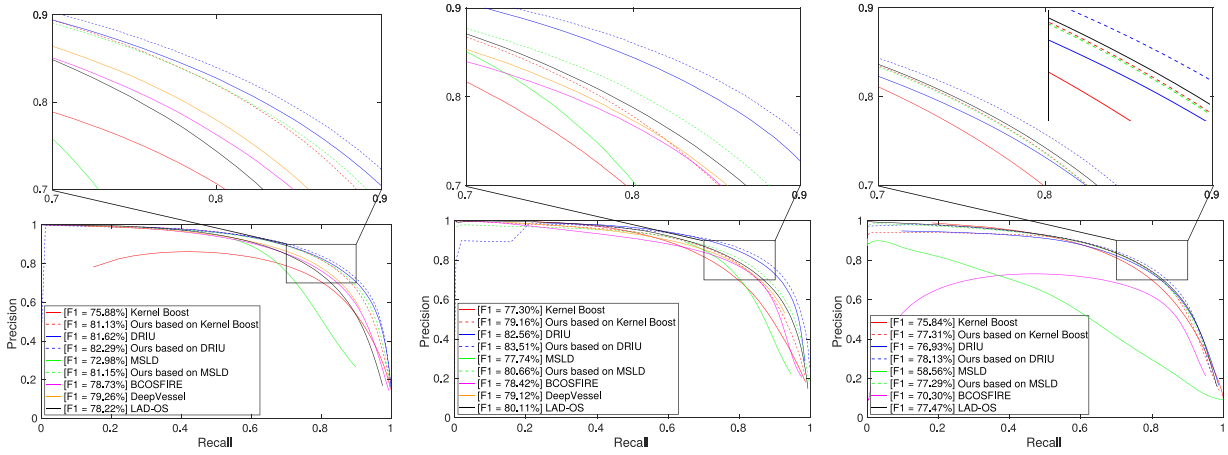
This section is to examine the segmentation performance on the full image level, which evaluates the overall performance including both the main trunks and the fine vessels. As mentioned above, three baseline methods (*i.e.* DRIU, Kernel Boost and MSLD) are engaged to provide the initial score maps for our approach.

Table 1 displays the quantitative results of these baseline methods (*i.e.* Baseline), as well as the corresponding results after applying our approach (*i.e.* Ours). To characterize the vessel segmentation performance, four evaluation metrics are considered, including the F1-score, sensitivity, precision, and specificity. The metric of *sensitivity* is also referred to as *recall*. F1-score delivers an overall performance summary, while the other three can form two groups of paired metrics. One pair of metrics is precision and recall, and another pair is sensitivity and specificity. These two pairs of metrics can provide detailed performance indications from two different perspectives.

Moreover, to better connect our performance on the benchmark datasets to the literature, the results of a number of additional state-of-the-art segmentation methods have also been included, which are FC-CRF [27], BCOSFIRE [15], LCMBoost [26], DeepVessel [33] and LAD-OS [22]. Here, the reported performance of most comparison methods (FC-CRF, LCMBoost, DeepVessel, LAD-OS) are obtained by evaluating the results provided by authors of the respective papers. As to BCOSFIRE, its performance is evaluated by executing the original implementation with default parameters on our side.

Results based on the Kernel Boost baseline are shown in 3rd to 4th columns, while those of DRIU are displayed in 5th and 6th columns and those of MSLD are presented in 7th and 8th columns. Specifically, under the DRIU-tagged columns, *Baseline* is from the original DRIU model, while *Ours* report our results with initial trimaps based on these DRIU score maps. Take the DRIU baseline on DRIVE dataset as an example, which achieves a rather high F1-score of 81.62%, while our approach still improves to 82.29%. Notably, the recall rate of our approach gains an additional 0.83% over that of baseline, suggesting the capability of extracting more vessel pixels. At the same time, its precision rate is also increased from 80.80% to 81.31%. This clearly demonstrates the overall improvement from the baseline. Similar observations are also drawn from the STARE and HRF benchmarks, where we usually witness gains from most of the four metrics after adopting our approach. Furthermore, for the Kernel Boost baseline, our approach produces more visible gains: 5.25% and 1.86% for DRIVE and STARE, respectively. This we attribute to the large performance deficit of the Kernel Boost baseline from the DRIU baseline, which achieves only 75.88% vs. 81.62% of DRIU. For the unsupervised baseline of MSLD, a F1-score of 72.98% is obtained. And our approach is able to lift up to 81.15%. Overall, our approach seems capable of enhancing the performance over various baselines with score map, being supervised or unsupervised.

More detailed information is provided from the precision-recall curves of Fig. 3(a–c) are the PR curves attained on DRIVE, STARE, and HRF datasets, respectively. Our approach outperforms the baseline methods, especially in the range of [0.7, 0.9] over both precision and recall axes, which is the central important zone in most applications. The zoomed-in view of this region is also displayed in figures. The clear margin between the dotted lines over the solid lines of the same colors demonstrate that our approach indeed helps in advancing the performance over the baselines, not only on a single point, but also on the PR curve.



(a) DRIVE (b) STARE (c) HRF

Fig. 3. (a–c) are the three plots of precision-recall curves about the competing segmentation methods, for DRIVE, STARE and HRF datasets, respectively. To demonstrate the effectiveness of our approach, we consider as baselines three state-of-the-art methods of DRIU [4], Kernel Boost [5], and MSLD [6], with their results presented in solid lines of specific colors. Results of our approach augmented with the baselines are also provided in dotted lines of the same colors. For clarity, we also show the zoomed-in view focusing on the important area (*i.e.* [0.7, 0.9] of both axes) of the PR curves.

Fig. 4 displays the visual segmentation results. Original retinal images are presented in the first column. It is followed by the baseline results in second column, our results in the third column, and the ground-truths in the fourth column. The zoomed-in views are displayed beside the segmentation results in (b-d) for more local details. It is observed that in general our approach preserves more fine details along the boundaries and capture more small vessels than the counterpart baselines. Moreover, different from the DRIU baseline, the Kernel Boost baseline tends to falsely detect more of the background pixels as the foreground, these false alarms subsequently lead to a low precision. In this situation, our approach is shown to not only capture more small vessels, but also have much less of these false alarms, which results in a visually cleaner segmentation map. This observation is also quantitatively supported by Table 1, where higher values of both precision and recall are achieved by our approach.

To put our segmentation results into perspective, we also compare with other retinal vessel segmentation methods. Among all these competing methods, our DRIU-based approach achieves the best results with F1-score of 82.29%. Consider Kernel Boost method: it has a F1-score of 75.88% on DRIVE, which falls behind the 78.57% of *FC-CRF*. Our Kernel Boost-based approach brings up the performance to 81.13%, overpassing that of FC-CRF by a large margin. It is worth mentioning that this improvement is obtained with a good balance between precision and recall as is reflected in Table 1. These observations are also held true for the other two benchmarks, STARE and HRF.

4.3. Ablation tests

To further examine the advantage of the proposed loss terms, we also carry out the ablation tests on the DRIU baseline and the DRIVE dataset. Table 2 displays the quantitative results of the models trained on different loss terms. We observe that models trained with only one loss term can hardly boost the baseline method. We start with the model trained with only the global pixel loss. It has a F1-score of 81.89%, a slight 0.27% increase over the baseline performance. In the meanwhile, the recall rate improves around 0.5%, which comes at a cost of decrease in precision of 0.43%. Similarly, the model trained with only the matting loss attains a much higher recall score of 84.52%, and gains 2.06% improvement. On the other hand, more wrong detections are taking place and lead to a larger decrease of precision. The F1-score of 81.53% is slightly worse comparing to the baseline. The trained model is able to achieve the best F1-score of 82.29%, with a more balanced precision and recall values when both loss terms are incorporated

We also study the influence of thresholds of bi-level trimap. Table 3 displays the experimental tests with different thresholds on DRIVE dataset and DRIU baseline method. We can observe from the table that such pairs of γ_l and γ_h can improve the performance of baseline method. One extreme situation is also considered, in which $\gamma_l = 0$ and $\gamma_h = 1$. When $\gamma_l = 0, \gamma_h = 1$, values of all the pixels are 0.5 and the whole image is considered as the unknown region. In such situation, our approach can still produce a satisfied result with F1-score of 80.92. In our experiments, we adopt $\gamma_l = 0.2$ and $\gamma_h = 0.9$ based on this parameter study.

4.4. Segmentation performance evaluation on small vessels vs. big vessels

Empirically it has been observed that although big vessels including the main trunks are relatively easy to be dealt with, segmenting small vessels is always a challenging issue. The above evaluation on the whole image level demonstrates the superior

Table 2
Ablation tests where results are produced by models trained with different loss terms on the DRIU baseline and the DRIVE benchmark dataset. Quantitative evaluation metrics considered here include F1-score (%), Recall (i.e Sensitivity) (%) and Specificity (%).

	F1-score	Recall	Precision	Specificity
Baseline	81.62	82.46	81.31	97.61
Ours with only the global pixel loss	81.89	82.94	80.88	97.60
Ours with only the local matting loss	81.53	**84.52**	78.85	97.23
Ours with both loss terms	**82.29**	83.29	**83.34**	**97.67**

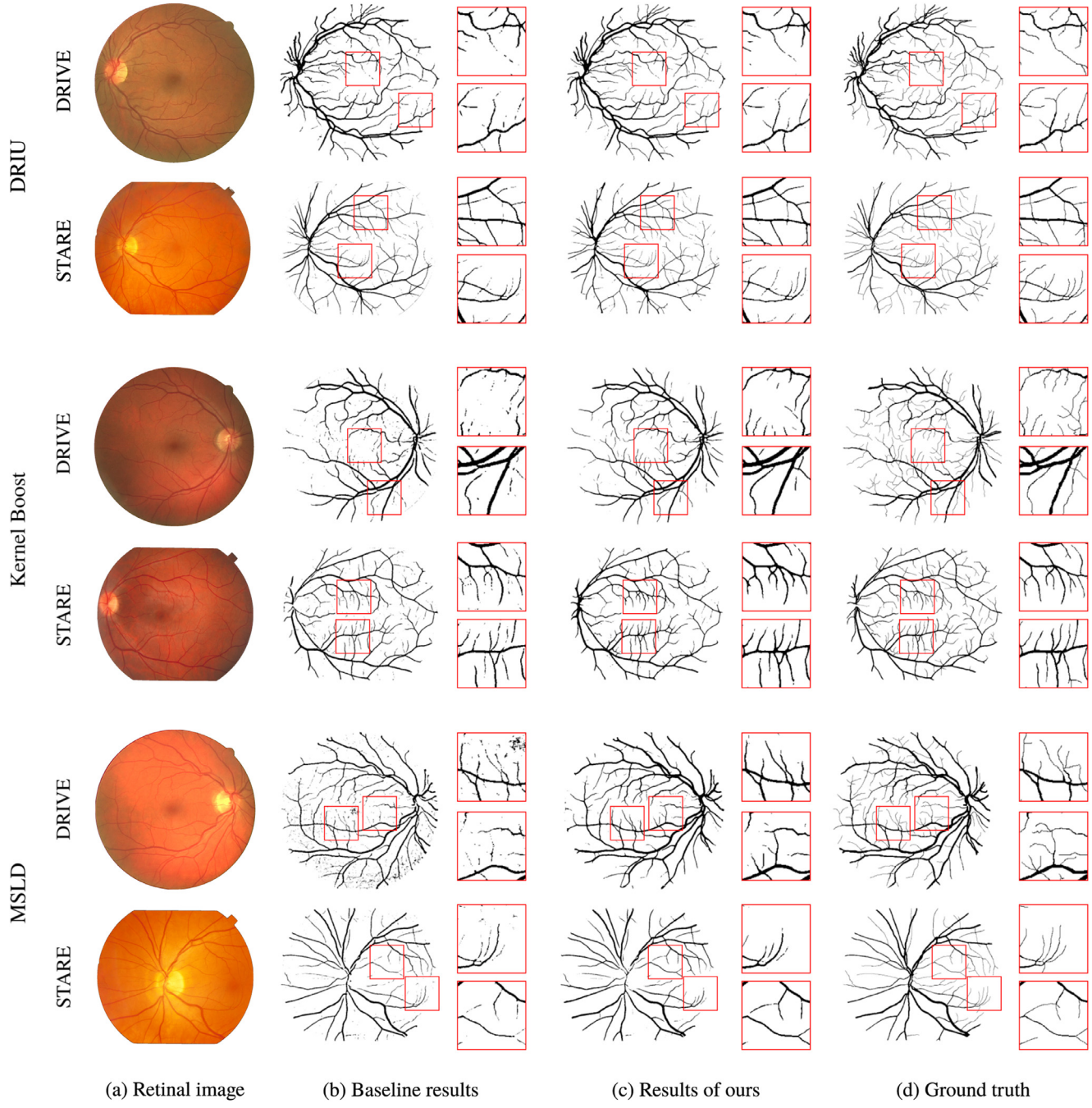


Fig. 4. Visual exemplar results of the baseline methods and our approach. The first panel is the results of the DRIU baseline and ours on DRIVE and STARE. Similarly, the second panel displays the results of Kernel Boost and ours, while the last one shows the results of MSLD and ours. See text for details.

performance of our approach. Here we further examine how our approach performs on these small vs. big vascular structures.

This calls for a new segmentation ground-truth annotation such that the vessel pixels are categorized into either small or big vessels. In this paper, we utilize the following automated process to generate these annotations. We first utilize morphological opening operation which engages a disk structure element with a radius of 0.5 to extract small vessels from the segmentation ground-truths. The connected pixels with area larger than 100 pixels are then removed from the remaining segments. The segmentation map is further cleaned by discarding regions smaller than 8 pixels. Empirically this automatic pipeline can produce satisfactory small vessel extraction results. Two exemplar annotated segmentation maps of small vessels are presented in Fig. 5, where *blue* refers to the collection of small vessel pixels, *black* represents the big ones.

Table 3
Parameter sensitivity tests where results are produced by models trained with different thresholds of bi-level (γ_l / γ_h) trimap on the DRIU baseline and the DRIVE benchmark dataset. Quantitative evaluation metrics considered here is F1-score (%).

γ_l / γ_h	0.1 / 0.8	0.2 / 0.8	0.1 / 0.9	0.2 / 0.9	0 / 1
F1-score	81.89	81.92	81.87	**82.29**	80.92

Table 4 quantitatively summarizes the results of three baselines and ours on small vs. big vessels. Note the numbers here are not

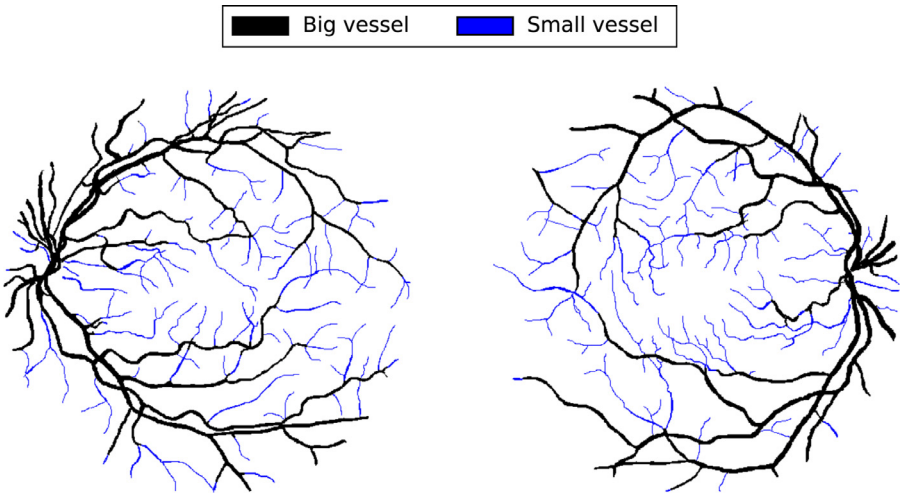


Fig. 5. Two exemplar segmentation ground-truths of small vs. big vessels. Here *blue* represents the small vessel pixels, while *black* shows the big vessel pixels. See text for details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to be directly compared with those in the other sections. For large vessels, our approach leads to slightly better performance. More significant gains over the baselines are evident from the experiments of small vessels, that include not only the overall F1-score, but also the individual precision and recall rates. The increase of the recall rates is particularly prominent, which demonstrates a strong capability in retrieving those more challenging foreground vessel pixels. The results of Table 4 suggest that our approach with the proposed loss function is good at restoring the fine vessels, as is visually presented in Fig. 4.

4.5. Evaluation of vessel structure using the CAL metric of [44]

So far, we have considered several evaluation metrics including F1-score, recall, precision, and specificity. These metrics are based on pixel-wise comparison between the segmentation map and the corresponding ground-truth, which are criticised by some works including [44] as not matching well with the human quality perception in delineating the vessel trees. Instead, one may consider the connectivity-area-length (CAL) score [44] that has been also considered in e.g. [45,46] as an alternative evaluation metric of vessel segmentation. The CAL score contains three components: C is the vessel connectivity; A refers to overlapping area between segmented image and ground-truth; and L represents the length of the extracted vessel. Finally, it is summarized by a single CAL score, which is computed by the production of these three components. The value of all these values is within the range of [0, 1], and the higher the better.

Table 5 thus provides a quantitative evaluation in terms of the CAL metric over the well-known benchmarks of DRIVE and STARE. It is observed that our approach excels in all the three distinct aspects of C, A, and S, as well as the final CAL score. The better performance in C suggests that the vessels extracted by our approach contains less isolated fragments. The higher values of A and L indicate better detection of vessels in regard to both region area and vessel skeleton length. For example, the DRIU baseline achieves an overall CAL score of 0.8205 on DRIVE, a result better than the Kernel Boost baseline. Meanwhile, our DRIU-based approach further advances the performance to 0.8471.

4.6. Evaluation on a different imaging modality

In addition to the fundus cameras, empirical experiment is also carried out on a different imaging device, the scanning laser opthalmoscopy, also known as SLO. More specifically, the IOSTAR [47] dataset is considered, which contains 24 images taken

Table 4
Quantitative evaluation on small vs. big vessels over baselines and ours on DRIVE and STARE datasets. The baselines include the state-of-th-art supervised methods of Kernel Boost [5] and DRIU [4]. Results are reported in F1-score (%), Recall (i.e. Sensitivity) (%), Precision (%) and Specificity (%).

			Kernel Boost		DRIU		MSLD	
			Baseline	Ours	Baseline	Ours	Baseline	Ours
Small vessels	DRIVE	F1-score	46.59	55.46	57.24	59.75	56.04	57.24
		Recall	48.37	57.22	61.61	64.29	58.81	61.61
		Precision	45.10	54.02	53.61	55.99	53.67	53.61
		Specificity	98.56	98.80	98.70	98.77	98.70	98.75
	STARE	F1-score	48.50	51.98	61.46	61.83	49.66	56.95
		Recall	50.08	51.64	66.62	67.43	51.37	60.02
		Precision	47.95	52.70	57.21	57.55	49.57	54.46
		Specificity	99.43	99.50	99.47	99.47	99.46	99.47
Big vessels	DRIVE	F1-score	86.61	88.29	88.57	88.81	82.66	88.01
		Recall	86.32	87.67	89.57	89.71	73.69	89.36
		Precision	86.98	88.99	87.70	87.99	94.39	86.76
		Specificity	98.80	98.99	98.83	98.86	99.59	98.73
	STARE	F1-score	83.79	84.85	87.06	87.85	84.25	85.79
		Recall	79.85	82.28	86.13	86.65	77.45	84.18
		Precision	88.33	87.77	88.04	89.10	92.57	87.55
		Specificity	99.27	99.23	99.20	99.27	99.58	99.17

Table 5
Performance evaluation measured by the connectivity-area-length (CAL) score of [44]. See text for details.

	Method		C	A	L	CAL
DRIVE	Kernel Boost	Baseline	0.9903	0.8501	0.7697	0.6485
		Ours	0.9971	0.9248	0.8598	0.7933
	DRIU	Baseline	0.9962	0.9377	0.8780	0.8205
		Ours	0.9969	0.9476	0.8964	0.8471
	MSLD	Baseline	0.9687	0.8685	0.7847	0.6614
		Ours	0.9968	0.9245	0.8581	0.7913
STARE	Kernel Boost	Baseline	0.9905	0.8610	0.8103	0.6966
		Ours	0.9970	0.8810	0.8432	0.7438
	DRIU	Baseline	0.9943	0.9184	0.8829	0.8068
		Ours	0.9977	0.9306	0.9114	0.8466
	MSLD	Baseline	0.9795	0.8774	0.8199	0.7088
		Ours	0.9966	0.9054	0.8734	0.7897

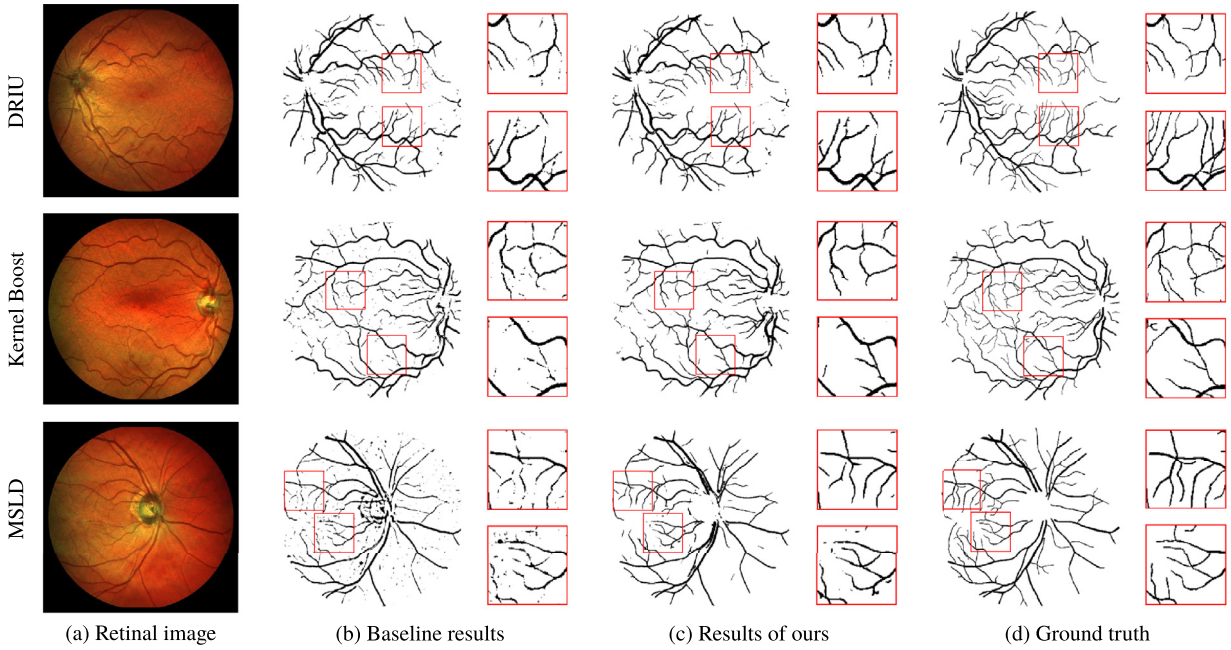


Fig. 6. Visual exemplar results of baseline methods and our approach on IOSTAR dataset. Rows from top to bottom display three baseline respectively.

Table 6
Quantitative evaluation on IOSTAR dataset acquired by SLO imaging (i.e. not the fundus imaging). Here the Kernel Boost and DRIU baselines are considered. Results are reported in F1-score (%), Recall (i.e. Sensitivity) (%), Precision (%) and Specificity (%).

	Kernel Boost		DRIU		MSLD	
	Baseline	Ours	Baseline	Ours	Baseline	Ours
F1-score	75.14	76.69	78.80	79.31	71.32	76.88
Recall	73.60	76.70	80.28	80.68	68.68	77.25
Precision	76.86	76.81	77.44	78.02	74.37	76.55
Specificity	97.84	97.73	97.72	97.80	97.70	97.70

with a FOV of 45° and an image size of 1024×1024. Here the first 12 images are employed as the training set, with the rest serving as the testing images. Table 6 displays the quantitative results, where our approach is shown to also improve over the three different baselines, Kernel Boost, DRIU and MSLD. Quantitative comparisons with the baseline results are displayed in Fig. 6, the improvement over the small vessels are particularly noticeable.

5. Conclusion and outlook

In this paper, we present a new approach to improve existing retinal image segmentation methods. Our approach transforms the segmentation problem into a matting task with a trimap. It is achieved by the proposed local matting loss and global pixel loss as well as the matting network. Experiments on different datasets demonstrate the effectiveness of our approach that works particularly well in delineating small vessels using the proposed local matting loss. Our approach also works well with different baseline methods, which provides a wide application prospect. In addition to widely-used fundus image benchmarks, our approach is also demonstrated to work well with the SLO images from the IOSTAR benchmark. Looking forward, we would like to continue working with eye angiography images, as well as 3D OCT images.

Acknowledgment

The project is partially supported by the Singapore A*STAR grants. We thank Zhen Guan for his help with this project.

References

[1] J.J. Kanski, B. Bowling, Clinical Ophthalmology: a Systematic Approach, Elsevier Health Sciences, 2011.
[2] E.J. Sussman, W.G. Tsiaras, K.A. Soper, Diagnosis of diabetic eye disease, JAMA 247 (23) (1982) 3231–3234.
[3] T.Y. Wong, R. Klein, F.J. Nieto, B.E. Klein, A.R. Sharrett, S.M. Meuer, L.D. Hubbard, J.M. Tielsch, Retinal microvascular abnormalities and 10-year cardiovascular mortality: a population-based case-control study, Ophthalmology 110 (5) (2003) 933–940.

[4] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, L. Van Gool, Deep retinal image understanding, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 140–148.
[5] C. Becker, R. Rigamonti, V. Lepetit, P. Fua, Supervised feature learning for curvilinear structure segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 526–533.
[6] U.T. Nguyen, A. Bhuiyan, L.A. Park, K. Ramamohanarao, An effective retinal blood vessel segmentation method using multi-scale line detection, Pattern Recognit. 46 (3) (2013) 703–715.
[7] B. Fang, W. Hsu, M.L. Lee, Reconstruction of vascular structures in retinal images, in: Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, 2, IEEE, 2003, pp. II–157.
[8] F. Zana, J.-C. Klein, Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation, IEEE Trans. Image Process. 10 (7) (2001) 1010–1019.
[9] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 1998, pp. 130–137.
[10] P. Bankhead, C.N. Scholfield, J.G. McGeown, T.M. Curtis, Fast retinal vessel detection and measurement using wavelets and edge location refinement, PloS One 7 (3) (2012) e32435.
[11] M.M. Fraz, S.A. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, An approach to localize the retinal blood vessels using bit planes and centerline detection, Comput. Methods Programs Biomed. 108 (2) (2012) 600–616.
[12] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, M. Goldbaum, Detection of blood vessels in retinal images using two-dimensional matched filters, IEEE Trans. Med. Imaging 8 (3) (1989) 263–269.
[13] Y. Wang, G. Ji, P. Lin, E. Trucco, Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition, Pattern Recognit. 46 (8) (2013) 2117–2133.
[14] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, IEEE Trans. Med. Imaging 19 (3) (2000) 203–210.
[15] G. Azzopardi, N. Strisciuglio, M. Vento, N. Petkov, Trainable cosfire filters for vessel delineation with application to retinal images, Med. Image Anal. 19 (1) (2015) 46–57.
[16] E. Ardizzone, R. Pirrone, O. Gambino, F. Scaturro, Automatic extraction of blood vessels, bifurcations and end points in the retinal vascular tree, in: 13th International Conference on Biomedical Engineering, Springer, 2009, pp. 22–26.
[17] Y. Yin, M. Adel, S. Bourennane, Retinal vessel segmentation using a probabilistic tracking method, Pattern Recognit. 45 (4) (2012) 1235–1244.
[18] J. Zhang, H. Li, Q. Nie, L. Cheng, A retinal vessel boundary tracking method based on Bayesian theory and multi-scale line detection, Comput. Med. Imaging Graph. 38 (6) (2014) 517–525.
[19] C.L. Srinidhi, P. Aparna, J. Rajan, Automated method for retinal artery/vein separation via graph search metaheuristic approach, IEEE Trans. Image Process. 28 (6) (2019) 2705–2718.
[20] B. Yin, H. Li, B. Sheng, X. Hou, Y. Chen, W. Wu, P. Li, R. Shen, Y. Bao, W. Jia, Vessel extraction from non-fluorescein fundus images using orientation-aware detector, Med. Image Anal. 26 (1) (2015) 232–242.
[21] Y. Zhao, L. Rada, K. Chen, S.P. Harding, Y. Zheng, Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images, IEEE Trans. Med. Imaging 34 (9) (2015) 1797–1807.
[22] J. Zhang, B. Dashtbozorg, E. Bekkers, J.P. Pluim, R. Duits, B.M. ter Haar Romeny, Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores, IEEE Trans. Med. Imaging 35 (12) (2016) 2631–2644.
[23] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501–509.
[24] J.V. Soares, J.J. Leandro, R.M. Cesar, H.F. Jelinek, M.J. Cree, Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification, IEEE Trans. Med. Imaging 25 (9) (2006) 1214–1222.
[25] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, S.A. Barman, An ensemble classification-based approach applied to retinal blood vessel segmentation, IEEE Trans. Biomed. Eng. 59 (9) (2012) 2538–2548.
[26] L. Gu, L. Cheng, Learning to boost filamentary structure segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 639–647.
[27] J.I. Orlando, E. Prokofyeva, M.B. Blaschko, A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images, IEEE Trans. Biomed. Eng. 64 (1) (2017) 16–27.
[28] L. Gu, X. Zhang, H. Zhao, H. Li, L. Cheng, Segment 2d and 3d filaments by learning structured and contextual features, IEEE Trans. Med. Imaging 36 (2) (2017) 596–606.
[29] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[30] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, T. Wang, A cross-modality learning approach for vessel segmentation in retinal images, IEEE Trans. Med. Imaging 35 (1) (2016) 109–118.
[31] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1395–1403.
[32] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, P.H. Torr, Conditional random fields as recurrent neural networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1529–1537.
[33] H. Fu, Y. Xu, S. Lin, D.W.K. Wong, J. Liu, Deep vessel: retinal vessel segmentation via deep learning and conditional random field, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 132–139.
[34] H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, L. Cheng, Supervised segmentation of un-annotated retinal fundus images by synthesis, IEEE Trans. Med. Imaging 38 (1) (2019) 46–56.
[35] J. Wang, M.F. Cohen, Image and video matting: a survey, Found. Trends Comput. Graph.Vis. (2008) 97–175.
[36] Y.Y. Chuang, B. Curless, D.H. Salesin, R. Szeliski, A Bayesian approach to digital matting, in: IEEE Conference on Computer Vision and Pattern Recognition, 2001, pp. 264–271.
[37] A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2008) 228–242.
[38] Q. Chen, D. Li, C.-K. Tang, KNN matting, IEEE Trans. Pattern Anal. Mach. Intell. 35 (9) (2013) 2175–2188.
[39] D. Cho, Y.-W. Tai, I. Kweon, Natural image matting using deep convolutional neural networks, in: European Conference on Computer Vision, Springer, 2016, pp. 626–643.
[40] X. Shen, X. Tao, H. Gao, C. Zhou, J. Jia, Deep automatic portrait matting, in: European Conference on Computer Vision, Springer, 2016, pp. 92–107.
[41] N. Xu, B. Price, S. Cohen, T. Huang, Deep image matting, in: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, IEEE, 2017, pp. 311–320.
[42] T. Köhler, A. Budai, M.F. Kraus, J. Odstrčilik, G. Michelson, J. Hornegger, Automatic no-reference quality assessment for retinal fundus images using vessel segmentation, in: Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on, IEEE, 2013, pp. 95–100.
[43] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv:1412.6980.
[44] M.E. Gegúndez-Arias, A. Aquino, J.M. Bravo, D. Marin, A function for quality evaluation of retinal vessel segmentations, IEEE Trans. Med. Imaging 31 (2) (2012) 231–239.
[45] W.S. Oliveira, J.V. Teixeira, T.I. Ren, G.D. Cavalcanti, J. Sijbers, Unsupervised retinal vessel segmentation using combined filters, PloS One 11 (2) (2016) e0149943.
[46] B. Sheng, P. Li, S. Mo, H. Li, X. Hou, Q. Wu, J. Qin, R. Fang, D.D. Feng, Retinal vessel segmentation using minimum spanning superpixel tree detector, IEEE Trans. Cybern. (2018).
[47] S. Abbasi-Sureshjani, I. Smit-Ockeloen, J. Zhang, B.T.H. Romeny, Biologically-inspired supervised vasculature segmentation in slo retinal fundus images, in: International Conference Image Analysis and Recognition, Springer, 2015, pp. 325–334.

He Zhao received B.E. degree from Beijing Institute of Technology, China in 2014. He is currently a Ph.D. candidate at Beijing Institute of Technology, China. His research interest is medical image processing, deep learning, computer vision.

Huiqi Li received Ph.D. degree from Nanyang Technological University, Singapore in 2003. She is currently a professor at Beijing Institute of Technology. Her research interests are image processing and computer-aided diagnosis.

Li Cheng received Ph.D. degree from University of Alberta, Canada in 2004. He is currently an associate professor at University of Alberta. His research interests are computer vision and machine learning.