Biomedical Signal Processing and Control 44 (2018) 110–126

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

A visual attention guided unsupervised feature learning for robust vessel delineation in retinal images

Chetan L. Srinidhi[a,*], P. Aparna[a], Jeny Rajan[b]

[a] Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, India
[b] Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal, India

ARTICLE INFO

Article history:
Received 8 January 2018
Received in revised form 2 April 2018
Accepted 15 April 2018

Keywords:
Retinal image
Vessel segmentation
Visual attention
Unsupervised feature learning

ABSTRACT

Background and objective: Accurate segmentation of retinal vessels from color fundus images play a significant role in early diagnosis of various ocular, systemic and neuro-degenerative diseases. Segmenting retinal vessels is challenging due to varying nature of vessel caliber, the proximal presence of pathological lesions, strong central vessel reflex and relatively low contrast images. Most existing methods mainly rely on carefully designed hand-crafted features to model the local geometrical appearance of vasculature structures, which often lacks the discriminative capability in segmenting vessels from a noisy and cluttered background.

Methods: We propose a novel visual attention guided unsupervised feature learning (VA-UFL) approach to automatically learn the most discriminative features for segmenting vessels in retinal images. Our VA-UFL approach captures both the knowledge of visual attention mechanism and multi-scale contextual information to selectively visualize the most relevant part of the structure in a given local patch. This allows us to encode a rich hierarchical information into unsupervised filtering learning to generate a set of most discriminative features that aid in the accurate segmentation of vessels, even in the presence of cluttered background.

Results: Our proposed method is validated on the five publicly available retinal datasets: DRIVE, STARE, CHASE_DB1, IOSTAR and RC-SLO. The experimental results show that the proposed approach significantly outperformed the state-of-the-art methods in terms of sensitivity, accuracy and area under the receiver operating characteristic curve across all five datasets. Specifically, the method achieved an average sensitivity greater than 0.82, which is 7% higher compared to all existing approaches validated on DRIVE, CHASE_DB1, IOSTAR and RC-SLO datasets, and outperformed even second-human observer. The method is shown to be robust to segmentation of thin vessels, strong central vessel reflex, complex crossover structures and fares well on abnormal cases.

Conclusions: The discriminative features learned via visual attention mechanism is superior to hand-crafted features, and it is easily adaptable to various kind of datasets where generous training images are often scarce. Hence, our approach can be easily integrated into large-scale retinal screening programs where the expensive labelled annotation is often unavailable.

1. Introduction

Retinal fundus image provides rich information about the early manifestation of various diseases related to the eye, cardiovascular and neurodegenerative diseases [1–3]. These images are often routinely acquired for non-invasive examination of its anatomical components such as vessel tree, optic disc and the fovea. Segmentation and quantification of retinal vessel tree provide important clinical biomarkers through the analysis of its geometrical properties, that aid in early diagnosis of various diseases such as diabetes [1], stroke [4], hypertension [5], arteriosclerosis [6] and cerebral small vessel diseases. Numerous longitudinal studies have been conducted in the past, that shows a strong and consistent link between retinal microvasculature with incident clinical stroke, hypertension, cardiovascular and various neurodegenerative diseases [4,5,7]. Changes in microvascular geometrical patterns such as vessel width, tortuosity, fractal dimension and branching angle provide an early insights into the progression of aforementioned diseases. Therefore, an accurate delineation of retinal vessel

* Corresponding author.
E-mail addresses: srinidhipy@gmail.com (C.L. Srinidhi), p.aparnadinesh@gmail.com (P. Aparna), jenyrajan@gmail.com (J. Rajan).

https://doi.org/10.1016/j.bspc.2018.04.016
1746-8094/

structures from color fundus image is of interest. Manual segmentation of vessel tree is often tedious, time-consuming and prone to large intra and inter-observer variability. In addition, manual delineation often requires careful interpretation of images, which is very cumbersome and painful especially for large population-based screening programs. Hence, there is a need for the automated segmentation of retinal vessels for the accurate quantification of vascular changes, along the entire disease course.

Segmenting retinal vessel tree is extremely challenging due to multi-scale nature of varying vessel caliber, the presence of strong central vessel reflex, the close proximity of pathological lesions, close parallel and highly curved vessels and complex crossover regions. In the past decades, many solutions have been proposed with techniques ranging from conventional matched filtering [8] to more recent convolutional neural network [9,10] based approaches. The existing techniques can be broadly divided into two main categories: unsupervised and supervised methods [11]. Supervised methods require a set of manually labelled training images for classifying a pixel in a previously unseen image. Whereas, unsupervised methods can segment the vessels without requiring any manual labelled annotations. In general, most of the existing techniques mainly rely on carefully designed hand-crafted filters to inherently model the local geometrical appearance of vessel structures. For example, the hand-crafted filters such as Gabor filters [12], multi-scale derivative of Gaussian [13], matched filters [8], ridge detector [14], line detector [15], wavelet transform [16,17], moment invariant features [18], first and second order derivatives of Gaussian [19] and response of COSFIRE filters [20,21] to name a few. These hand-crafted filters were designed based on complex domain knowledge and require careful parameter tuning to achieve optimal segmentation performance, across a wide variety of data. Besides, the response of these filters often poorly represents the appearance of thin vessel structures, crossover and bifurcation regions, highly curved tortuous vessels and susceptible to non-illumination present in an image. The other drawback of these methods is that the extracted features often lack the discriminative capability to predict the actual class label, even in the presence of similar looking cluttered objects.

To overcome the aforementioned limitations, various automatic feature learning algorithms [9,10,22,23] have been proposed to learn the feature representations directly from the training data. These approaches are primarily motivated by the success of deep learning (DL) methods, which is applied in various computer vision applications such as object recognition, scene classification, semantic segmentation, etc. The success of these methods is critically dependent on an enormous amount of labelled training data which is typically expensive in medical imaging applications. To address these shortcomings, various unsupervised feature learning (UFL) algorithms [24–28] have been proposed to automatically learn the feature representations, only from a set of unlabelled data. Automatic feature learning enables to encode rich hierarchical information that learns to map complex functions from input to output, directly from the data, without depending on hand-crafted features.

The main bottleneck for the accurate segmentation of retinal vessels comes from the multiscale nature of varying vessel caliber, poor visibility of low contrast thin vessels, the proximal presence of pathological structures and poor vessel connectivity at complex junction locations. The automatic features learned from these challenging locations often lack discriminative capability in accurately identifying vessel pixels from similar-looking cluttered objects. This is mainly because, the traditional UFL approaches encodes feature representation from a limited input patch size, which is often referred to as "*receptive field*". The selection of this receptive field mainly depends on the object of interest, which we are trying to encode. In case of retinal vessels, the structure of interest varies significantly, resulting in difficulty in choosing an appropriate receptive field size, that fits for a wide range of input data. Further, various size of receptive fields encodes different *contextual* information (inter-pixel dependencies), thereby resulting in different feature representation for the same pixel centred on a patch.

To alleviate this problem, we propose a novel UFL approach which is primarily inspired by the visual attention mechanism in the human visual system where we humans have the capability to pay attention *selectively* to the part of the image, instead of processing the whole scene in its entirety [29,30]. Such a selection mechanism is often referred to as "*visual attention prediction*". In this work, we leverage this idea of visual attention mechanism with the unsupervised feature learning approach, to automatically learn the most relevant hierarchical features from unlabelled data. The proposed visual attention guided unsupervised feature learning (VA-UFL) approach automatically learns to selectively pay attention to the most relevant part of the structure in a given input patch, and use this information to selectively encode features for subsequent classification. The feature learned via this approach offers an edge over tradition UFL methods – by exploring both the notion of *selection mechanism* and the *multi-scale contextual information* under a single framework. This allows us to learn the most relevant hierarchical features at multiple scales, which encodes sufficient information about the most salient region as well as the multi-scale context for a given input patch. The advantage of our proposed approach is that we encode a complex hierarchical level of information with such a simple approach, without relying on sophisticated feature learning architectures.

The key contributions of this paper can be summarized as follows.

1. We propose to explore the idea of visual attention mechanism that learns to selectively pay attention to the most relevant structure *and* capture multi-scale contextual information from a given local patch. This, in turn, drives the subsequent UFL approach to encode the rich hierarchical information for the automatic feature discovery in retinal images.
2. We show that the proposed visual attention model, when leveraged with the UFL approach, aims to automatically learn the most discriminative set of features that underscores the inter-class differences (between the vessel and background pixels) to be large, while keeping intra-class differences (between vessel pixels) to be small, resulting in the accurate segmentation of retinal vessels.
3. We validated our approach on five publicly available retinal datasets, including both the RGB and SLO images. The extensive experimental analysis demonstrates the effectiveness of the proposed approach in handling all the challenging cases, compared with the state-of-the-art methods.

The rest of this paper is organized as follows. In Section 2, relevant literature to this work is briefly reviewed. Section 3 explains in detail the proposed framework. In Section 4, we provide information about the dataset used, quantitative measures employed, parameter settings, followed by experimental results. Section 5 discusses the key findings, followed by future directions and finally, we conclude in Section 6.

2. Background

In past decades, many methods have been proposed for the automatic segmentation of vessels from retinal color fundus images. These methods can be broadly divided into two main

categories: unsupervised [8,16,20,31–41] and supervised methods [13,14,12,15,42,18,19,43,44,10,45,17].

Unsupervised methods rely mostly on rule-based techniques such as matched filtering [8,20,31,32,16,33], vessel tracking [34–36,46], thresholding [37] and model-based approaches [38–41]. These methods do not depend on prior knowledge of manual annotations and typically faster, since no training is involved. Methods based on matched filtering (MF) involves convolving an image with a 2D-Gaussian template. These methods generally model vessel cross-sectional intensity profile as a 2D-Gaussian function [8]. Several variants of MF have been proposed in combination with the first order derivative of Gaussian, to reduce false responses at non-vessel locations [33]. Azzopardi et al. [20] proposed a combination of shifted Gabor filter responses that selectively respond to linear bar-shaped structures, such as vessels. Kovacs et al. [31] proposed a self-calibration based template matching and contour reconstruction technique that can be used to segment vessels in images of different resolutions. Yin et al. [32] developed an orientation-aware detector, which is designed based on the fact that the vessels are locally oriented and elongated structures. Most of the conventional MF techniques often fail to model vessels in the presence of strong central vessel reflex (CVR) and highly curved tortuous vessels, where the Gaussian vessel profile assumption does not hold anymore. To overcome this limitation, Zhang et al. [16] proposed a filtering technique based on maximizing the multi-scale second-order Gaussian derivative filter responses in the orientation score domain. Their approach was successful in segmenting vessels with CVR, thin vessels and complex crossover structures.

Vessel tracking methods starts from an initial seed point and iteratively trace the entire vasculature by following the vessel centerline, based on local information. These methods generally provide accurate vessel width information and precise vessel connectivity at branching and crossover locations. In Yin et al. [35], an initial set of vessel edges are identified, followed by Gaussian curve fitting to the cross-sectional intensity profile for estimating the local vessel appearance. Bekkers et al. [34] proposed a multi-orientation framework to automatically track the local vessel edges. De at al. [36] proposed a transductive learning-based approach to solve the crossover issue encountered during vessel tracing. In general, the performance of tracking-based approaches rely heavily on proper initialization of initial seed points and also depends on the robustness of the tracking algorithm. A fast iterative global thresholding based approach is proposed in [37], that iteratively adds new vessel pixels to the initial estimate of the segmented vasculature, until a stopping criterion is met. Model-based approaches consists of vessel profile [38] and deformable models [39,40] to segment the retinal vasculature. Lam et al. [38] proposed a multiconcavity modelling approach to handle both normal and abnormal images, simultaneously. Methods based on the active contour is presented in [39–41] which takes into account the local geometrical information as well as the image intensity, to simultaneously enhance and segment the retinal vessels.

In contrast, supervised methods aim to classify a pixel as vessel or background based on a prior knowledge learnt using a set of manually labelled training images. These methods often exhibit superior performance and are much more computationally expensive compared to unsupervised ones. Niemeijer et al. [13] presented a technique based on the multi-scale derivative of Gaussian matched filters to extract features, which are then trained using k-Nearest Neighbor (k-NN) classifier. Staal et al. [14] proposed a ridge based vessel extraction technique, where the ridge pixels are grouped into convex sets that approximate straight line elements. Soares et al. [12] extracted features based on multi-scale Gabor wavelet transform, followed by Gaussian mixture model (GMM) to classify the image pixels. A method based on basic line detector and support vector machines (SVM) is presented in Ricci et al. [15]. In Fraz et al. [42], a feature vector is designed based on the orientation analysis of gradient vector field, line strength measure, morphological transformation and Gabor filter response to train a decision tree classifier. The method based on the combination of invariant moments and gray-scale image intensity is proposed in [18] to train an artificial neural network (ANN) model. In Roychowdhury et al. [19], a hybrid technique is proposed based on first and second order gradient features, along with GMM classifier to segment both large and small vessel structures. Orlando et al. [43] proposed a fully connected conditional random field (CRF) approach to train a structured output SVM for learning the model parameters. Zhang et al. [17] proposed a vessel filtering and wavelet transform based features in orientation score domain, to simultaneously enhance and segment the vessel structures. Inspired by the success of DL methods, Li et al. [44] presented an auto-encoder based cross-modality approach to model the relationship between the retinal image and binary vessel map. The most recent method among DL based technique is presented in [10] which has shown to achieve an area under the curve (AUC) of 0.99, which is significantly greater than all previous methods. In general, most DL based approaches have shown to exhibit better performance in the presence of strong CVR, thin vessels, and complex crossover structures.

3. Methods

The overall framework of the proposed retinal vessel segmentation approach is illustrated in Fig. 1. Given an image, we first apply image preprocessing steps to remove non-uniform illumination and contrast variability. Next, the input patches are extracted at random locations from the preprocessed image. Each input patch is modelled with a visual attention mechanism by applying a retinal transformation. We then train a filter bank/dictionary using these patches based on K-means clustering. Given the learned filter bank and a set of labelled training images, we obtain features corresponding to each input patches. These features are then trained using random forest (RF) [47] classifier to predict the label of an unknown patch from the test image. The details are presented next.

3.1. Preprocessing

Retinal fundus image often exhibits non-uniform illumination acquired due to the geometrical properties of the retinal surface and various other complex imaging conditions such as, the presence of pathologies, pupil dilation and involuntary eye movement, etc. [48]. This greatly influences the performance of the segmentation methods mainly for the region close to periphery of the retina, which often exhibits many false detections. Further, luminosity and contrast variability pose a major problem for the accurate delineation of thin vessel structures. Therefore, in this work, we adopt the method proposed in [48] for retinal image enhancement, which is based on the geometrical property of the retinal surface. Compared to the most widely used method in the literature [49], the approach in [48] utilizes the non-uniform sampling scheme on the polar grid to better estimate the degradation component of the acquired image. A sample visualization of image pre-processing steps is shown in Fig. 2. The approach is shown to be robust in enhancing the thin vessel structures, which subsequently improves the vessel segmentation performance. In our experiments, we consider *only* the green channel of the retinal fundus image, since it exhibits better contrast between vessel and background, when compared to the other channels.

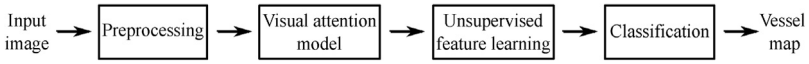

Fig. 1. An overview of the proposed retinal vessel segmentation approach.

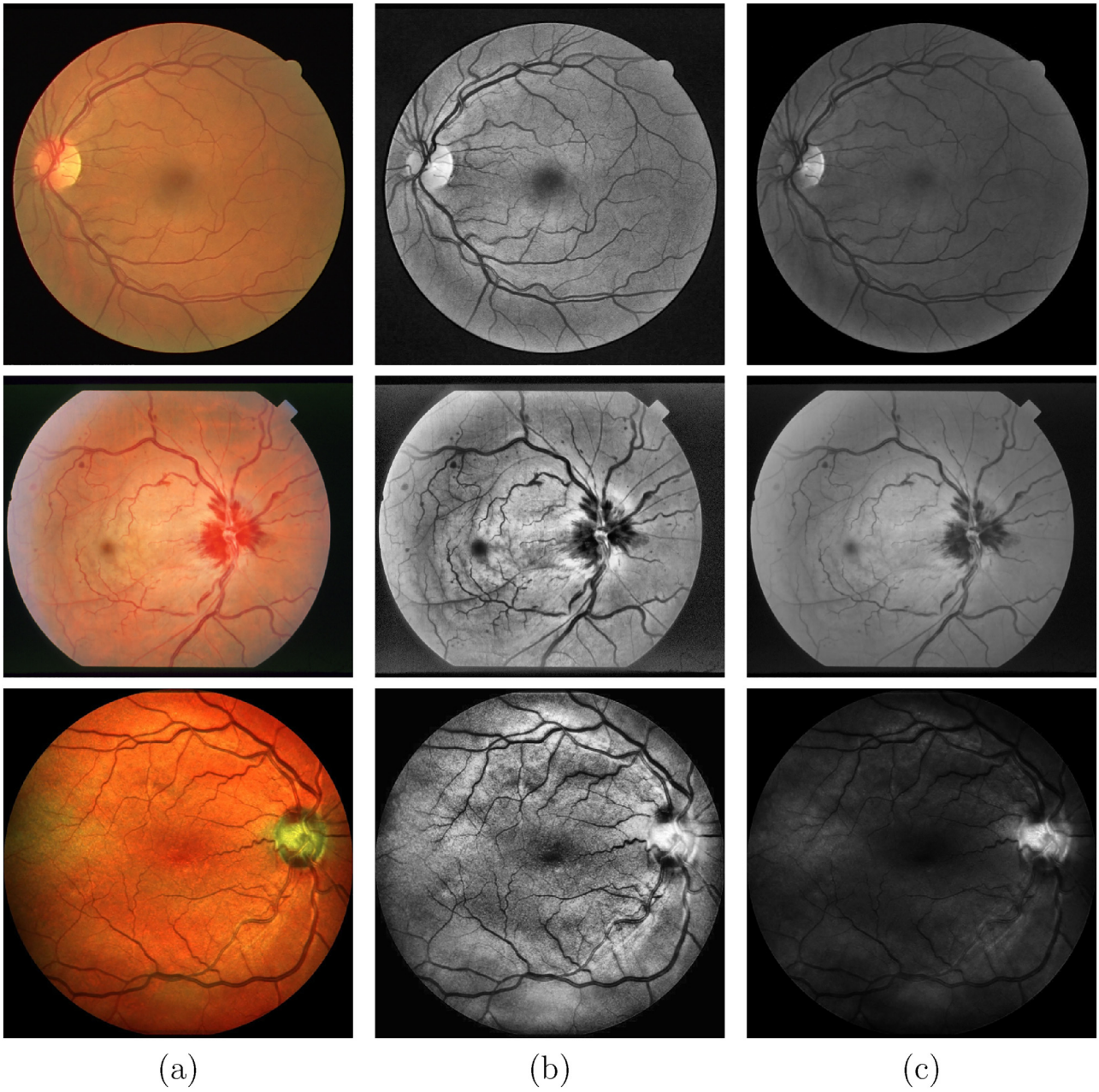

Fig. 2. Visualization of image pre-processing steps: (a) an example image, (b) enhanced image after applying the contrast limited adaptive histogram equalization (CLAHE) algorithm, (c) enhanced image after applying technique proposed in [48]. The images shown belongs to the DRIVE (first row), STARE (second row) and IOSTAR (third row) datasets.

3.2. Visual attention modelling

Given an image, the goal is to classify a patch centred on a pixel of interest, as belonging to a vessel or background. We sample patches $X = \{x_1, x_2, \ldots, x_m\}$ at random locations from an image as shown in Fig. 3. For each patch $x \in \mathbb{R}^p$, we assign a binary label $y \in \{0, 1\}$ where, 0 – denotes a background pixel; and 1 – denotes a vessel pixel. Let $Y = \{y_1, y_2, \ldots, y_m\}$ be the corresponding labels of the input data X.

We model the input patch $x \in \mathbb{R}^p$ centred on a pixel of interest (i, j) by constructing visual glimpses $G(i, j) = \{G_0, G_1, \ldots, G_m\}$. The visual glimpses consists of multi-scale patches $G_0, G_1, \ldots, G_m$ having dimension $p_0 \times p_0, p_1 \times p_1, \ldots, p_m \times p_m$ pixels, respectively, such that the dimension of the patch $(G_0 \ll G_1 \ll \cdots \ll G_m)$ as shown in Fig. 3. We then apply a retinal transformation T for each visual

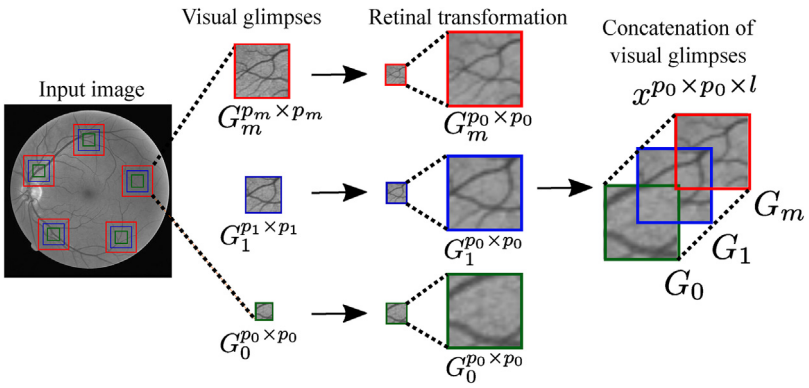

Fig. 3. An illustration of visual attention modelling.

glimpse G to model the visual attention mechanism. Retinal transformation is similar to fovea of the human retina where, it encodes a higher spatial resolution at a region close to the centre (pixel of interest), and progressively lower resolution as one moves away from the centre, in a non-linear fashion. This lower resolution representation is often referred to as "*visual glimpse*" [50]. The retinal transformation T is applied to visual glimpses G as $x = T(G(i, j))$; which is computed by down-sampling each multi-scale patches to the dimension equal to $(p_0 \times p_0)$ pixels using bi-cubic interpolation. Where, the glimpse $G_0^{p_0 \times p_0}$ – represents the patch with highest spatial resolution, and $(G_1^{p_0 \times p_0}, G_2^{p_0 \times p_0}, \ldots, G_m^{p_0 \times p_0})$ – corresponds to the successive lower resolution patches as shown in Fig. 3. This lower resolution patches often provide *contextual* information and captures complex vessel structures at multiple *scales* and multiple *orientations*. We then concatenate these visual glimpses G to form an input patch x of dimension $(p_0 \times p_0 \times l)$ where, $l = 0, 1, \ldots, m$ denotes the number of levels of visual glimpse G for a pixel location (i, j).

The intuition behind our approach is that, when humans tend to focus selectively on a particular object/scene, the region centred to the eye fixation point will exhibit higher spatial resolution than the region surrounding the object of interest [29]. The region centred to the eye fixation is regarded as the most important part of a scene (salient region) while, the region surrounding the object of interest often provides contextual information (inter-object relationship), which aid in accurate localization/classification of an object in the presence of clutter. This allows us to draw conclusion that at what level of spatial context, the extracted features contribute to the true label prediction of a patch centred on a pixel of interest. The visualization of sample visual glimpse patches for different retinal regions of interest is shown in Fig. 4. The goal is to classify the object of interest, which is the centre pixel of a patch, as either belonging to a vessel or background. The contextual information present in the region of interest (ROI) (see Fig. 4(a)) often lack the discriminative capability in predicting the true class label which is often surrounded by similar looking cluttered background. For instance – a dark lesion such as microaneurysm (see the first row of Fig. 4) which usually appears near the region of thin vessels, exhibits the similar visual appearance of high-curvature and junction points of thin vessels. Thus, posing a significant challenge in predicting the actual class label, given the limited neighbourhood information. While, the patches (see first row of Fig. 4(b) and Fig. 4(c)) provides much larger contextual information in which a lesion appears to be at the isolated location, often not connected to any vessel fragments.

Similarly, the region consisting of low contrast thin vessels, large and small vessel crossings, poor vessel connectivity at complex junction locations (see the second and third row of Fig. 4), often hinder the segmentation performance mainly for the detection of thin vessels. In contrast, by taking into account the most salient region – encoded by higher resolution glimpse (G_0) and the contextual information – encoded by lower resolution glimpses (G_1, G_2), provides the most discriminative features – by encoding the knowledge of both inter-structure relationship, as well as the most salient region in a given local patch. Thus, facilitating the subsequent classifier in accurately predicting the vessel pixels, even in the presence of similar looking background structures.

3.3. Visual attention guided unsupervised feature learning (VA-UFL)

The goal of any UFL algorithm is to learn a good latent representation from only unlabelled data U. Given an input patch x, we aim to train a dictionary D, which encodes the knowledge of the distribution of x. This trained dictionary D is often referred to as learned filter bank/weights. Next, we define an encoder $E(x;D) = s$, that maps an input patch $x \in \mathbb{R}^p$ to a new representation $s \in \mathbb{R}^K$ where, s is known as latent representation or feature vector. The feature vector (s) along with their true label can be passed to any machine learning classifier to predict the label of an unknown patch from s. The overall framework of the UFL approach is illustrated in Fig. 5.

Off-the-shelf, several unsupervised feature learning algorithms have been proposed in the past, such as sparse coding [24], sparse auto-encoders [51], restricted Boltzmann machines (RBMs) [26], denoising auto-encoders [27], K-means clustering [28], deep belief networks (DBN) [52] and many others as well. Among the approaches, K-means clustering [28] has been successfully applied for unsupervised filter learning, often competing with the state-of-the-art techniques due to its speed and scalability. In this work, we adopt the unsupervised learning module based on K-means clustering proposed in [28] for extracting a discriminative feature set, for the task of retinal vessel segmentation. The main reason behind choosing the K-means clustering as unsupervised filter learning is of two folds: K-means tends to learn the sparse projections of the input data – (i) given a sufficiently large amount of training images corresponding to the input dimensionality of the image; (ii) apply whitening to remove correlation between the data points. The above two assumptions hold true in our case where, there is an abundant amount of unlabelled data which is a typical scenario in the medical imaging domain – due to its expensive manual annotations. The major advantages of K-means over other UFL approaches are its speed, scalability and no hyper-parameter tuning involved.

Given a set of randomly sampled visual glimpse patches x, we construct a dataset $X = \{x_1, x_2, \ldots, x_m\}$, where $x \in \mathbb{R}^p$. Having

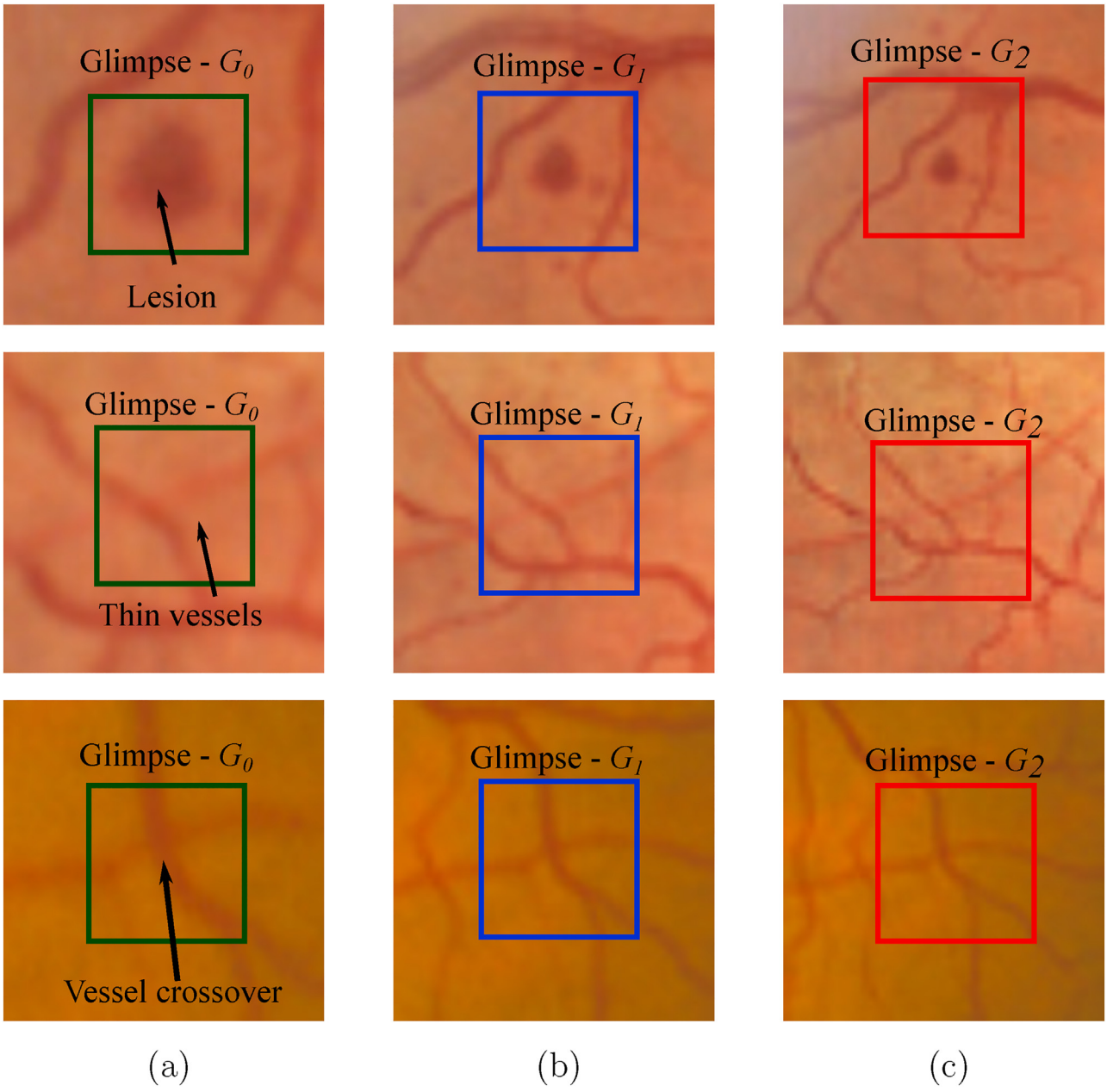

Fig. 4. Visualization of sample visual glimpse patches containing different retinal structures of interest (from top to bottom: retinal lesion (first row), area surrounded by thin vessels (second row) and a crossover location (third row). The object of interest is positioned at the centre pixel of a patch): (a) the higher resolution patch (20×20 pixels), (b) and (c) corresponds to the progressive lower resolution patches, obtained by first extracting a higher resolution patch of size 40×40 and 60×60 pixels respectively, followed by down-scaling to 20×20 pixels.

obtained input data X, we perform data preprocessing followed by unsupervised filter learning to learn a Dictionary D.

3.3.1. Data preprocessing

It is a standard practice in conventional deep learning frameworks to carry-out local brightness and contrast normalization, before generating features from the input data. Two types of preprocessing steps are generally performed: global contrast normalization and zero-phase component analysis (ZCA) whitening [53].

In global contrast normalization, we perform local brightness and contrast normalization by normalizing every patch x by subtracting the mean and dividing by the standard deviation of its elements. After normalization, we perform the patch level whitening using ZCA transform to remove the linear correlations among

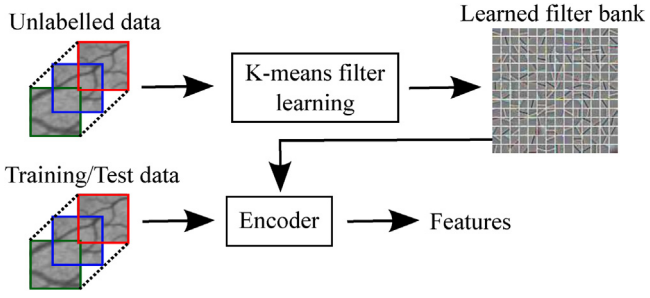

Fig. 5. An overview of the visual attention guided unsupervised feature learning (VA-UFL) framework.

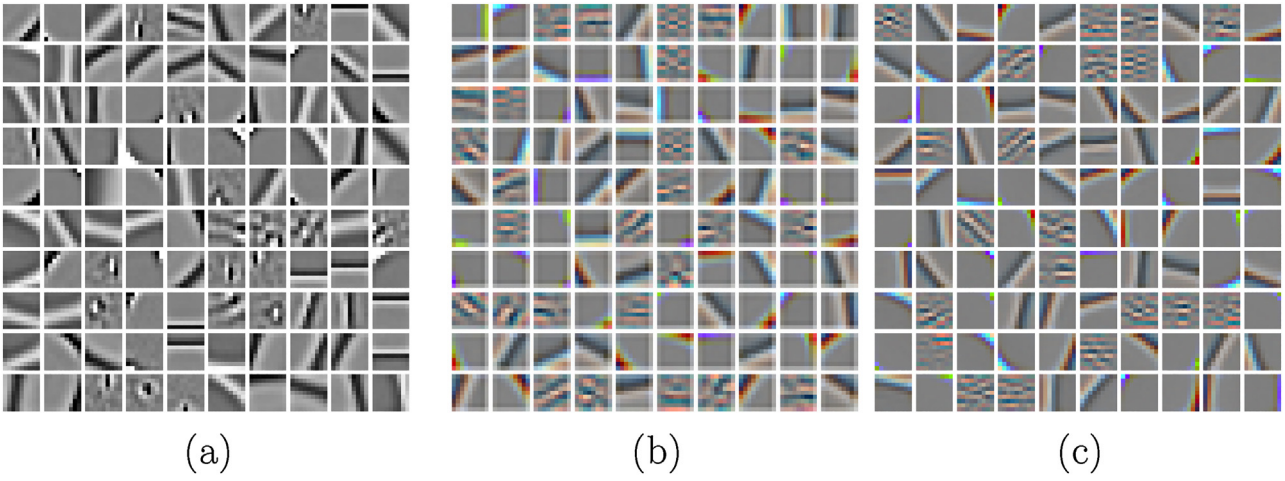

Fig. 6. Example bases learned using K-means for varying level (l) of input visual glimpses: (a) visual glimpse with $l = 0$, (b) visual glimpse with $l = 1$, (c) visual glimpse with $l = 2$.

neighbouring pixels. The ZCA transformation is summarized as follows. Given a normalized data X:

(i) find μ and Σ such that

$$\mu = \text{mean}\ (X) \quad \text{and} \quad \Sigma = \text{cov}\ (X) = \frac{1}{n}\sum_{i=1}^{n} XX^T; \tag{1}$$

(ii) find eigenvalue and eigenvector such that

$$\sum = V\Lambda V^T; \tag{2}$$

(iii) finally, the patch level whitening is computed as

$$x_{ZCA} = V(\Lambda + \epsilon_{ZCA}\mathbb{I})^{-1/2}V^T x, \tag{3}$$

where ϵ_{ZCA} is a small constant controlling the trade-off between whitening and noise amplification. For simplicity, we denote x_{ZCA} as x – which we use as an input to the rest of our pipeline.

3.3.2. K-means filter learning

Given the whitened data x, we learn a reconstructive dictionary $D \in \mathbb{R}^{p\times K}$ with K elements or atoms, which is accomplished by solving the following optimization problem

$$\langle D, s\rangle = \arg\min_{D,s} \|x - Ds\|_2^2 \quad \text{s.t.} \|s\|_0 \leq 1, \tag{4}$$

where $\|x - Ds\|_2^2$ denotes the reconstruction error, $s \in \mathbb{R}^K$ is the code vector corresponding to the input x, and $D = [d_1, d_2, \ldots, d_K] \in \mathbb{R}^{p\times K}$ ($K > p$, makes the dictionary over complete). Here each columns of dictionary are the centroids learned by K-means. A sample visualization of the learned dictionary (filter bank) for varying level (l) of input visual glimpses (G) is shown in Fig. 6. It is observed that the higher levels of input visual glimpses produces a more discriminative visual features such as sharp vessel edges (see Fig. 6(b) and (c)) compared to the noisy filter bases learned without attention modelling as shown in Fig. 6(a).

3.3.3. Encoder

An encoder is a function $E(x;D) = s$, which is defined as a non-linear mapping function that transforms a new input patch $x_i \in \mathbb{R}^p$ to a latent representation $s_i \in \mathbb{R}^K$ with D fixed. Where, s_i is the feature vector corresponding to a new input patch x_i.

More concisely, given D the learned filter bank using K-means, we find the latent representation s_i for a new input patch x_i by solving the following optimization problem

$$s_i = \arg\min_{s} \|x_i - Ds\|_2^2 + \lambda\|s\|_1, \tag{5}$$

where λ is a parameter that controls the reconstruction error and sparsity. The following L_1 optimization can be solved efficiently using approaches [24,54]. We then consider both the positive and negative components of sparse code s into separate features as: $s_K = max\{0, s_K\}$ and $s_{K+n} = max\{0, -s_K\}$. Note: the resulting feature vector dimension is $s \in \mathbb{R}^{2K}$. The overall steps our proposed VA-UFL algorithm is summarized in Algorithm 1.

Algorithm 1. Visual attention guided unsupervised feature learning (VA-UFL) approach.

Input: U – set of unlabelled images; x – the input patch centred on a pixel of interest; l – number of levels of visual glimpse; K – number of atoms in the dictionary.
Output: $\boldsymbol{D}$ – learned filter bank/dictionary; $\boldsymbol{s_i}$ – feature vector for a new input patch x_i.
1 Extract input patches $\boldsymbol{x}$ at random locations from a set of unlabelled images $\boldsymbol{U}$;
2 Apply retinal transformation $\boldsymbol{T}$ to each input patch x to construct visual glimpses $\boldsymbol{G}$ of $\boldsymbol{l}$ levels;
3 Pre-process the patch by normalizing brightness and contrast, as well as ZCA whitening defined in Eq. (3);
4 Train a dictionary $\boldsymbol{D}$ using K-means clustering defined in Eq. (4);
5 Compute the feature vector $\boldsymbol{s_i}$ for a new input train/test visual glimpse patch x_i via sparse encoding defined in Eq. (5);
6 **return** $\boldsymbol{D}$, $\boldsymbol{s_i}$.

3.4. Classification

To distinguish between the vessel and non-vessel pixels, we use a RF classifier [47] due to its capability of performing both classification and feature selection implicitly. It is robust against overfitting, outliers and high dimensional imbalanced data. This is true in our case, where only 9–14% of the total pixels belongs to vessels while, the rest belongs to non-vessels, leading to a highly skewed dataset.

A RF is a combination of N_T decision trees which are trained independently using bootstrap samples drawn with replacement from the training set. Each node in a tree is split using a randomly selected subset of m features ($m = \sqrt{d}$) (where d – is the dimensionality of the feature vector), which is chosen according to the decrease in the Gini index as recommended in [47]. The RF returns, for each selected feature input, the probability of being a vessel or non-vessel, based on the majority of the trees returning a positive

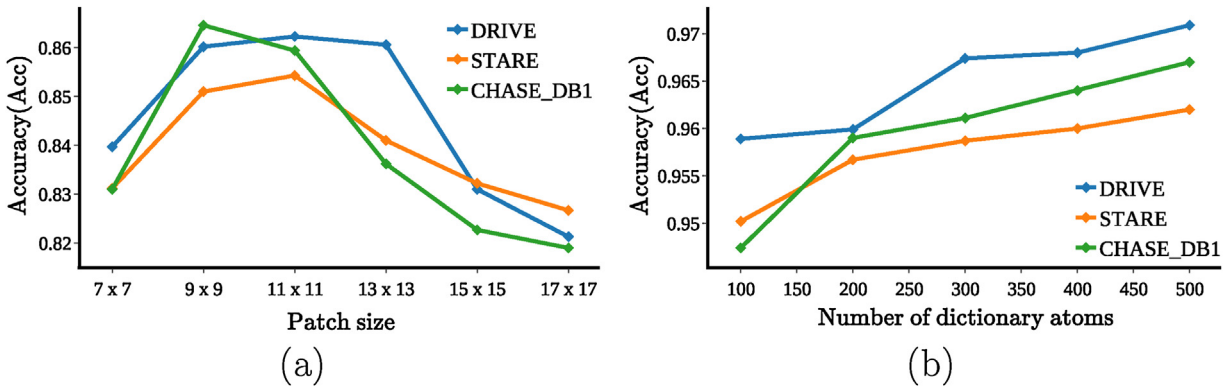

Fig. 7. Influence of parameter selection: (a) classification accuracy vs. varying input patch size (p_0), (b) classification accuracy vs. varying number of learned filters K.

response. We used publicly available RF implementation [55] for supervised classification. The default setting for the classifier is the number of trees $N_T = 100$ trees; and the feature vector dimension $d = (2 \times K) = (2 \times 100) = 200$ (where K – is the number of dictionary atoms).

During training, given a set of labelled training images $X_{(i)}$, $Y_{(i)}$, $i = 1, 2, \ldots, m$; where, $X_{(i)}$ are the training images; and $Y_{(i)} \in \{0, 1\}$ are the corresponding class labels (where, 0 – indicates a background pixel, and 1 – indicates a vessel pixel); the goal is to classify a patch $x_{(i)}$ centred on a pixel of interest as either vessel or background. In order to do this, we first extract visual glimpse patches $x_{(i)}$ for every pixels from each of the training images $X_{(i)}$. Given the learned dictionary D using unlabelled images U, we extract the features $s_{(i)}$ corresponding to each input visual glimpse patch $x_{(i)}$ using the encoder defined in Section 3.3.3. The feature set $s_{(i)}$ corresponding to each input patch $x_{(i)}$ along with their true label $y_{(i)}$ is used to train a RF classifier of N_T trees. In the testing stage, the probability of a patch $P(vessel = 1|s_{(i)})$, $\forall (x_{(i)}, y_{(i)})$ that belongs to a set of test image is estimated by feeding into the trained RF classifier.

4. Experiments and results

4.1. Datasets

To validate the proposed approach, five publicly available retinal datasets were used: DRIVE [14], STARE [56], CHASE_DB1 [42], and the two new Scanning Laser Ophthalmoscopy (SLO) image datasets namely IOSTAR [57], and RC-SLO [58].

The DRIVE dataset consists of 40 color fundus images with a resolution of 565 × 584 pixels (px). The dataset is divided into training and test sets, each of which contains 20 images. The STARE dataset includes 20 fundus images of resolution 700 × 605 px, out of which 10 contains pathological signs. A leave-one-out cross-validation was performed on STARE. The CHASE_DB1 contains 28 fundus images of resolution 1280 × 960 px, out of which first 20 images are used for testing and the last 8 images for training [42]. Manually segmented binary vessel maps are provided by two human annotators for DRIVE, STARE and CHASE_DB1 datasets. In addition, our method is also validated on the two new publicly available: IOSTAR and RC-SLO datasets. The IOSTAR consists of 30 – SLO images of resolution 1024 × 1024 px and the RC-SLO contains 40 image patches of resolution 360 × 320 px. All the vessel pixels in both the datasets are annotated by a group of experts. Half random split was employed for both IOSTAR and RC-SLO datasets. The performance metrics of the proposed segmentation approach are computed by considering all the pixels within the field of view (FOV).

4.2. Evaluation metrics

In-order to quantitatively compare the performance of our binary segmentation results with the corresponding manual ground truths, we obtain five different performance measurements based on: the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The following are the evaluation metrics that were used to compare the proposed approach with the state-of-the-art vessel segmentation methods: *Sensitivity* (*Se*) (also know as *Recall* (*Re*)), *Specificity* (*Sp*), *Accuracy* (*Acc*), *Matthews correlation coefficient* (*MCC*), *F1-score* (*F*1) and *Area under the ROC curve* (*AUC*).

$$Se = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP}, \quad Acc = \frac{TP + TN}{N},$$
$$MCC = \frac{TP/N - S \times P}{\sqrt{P \times S \times (1 - S) \times (1 - P)}}, \quad F1 = \frac{2 \times Pr \times Re}{Pr + Re}, \tag{6}$$

where $N = TN + TP + FN + FP$, $S = (TP + FN)/N$, $P = (TP + FP)/N$ and $Pr = TP/(TP + FP)$.

The *MCC* and *F1-score* are often used in the performance analysis of segmentation result on class imbalance datasets. This is true in case of retinal images, where only a small portion of pixels belongs to vessels (only around 9–14% vessel pixels), while the others are counted as background pixels. The *MCC* returns a value between −1 and +1, with +1 indicating a perfect prediction, and −1 a completely incorrect prediction. Finally, the receiver operating characteristic (ROC) curves are computed with the true positive fractions (*Se*) versus false positive fractions ($1 - Sp$) by varying the threshold on the probability map. The *AUC* is calculated to quantify the performance of the segmentation algorithm where, the *AUC* value of 1 indicates a perfect segmentation.

4.3. Parameters setting

The proposed retinal vessel segmentation approach mainly depends on the three main parameters: p_0 – (the size of input visual glimpse patch x), l – (the number of levels of visual glimpse) and K – (the number of dictionary atoms (learned filter size)). These parameters play a critical role in obtaining the optimal performance of retinal vessel segmentation approach. The details are presented next.

In our experiments, we choose $K = 100$ filters by randomly sampling 100000 patches from the training sets of three datasets namely – DRIVE, STARE and CHASE_DB1. Fig. 7(a) shows the effect of varying the input patch size (p_0), without visual attention mechanism ($l = 0$), on the accuracy of segmentation performance. It is observed that the input patch size of $(9 \times 9)px$ gave the best results across all three datasets of varying image resolutions. This

Table 1
Performance analysis for varying levels (l) of visual glimpse. *Note:* the input patch with $l = 0$ is the one *without* having any visual attention mechanism.

Datasets	l	*Se*	*Sp*	*Acc*	*AUC*	*MCC*	*F1*
DRIVE	0	0.7665	0.9088	0.8602	0.8990	0.6444	0.6737
	1	0.8387	0.9296	0.9011	0.9405	0.6886	0.7077
	2	**0.8644**	**0.9667**	**0.9589**	**0.9701**	**0.7421**	**0.7607**
STARE	0	0.7894	0.8832	0.8510	0.9010	0.6391	0.6442
	1	0.8134	0.9178	0.8945	0.9372	0.6732	0.6818
	2	**0.8325**	**0.9746**	**0.9502**	**0.9670**	**0.7398**	**0.7698**
CHASE_DB1	0	0.7576	0.8815	0.8646	0.9108	0.6485	0.6695
	1	0.7823	0.9297	0.9173	0.9303	0.6653	0.7012
	2	**0.8297**	**0.9663**	**0.9474**	**0.9591**	**0.6927**	**0.7189**
IOSTAR	0	0.7640	0.9074	0.8543	0.9094	0.6536	0.6831
	1	0.7853	0.9393	0.9236	0.9453	0.6739	0.7032
	2	**0.8269**	**0.9669**	**0.9564**	**0.9663**	**0.7057**	**0.7354**
RC-SLO	0	0.7756	0.9176	0.9056	0.9165	0.6598	0.6854
	1	0.8149	0.9432	0.9304	0.9488	0.6812	0.7019
	2	**0.8488**	**0.9666**	**0.9581**	**0.9678**	**0.7029**	**0.7200**

is because the selected patch size is able to capture the varying nature of vessel caliber (such as thin and thick vessels) across multiple scales and multiple orientations. Further, the chosen patch size is found to be strongly dependent on average vessel caliber across each dataset. It is experimentally found that, the mean vessel caliber varies from 3.4 ± 1.6 *px* for the DRIVE, 4.4 ± 2.6 *px* for the STARE, 5.4 ± 3.6 *px* for the CHASE_DB1, 6.3 ± 2.5 *px* for the IOSTAR and 5.2 ± 2.5 *px* for the RC-SLO datasets [17]. Hence, the chosen patch size is well within the range of average vessel caliber across all five datasets and thus exhibits an improved performance for a fixed patch size (p_0), even with varying image resolutions across datasets.

Next, to assess the influence of visual attention mechanism on filter learning, we adopted *two* levels (l) of visual glimpse – G_1, G_2 where, $l = 2$; on a fixed input patch (G_0), which is of size $p_0 = (9 \times 9)$ *px*. Table 1 depicts the vessel segmentation performance for varying levels of visual glimpse across all five datasets. It is observed that there is a significant boost in the segmentation performance with an average *Se* > 10/5/7/6/7%, *Sp* > 6/9/8/6/5%, *Acc* > 10/10/8/10/5%, *AUC* > 7/7/5/6/5%, *MCC* > 10/10/4/5/4%, and *F*1 > 9/13/5/5/4% across DRIVE/STARE/CHASE_DB1/IOSTAR/RC-SLO datasets, respectively. This is mainly because the higher level of visual glimpses learns to selectively pay attention to the most relevant part of a scene (such as vessel structures) in a given local patch. Thus providing a discriminative capability in predicting the true class label (*TP's*), even in the presence of noisy and cluttered background. The ability of our approach to correctly identify the true positives is clearly reflected in the values of *Se*, *Acc*, *MCC*, *F*1 with an average increase in 4% across all five datasets, when compared with the performance obtained without any visual attention mechanism ($l = 0$ in Table 1). Further, there is a boost in the values of all performance metrics of ≈4% at every level (l) of the visual glimpse, clearly indicating an influence of visual attention mechanism on vessel segmentation performance. Thus in all our experiments, we choose input patch size of (9×9) *px* with $l = 2$ levels of visual glimpse for evaluating the segmentation performance against the state-of-the-art methods.

We also evaluated the segmentation performance by varying the number of dictionary atoms ranging from $K = \{100, 200, 300, 400, 500\}$ as shown in Fig. 7(b). For this experiment, we randomly sampled 100000 patches of size (9×9) *px* with $l = 2$ glimpses for learning the filter size K. It is shown that our segmentation approach achieves an average *Acc* > 0.95, when chosen filter size $K = 500$. But this improved segmentation accuracy comes at the price of a slower prediction time, which is inappropriate for a standard clinical setting. In our experiments, we further observed that there is a very minute improvement in the *Acc* of the segmentation performance between the filter size of $K = 100$–500. Hence, we empirically set the size of filter $K = 100$ throughout our experiments – to offer a good compromise between training/testing speed and segmentation performance.

In addition, there are few other parameters such as ϵ_{ZCA} which is set at 0.1, that provides a good balance between filter sharpness and noise amplification. The sparsity parameter λ is chosen to be 1, which yield best results on all datasets at fairly faster convergence rate. We trained a RF classifier of $N_T = 100$ trees using a $2 \times K = 200$ dimension feature vector. Further, increasing the number of trees have shown to result in better segmentation performance at a cost of increased time complexity [47]. Hence, we empirically set RF of 100 trees for classification across all five datasets.

4.4. Visualization of discovered features

In this section, we provide an in-depth experimental validation of the central hypothesis of this paper by visualizing the effect of visual attention mechanism on unsupervised filter learning. We visualize these learned projections using a fast implementation of t-distributed Stochastic Neighbor Embedding (t-SNE) [59] using default set of parameters. We adopt this technique due to its extensive ability to preserve neighborhoods and clusters in lower dimensional subspace, compared to other dimensionality reduction techniques such as PCA, Isomap, etc.

For visualizing the learned features, we select a random subset of 100000 observations from test sets of all five datasets, due to the extensive computational complexity of t-SNE. The features are extracted from an input visual glimpse patch of size (9×9) *px* with $l = 2$ glimpse to mimic the visual attention mechanism, as described in Section 3.2. Fig. 8 depicts the projection of a subset of observations for varying levels of visual glimpses. It is observed that there is a clear visual separation between the vessel and background classes obtained with $l = 2$ visual glimpse, as shown in Fig. 8(c). This infers that the higher levels of visual glimpse generally captures larger contextual information and thereby providing more discriminative features for subsequent class-assignment. Whereas, for the lower levels of visual glimpse (see Fig. 8(a) and (b)), class separation is clearly inferior, indicating that there is a strong overlap between the vessel and background classes, due to very limited neighborhood information encountered during filter learning process.

To further substantiate the hypothesis, we visualized the learned features of sample ROI's for various challenging cases such as segmentation in the presence of pathology, segmenting thin vessels and complex crossover structures, as shown in Fig. 9. The learned features via visual attention mechanism have shown to be

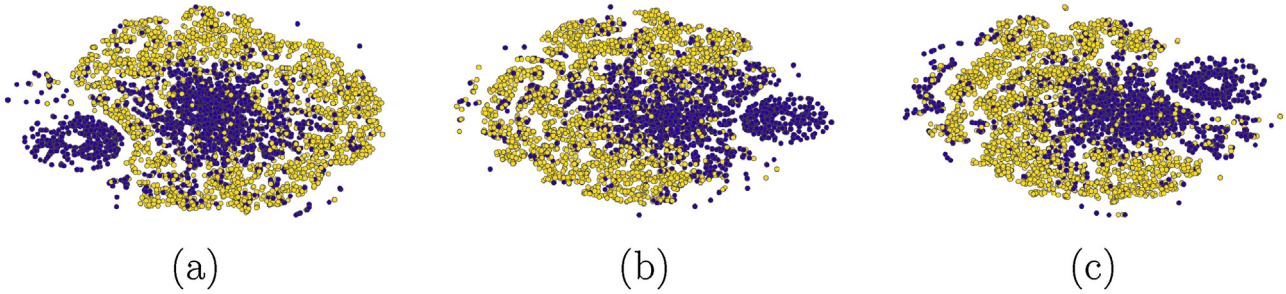

Fig. 8. Feature embedding visualization via (t-SNE) for subset of observations obtained from test sets of all five datasets, with varying levels (l) of visual glimpse: (a) visual glimpse with $l = 0$, (b) visual glimpse with $l = 1$, (c) visual glimpse with $l = 2$. The violet and yellow colors correspond to the vessel and background, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

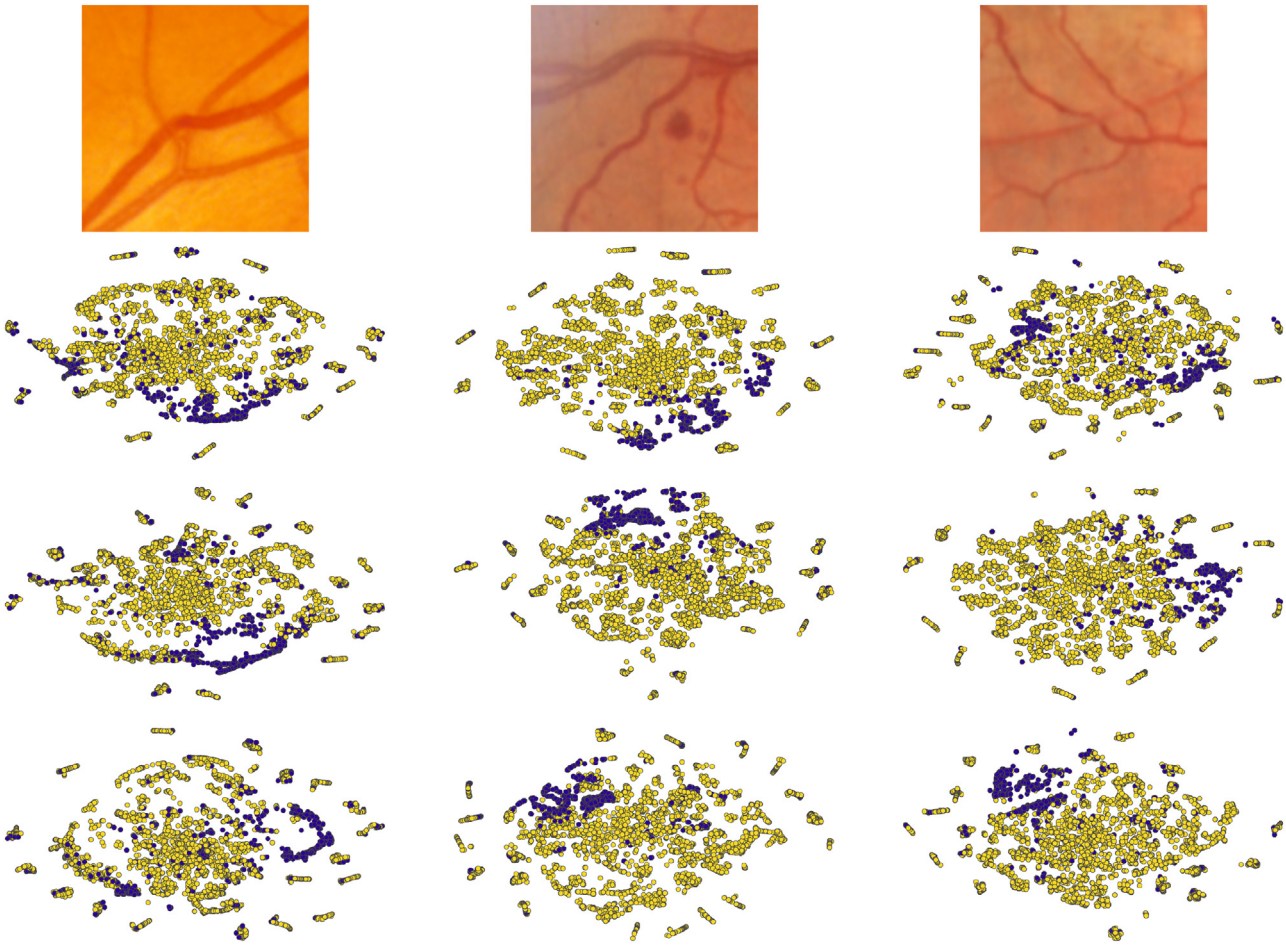

Fig. 9. Feature embedding visualization via (t-SNE) of sample ROI's for varying levels (l) of visual glimpse. First row: original image, second row: visual glimpse with $l = 0$, third row: visual glimpse with $l = 1$, fourth row: visual glimpse with $l = 2$. The violet and yellow colors correspond to the vessel and background, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

superior in exhibiting a discriminative class separation between vessel and background (see second and third row) compared to the one without attention mechanism (see the first row). Our approach is also shown to be robust to pathology and thin vessels structures, where higher levels of visual glimpse often aid in predicting the true class label even in the presence of similar looking cluttered objects (see second and third column). This is consistent with the hypothesis which we claimed in Section 3.2, that the region containing thin vessels and pathology often exhibits similar visual appearance, there by posing a significant challenge for accurate segmentation of retinal vessels. Hence, visual attention mechanism is shown to capture the most relevant structure (salient region), as well as the neighbourhood information (context), across multiple scales in a given local patch. Thus, the visual attention mechanism aims to drive the unsupervised filter learning process to generate a more compact and distinctive feature set for the subsequent classification.

4.5. Vessel segmentation results

The sample qualitative results of the proposed approach are shown in Fig. 10 on five datasets namely DRIVE, STARE, CHASE_DB1, IOSTAR and RC-SLO. It is observed that the proposed approach segments most of the thin vessel structures, preserves vessel con-

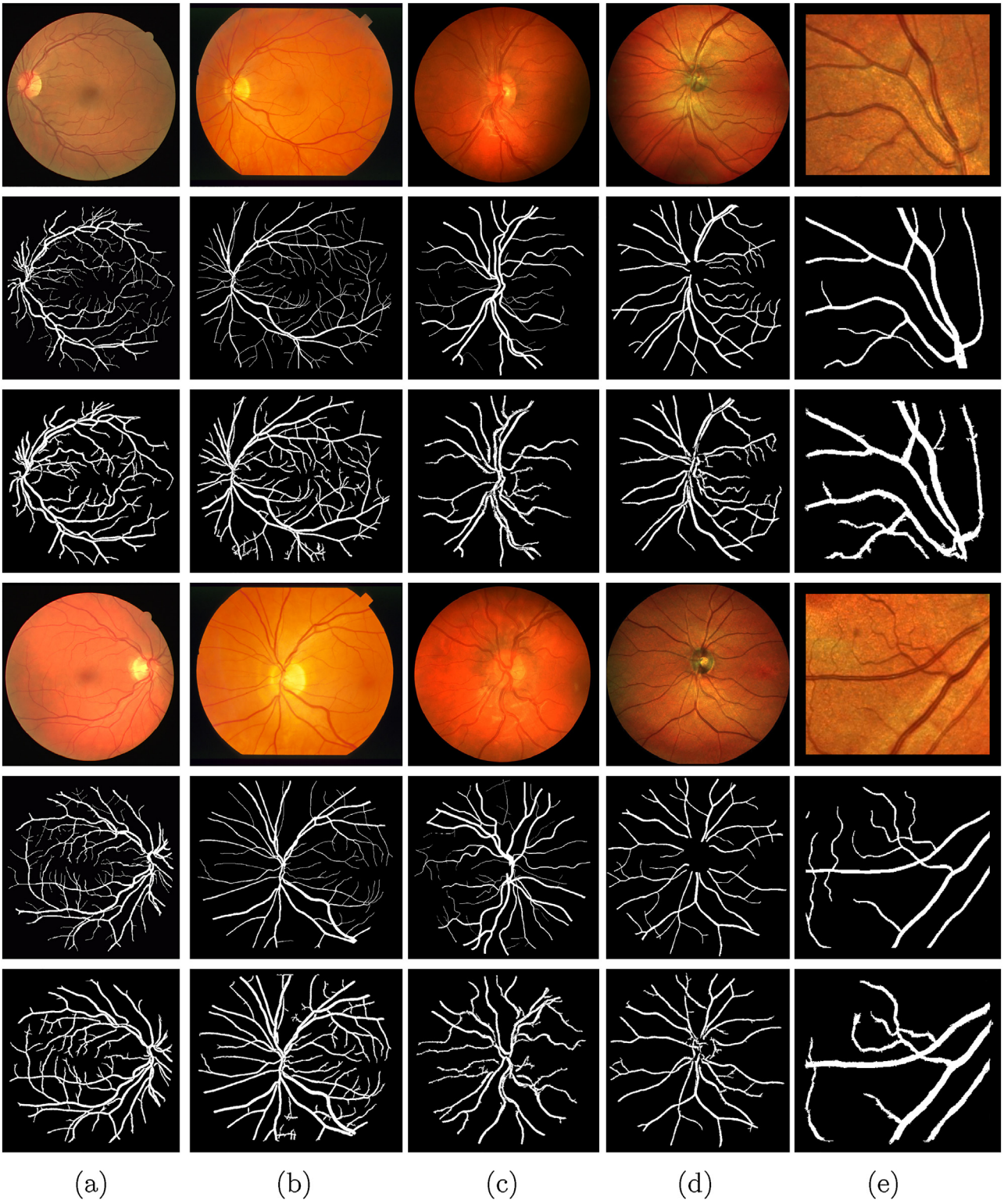

Fig. 10. Segmentation results obtained on DRIVE (a), STARE (b), CHASE_DB1 (c), IOSTAR (d) and RC-SLO (e) datasets. First and fourth row: original images. Second and fifth row: manual segmentation results. Third and sixth row: segmentation results of our proposed approach.

nectivity at junction locations and also shown to be robust even in the presence of illumination artefacts (see third and sixth row). The improved performance is largely due to the discriminative capability of unsupervised filter learning, which is combined with the visual attention mechanism to leverage both the salient and contextual information from multiple scales. This hybrid information facilitate to accurately identify vessel pixels from the clutter background (such as lesions), differentiating thin and thick vessels and varying crossover vessel structures.

Tables 2 and 3 depicts the performance of the proposed approach with the existing state-of-the-art supervised and unsupervised approaches. Our method achieves an average *Se* of greater than 7% compared with the existing approaches validated on DRIVE, CHASE_DB1, IOSTAR and RC-SLO datasets, and a higher *Se* value than

Table 2
Performance comparison on the DRIVE, STARE and CHASE_DB1 datasets.

	Methods	Year	DRIVE				STARE				CHASE_DB1			
			Se	*Sp*	*Acc*	*AUC*	*Se*	*Sp*	*Acc*	*AUC*	*Se*	*Sp*	*Acc*	*AUC*
	2nd human observer	–	0.7760	0.9724	0.9472	–	0.8952	0.9384	0.9349	–	0.8105	0.9711	0.9545	–
Unsupervised methods	Mendonca [60]	2006	0.7344	0.9764	0.9452	–	0.6996	0.9730	0.9440	–	–	–	–	–
	Martinez-Perez [61]	2007	0.7246	0.9655	0.9344	–	0.7506	0.9569	0.9410	–	–	–	–	–
	Al-Diri [39]	2009	0.7282	0.9551	–	–	0.7521	0.9681	–	–	–	–	–	–
	Lam [38]	2010	–	–	0.9472	0.9614	–	–	0.9567	0.9739	–	–	–	–
	Zhang [33]	2010	0.7120	0.9724	0.9382	–	0.7177	0.9753	0.9484	–	–	–	–	–
	You [62]	2011	0.7410	0.9751	0.9434	–	0.7260	0.9756	0.9497	–	–	–	–	–
	Fraz [63]	2012	0.7152	0.9759	0.9430	–	0.7311	0.9680	0.9442	–	–	–	–	–
	Roychowdhury [37]	2015	0.7395	0.9782	0.9494	0.9672	0.7317	0.9842[a]	0.9560	0.9673	0.7615	0.9575	0.9467[a]	0.9623[a]
	Azzopardi [20]	2015	0.7655	0.9704	0.9442	0.9614	0.7716	0.9701	0.9497	0.9563	0.7585	0.9587	0.9387	0.9487
	Yin [32]	2015	0.7246	0.9790	0.9403	–	0.8541[a]	0.9419	0.9325	–	–	–	–	–
	Zhao [40]	2015	0.7420	0.9820	0.9540	0.8620	0.7800	0.9780	0.9560	0.8740				
	Zhang [16]	2016	0.7743	0.9725	0.9476	0.9636	0.7791	0.9758	0.9554	0.9748	0.7626[a]	0.9661[a]	0.9452	0.9606
	Kovacs [31]	2016	0.7450	0.9793[a]	0.9494	0.9722[a]	0.8034	0.9786	0.9610[a]	0.9836[a]	–	–	–	–
	Zhao [41]	2017	0.7820[a]	0.9790	0.9570[a]	0.8860	0.7890	0.9780	0.9560	0.8850	–	–	–	–
Supervised methods	Niemeijer [13]	2004	–	–	0.9416	0.9294	–	–	–	–	–	–	–	–
	Staal [14]	2004	–	–	0.9441	0.9520	–	–	0.9516	0.9614	–	–	–	–
	Soares [12]	2006	0.7332	0.9782	0.9466	0.9614	0.7207	0.9747	0.9480	0.9671	–	–	–	–
	Ricci [15]	2007	–	–	0.9595	0.9558	–	–	0.9584	0.9602	–	–	–	–
	Lupascu [64]	2010	0.7200	–	0.9597[b]	0.9561	–	–	–	–	–	–	–	–
	Marin [18]	2011	0.7067	0.9801	0.9452	0.9588	0.6944	0.9819	0.9526	0.9769	–	–	–	–
	Fraz [42]	2012	0.7406	0.9807	0.9480	0.9747[b]	0.7548	0.9763	0.9534	0.9768	0.7224	0.9711	0.9469	0.9712
	Orlando [43]	2016	0.7897	0.9684	–	–	0.7680	0.9738	–	–	0.7277	0.9712	–	–
	Li [44]	2016	0.7569	0.9816[b]	0.9527	0.9738	0.7726	0.9844	0.9628	0.9879[b]	0.7507	0.9793	0.9581	0.9716
	Liskowski [10]	2016	0.7750	0.9795	0.9518	0.9747[b]	0.7766	0.9854[b]	0.9638[b]	0.9868	0.7544	0.9846[b]	0.9610[b]	0.9801[b]
	Zhang [17]	2016	0.7861	0.9712	0.9466	0.9703	0.7882	0.9729	0.9547	0.9740	0.7644	0.9716	0.9502	0.9706
	Our method	2017	**0.8644**[b]	**0.9667**	**0.9589**	**0.9701**	**0.8325**[b]	**0.9746**	**0.9502**	**0.9670**	**0.8297**[b]	**0.9663**	**0.9474**	**0.9591**

[a] Best values among unsupervised methods.
[b] Best values among supervised methods.

Table 3
Performance comparison on the IOSTAR and RC-SLO datasets.

Methods	Datasets	Year	*Se*	*Sp*	*Acc*	*AUC*	*MCC*
Zhang [16]	IOSTAR	2016	0.7545	**0.9740**	0.9514	0.9615	**0.7318**
Our method	IOSTAR	2017	**0.8269**	0.9669	**0.9564**	**0.9663**	0.7057
Zhang [16]	RC-SLO	2016	0.7787	**0.9710**	0.9512	0.9626	**0.7327**
Our method	RC-SLO	2017	**0.8488**	0.9666	**0.9581**	**0.9678**	0.7029

Table 4
Comparison of the proposed approach in terms of average *MCC* and *F*1-score.

Methods	Year	DRIVE		STARE		CHASE_DB1	
		MCC	*F*1	*MCC*	*F*1	*MCC*	*F*1
2nd human observer	–	0.7601	0.7881	0.7225	0.7401	0.7475	0.7686
Azzopardi [20]	2015	0.7475	–	0.7335	–	0.6802	–
Orlando [43]	2016	0.7556	0.7857	0.7417	0.7644	0.7046	0.7332
Zhang [17]	2016	**0.7673**	**0.7953**	**0.7608**	**0.7815**	**0.7324**	**0.7581**
Our method	2017	0.7421	0.7607	0.7398	0.7698	0.6927	0.7189

all the supervised approaches on STARE dataset. This increased *Se* is mainly because of the discriminative feature set, which takes into account of large contextual information in predicting the true class label from similar-looking background structures. For the DRIVE dataset, we obtain *Acc*/*AUC* of 0.9589/0.9701, which is significantly higher than all the unsupervised approaches. For the CHASE_DB1 dataset, we obtain *Acc* of 0.9474, which is higher than all the unsupervised approaches. We also assess the performance of the proposed approach on SLO images (see Table 3), which includes challenging cases like thin vessels, complex crossovers and closely spaced parallel vessels. Our method obtains slightly noticeable improvements in *Acc*/*AUC* of 0.9564/0.9663 and 0.9581/0.9678 on IOSTAR and RC-SLO dataset respectively, compared with Zhang et al. [16]. In Table 4, we also compare our method with the existing approaches in terms of average *MCC* and *F*1-score across DRIVE, STARE and CHASE_DB1 datasets. Our method has shown to perform better in terms of *MCC* and the *F*1-score, indicating that the proposed approach is also robust to class imbalance (fewer vessel pixels when compared to pixels belonging to background) in the dataset.

For the sake of fair comparison with the existing state-of-the-art methods, we extract sensitivity values from the ROC curve for a fixed specificity reported in two best methods among supervised and unsupervised approaches, as shown in Table 5. Compared with the unsupervised approaches, our method achieved best *Se* value for a fixed *Sp* reported in [41,31] on DRIVE, [32,31] on STARE and [19,16] on CHASE_DB1, respectively. Whereas, among the supervised ones, we achieve a higher *Se* in comparison with [10,44] on DRIVE and STARE, inferring that the proposed approach is also competent with other deep learning based approaches. The methods in [16,17,31] strongly rely on carefully designed hand-crafted features (such as Gabor and Wavelet-based filtering techniques) compared with the proposed approach that learns the feature representation directly from the raw data, without the explicit need for complex domain knowledge.

4.6. Cross-validation

We evaluate the performance by training and testing on a set of images from different databases, to check the robustness and practicability of the proposed approach. Table 6 shows the cross-validation performance between datasets with the existing state-of-the-art methods. We obtain slightly decrease in the *Acc* value of 0.9489 by training on STARE and testing on DRIVE dataset, when compared to training and testing on STARE alone. This is because the STARE contains pathological signs in almost 10 of its images, while DRIVE contains mostly healthier ones. Further, we obtain *Acc*/*AUC* values of 0.9640/0.9702 by training on DRIVE and tested on STARE, which is marginally greater than *Acc*/*AUC* of 0.9589/0.9701 obtained by training and testing on DRIVE alone. A decreasing trend is also observed in *Acc*/*AUC* values when trained on STARE and tested on CHASE_DB1, compared to training and testing on STARE alone. Overall, there is a clear indication that the learned feature set does not strongly rely on the specific training set, and are highly adaptable to varying image resolutions, robust to pathological signs and various other imaging artefacts present in the data.

4.7. Performance analysis on challenging cases

Segmenting the retinal vessels from color fundus image is challenging due to the presence of strong CVR, low contrast thin vessel structures, close parallel and highly curved vessels, close bifurcation and crossover regions and pathological lesions such as exudates, haemorrhages and microaneurysms.

Fig. 11 shows the qualitative performance of our approach on sample ROI's containing various challenging cases. Our method

Table 5
Comparison of sensitivity and specificity values.

	DRIVE			STARE			CHASE_DB1		
	Methods	*Se*	*Sp*[a]	Methods	*Se*	*Sp*[a]	Methods	*Se*	*Sp*[a]
Unsupervised	Zhao [41]	0.7820	0.9790	Yin [32]	0.8541	0.9419	Roychowdhury [19]	0.7615	0.9575
	Kovacs [31]	0.7450	0.9793	Kovacs [31]	0.8034	0.9786	Zhang [16]	0.7626	0.9661
	Our method	**0.7939**	**0.9790**	Our method	**0.8660**	**0.9419**	Our method	**0.7862**	**0.9575**
		0.7935	**0.9793**		**0.8060**	**0.9786**		**0.7670**	**0.9661**
Supervised	Liskowski [10]	0.7750	0.9795	Liskowski [10]	**0.7766**	**0.9854**	Liskowski [10]	**0.7544**	**0.9846**
	Li [44]	0.7569	0.9816	Li [44]	0.7726	0.9844	Zhang [17]	**0.7644**	**0.9716**
	Our method	**0.7930**	**0.9795**	Our method	0.7736	0.9854	Our method	0.7459	0.9846
		0.7647	**0.9816**		**0.7773**	**0.9844**		0.7512	0.9716

[a] Specificity cut-off: specific point on the ROC curve for extracting sensitivity values.

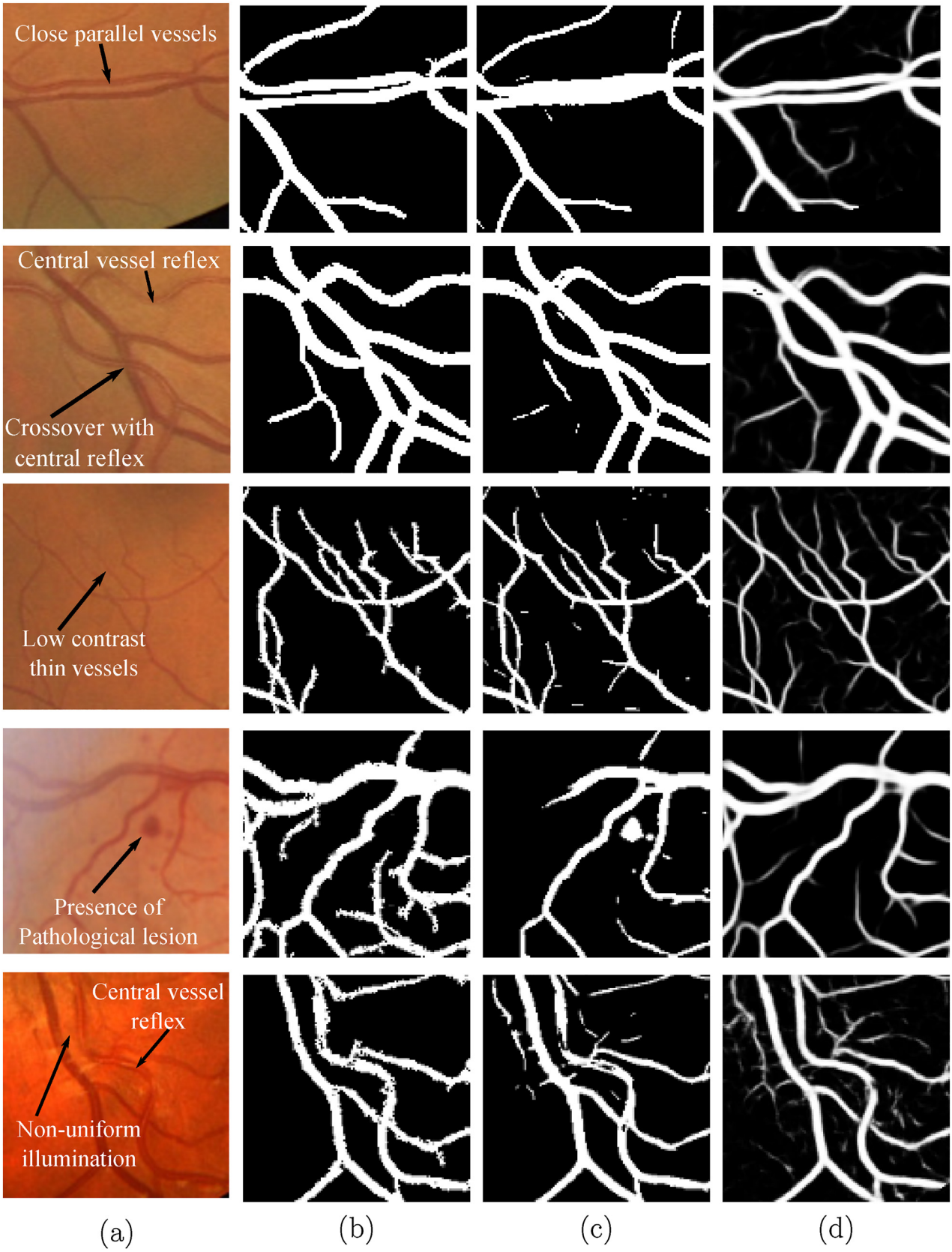

Fig. 11. Qualitative assessment of vessel segmentation results on challenging cases: (a) original image, (b) our approach, (c) Orlando et al. [43], (d) Liskowski et al. [10].

produces far less false positives, more clean segmentation compared to the most recent state-of-the-art methods [43,10]. The approach is also shown to preserve various complex vascular structures (first and second row) and segments thin vessels accurately even at very low contrast regions (third row). This robustness is mainly due to the learnt filters that leverage the visual atten-

Table 6
Performance comparison with cross-validation between databases.

	Methods	*Acc*	*AUC*
Test images from: DRIVE			
Model trained on	Soares [12]	0.9397	–
STARE	Ricci [15]	0.9266	–
	Marin [18]	0.9448	–
	Fraz [42]	0.9456	**0.9697**
	Li [44]	0.9486	0.9677
	Liskowski [10]	0.9416	0.9605
	Zhang [17]	0.9447	0.9593
	Our method	**0.9489**	0.9676
Model trained on CHASE_DB1	Li [44]	**0.9484**	**0.9605**
	Our method	0.9301	0.9476
Test images from: STARE			
Model trained on	Soares [12]	0.9327	–
DRIVE	Ricci [15]	0.9464	–
	Marin [18]	0.9528	–
	Fraz [42]	0.9493	0.9660
	Li [44]	0.9545	0.9671
	Liskowski [10]	0.9505	0.9595
	Zhang [17]	0.9488	0.9676
	Our method	**0.9640**	**0.9702**
Model trained on CHASE_DB1	Li [44]	**0.9536**	**0.9620**
	Our method	0.9334	0.9464
Test images from CHASE_DB1			
Model trained on DRIVE	Li [44]	0.9429	**0.9628**
	Our method	**0.9476**	0.9621
Model trained on	Fraz [42]	0.9415	**0.9565**
STARE	Li [44]	0.9417	0.9553
	Zhang [17]	**0.9458**	0.9538
	Our method	0.9403	0.9521

Table 7
Computational complexity vs. accuracy (*Acc*) of the state-of-the-art methods on DRIVE/STARE datasets, respectively. Note: the segmentation time is recorded for processing one DRIVE/STARE image.

Methods	Year	Running time	*Acc*
Soares [12]	2006	3 min	0.9466/0/0.9480
Marin [18]	2011	1.5 min	0.9452/0.9526
Fraz [42]	2012	2 min	0.9480/0.9534
Roychowdhury [19]	2015	3.11/6.7 s	0.9490/0.9560
Roychowdhury [37]	2015	2.45/3.95 s	0.9520/0.9510
Li [44]	2016	1.2 min	0.9527/**0.9628**
Zhang [17]	2016	23.4 s	0.9466/0.9547
Our approach	2018	**45 s**	**0.9589**/0.9502

tion mechanism – by taking into account of multi-scale contextual information, to produce a set of more discriminative features. Compared to the conventional methods such as [65,19], the proposed approach is capable of segmenting both large and small vessel structures within a single framework, without the explicit need for complex techniques to tackle them separately. Further, the method is also shown to be robust to pathological lesions such as microaneurysms (see the fourth row) and also the CVR phenomenon (see the fifth row) compared to [43]. The major advantage of our method is its ability to handle most of the challenging cases, without the explicit need for any complex pre/post-processing operations. Hence, our approach is more reliable and suitable for developing CAD tool for the automated analysis of retinal vasculature.

4.8. Computational cost

We report the running time for vessel segmentation approach (see Table 7) starting from visual attention modelling to the prediction of class labels using RF classifier, which takes approximately 45 s per image using an unoptimized MATLAB code, on a machine equipped with 2.2 GHz, Intel i3-2330 processor and 8 GB of RAM. The main computational bottleneck is the encoder part of the unsupervised filter learning approach, which takes roughly 20 s per image for solving L_1 optimization in the sparse coding step. This can be solved more efficiently using a very recently proposed soft thresholding scheme [66], which jointly learns both the dictionary and a linear classifier in a single optimization step. This could be a very interesting future work in this direction. It is observed that based on the current parameter setting (see Section 4.3), our proposed approach offers a considerable prediction time suitable for standard clinical setting, without compromising on segmentation accuracy across all five datasets.

5. Discussion

In this paper, we presented a visual attention guided unsupervised feature learning framework for the task of segmenting retinal vessels from color fundus images. Segmenting retinal vessels is challenging due to multi-scale nature of varying vessel caliber, complex crossover patterns, the close proximity of pathological lesions and non-uniform illumination inherited during image acquisition process. These challenges often pose a great difficulty in obtaining a discriminative feature set for the accurate classification of vessel pixels. Most earlier state-of-the-art approaches [16,17,19,20,31] have focused extensively on carefully designing hand-crafted features, which are generally based on complex domain knowledge, and requires an enormous amount of manual tuning of parameters to achieve optimal segmentation performance. We address these limitations, by leveraging the idea of visual attention mechanism into conventional unsupervised feature learning framework. Our idea aims to produce a more versatile segmentation approach through the design of automatic learned discriminative feature set, using a set of only unlabelled image samples. Therefore, the proposed approach is capable of capturing the geometrical richness of retinal vasculature – occurring at multiple-scales and multiple-orientations, without the need for any expensive manual annotations and complex domain expertise.

Our proposed approach has several appealing properties. First, the visual attention mechanism is capable of capturing the rich contextual information, which has the ability to ignore the clutter present in a local neighbourhood, by focusing only on the pixel of interest. Second, the integration of visual attention mechanism into unsupervised filter learning encourages the intra-class similarities to be small (between the vessel pixels), and emphasizes the inter-class differences to be large (between the vessel and background pixels). Third, our VA-UFL approach requires only three main parameters: p_0, l and K during its training, which is generally chosen by cross-validation. Further, the chosen parameters should be fixed only once, which is independent of varying image resolutions and requires practically no manual tuning for various kinds of datasets.

Extensive experiments are carried out on five publicly available retinal datasets containing various challenging cases to illustrate the broad applicability of the VA-UFL method. It is shown that our approach is highly competent and outperforms state-of-the-art methods by a value of $Se > 0.82$ across all five datasets as shown in Tables 2 and 3. In addition, the proposed approach consistently performed well in terms of global performance metrics such as *MCC* and *F*1-score (see Table 4), demonstrating that the method is also robust to class imbalance present in the data (fewer vessel pixels compared to background). The significant boost in *Se* value is mainly due to the discriminative power of learned filter bank which leverage the strength of visual attention mechanism, to accurately identify true positives from similar-looking noisy background.

The proposed approach is also shown to perform well on various challenging cases such as – segmenting thin vessels from low contrast regions, segmentation in the presence of pathological lesions and strong CVR as shown in Fig. 11. The improved segmentation

performance is mainly because of the visual attention mechanism that takes into account of complex neighbourhood information, to accurately segment vessels from its complex noisy background. This is also true in the case of projection of learned features, which is visualized via t-SNE in low-dimensional subspace as shown in Fig. 8. It is observed that there is a significant class separation between the vessel and background pixels with higher levels of visual glimpse, indicating that the visual glimpse is inherently capturing the most relevant structures in addition to the contextual information, to learn a set of most discriminative basis.

Although, the proposed approach successfully segments complex retinal vascular structures, we have observed some misclassification in the region of bright and large exudates. These false positives can further be removed by employing a simple pre/post-processing steps (such as image in painting) or by integrating vessel shape prior information into the filter learning approach. Besides that, the segmentation of pathological images can be even further improved by incorporating a more hybrid contextual models such as auto-context [67]. The auto-context iteratively utilizes the posterior distribution of labels along with the image features, to obtain a more compact and discriminative features suitable for highly overlapping classes. Furthermore, the discriminative capability of the UFL approach can be further improved by associating each class label information with the dictionary atom during the filtering process. This hybrid information effectively reduces both the reconstruction and classification error using a unified objective function.

As a future work, the proposed scheme can be used as an initial step in identifying vessel junctions [68], estimating vessel tortuosity, in painting vessels for lesion detection [69–71] and classifying artery-vein from retinal images. The method can be further improved by training an ensemble learning based classification model combining both hand-crafted and learned features as proposed in [72–80]. Due to the recent advancements in deep learning methods, it is worth investigating various newly introduced deep learning architectures [81,82] for the task of retinal vessel segmentation.

6. Conclusions

In this work, we have proposed a visual attention guided unsupervised feature learning approach for the purpose of vessel segmentation in retinal fundus images. The proposed approach inherits the advantages of visual attention mechanism and fully utilizes the potential of unsupervised feature learning for representing most discriminative features for classification. This allows us to explore the space of both selection mechanism and multi-scale contextual information under a single framework, without the need for any complex feature learning modules.

Extensive experimental analysis on five publicly available retinal datasets demonstrates the superior performance of our approach with respect to other state-of-the-art vessel segmentation methods. The effectiveness of the proposed approach is evident by the significant improvement in the value of sensitivity compared to all previously published methods and even outperforms second human observer. Further, the method is also shown to be robust to several challenging image structures such as central vessel reflex, complex crossover patterns, closely parallel and highly curved vessels, thin vessels and performs fairly well on pathological images. In conclusion, the excellent performance of our approach demonstrates the applicability for real-time computer-aided diagnosis and large-scale retinal disease screening programs.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgement

The authors gratefully acknowledge Prof. Jayanthi Sivaswamy (IIIT-Hyderabad), Samrudhdhi Rangrej and Karthik Gopinath for their valuable and constructive suggestions throughout this research work.

References

[1] M.D. Abramoff, M.K. Garvin, M. Sonka, Retinal imaging and image analysis, IEEE Rev. Biomed. Eng. 3 (2010) 169–208.
[2] A. London, I. Benhar, M. Schwartz, The retina as a window to the brain from eye research to CNS disorders, Nat. Rev. Neurol. 9 (2013) 44–53.
[3] C.Y. Cheung, M.K. Ikram, C. Chen, T.Y. Wong, Imaging retina to study dementia and stroke, Prog. Retin. Eye Res. 57 (2017) 89–107.
[4] M.L. Baker, J.J. Wang, G. Liew, P.J. Hand, D.A. De Silva, R.I. Lindley, P. Mitchell, M.-C. Wong, E. Rochtchina, T.Y. Wong, J.M. Wardlaw, G.J.a. Hankey, Differential associations of cortical and subcortical cerebral atrophy with retinal vascular signs in patients with acute stroke, Stroke 41 (10) (2010) 2143–2150.
[5] D. Rosenbaum, N. Kachenoura, E. Koch, M. Paques, P. Cluzel, A. Redheuil, X. Girerd, Relationships between retinal arteriole anatomy and aortic geometry and function and peripheral resistance in hypertensives, Hypertens. Res. 39 (7) (2016) 536–542.
[6] T.Y. Wong, R. Klein, D.J. Couper, L.S. Cooper, E. Shahar, L.D. Hubbard, M.R. Wofford, A.R. Sharrett, Retinal microvascular abnormalities and incident stroke: the atherosclerosis risk in communities study, Lancet 358 (9288) (2001) 1134–1140.
[7] S.M. Heringa, W.H. Bouvy, E. van den Berg, A.C. Moll, L.J. Kappelle, G.J. Biessels, Associations between retinal microvascular changes and dementia, cognitive functioning, and brain imaging abnormalities: a systematic review, J. Cereb. Blood Flow Metab. 33 (7) (2013) 983–995.
[8] 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.
[9] H. Fu, Y. Xu, S. Lin, D.W.K. Wong, J. Liu, DeepVessel: retinal vessel segmentation via deep learning and conditional random field, International Conference on Medical Image Computing and Computer-Assisted Intervention (2016) 132–139.
[10] P. Liskowski, K. Krawiec, Segmenting retinal blood vessels with deep neural networks, IEEE Trans. Med. Imaging 35 (11) (2016) 2369–2380.
[11] L. Srinidhi, P. Aparna, J. Rajan, Recent advancements in retinal vessel segmentation, J. Med. Syst. 41 (4) (2017) 70.
[12] J.V.B. Soares, J.J.G. 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.
[13] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff, et al., Comparative study of retinal vessel segmentation methods on a new publicly available database, SPIE Medical Imaging, vol. 5370 (2004) 648–656.
[14] J. Staal, M.D. Abramoff, 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.
[15] E. Ricci, R. Perfetti, Retinal blood vessel segmentation using line operators and support vector classification, IEEE Trans. Med. Imaging 26 (10) (2007) 1357–1365.
[16] J. Zhang, B. Dashtbozorg, E. Bekkers, J.P.W. 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.
[17] J. Zhang, Y. Chen, E. Bekkers, M. Wang, B. Dashtbozorg, B.M. ter Haar Romeny, Retinal vessel delineation using a brain-inspired wavelet transform and random forest, Pattern Recognit. 69 (2017) 107–123.
[18] D. Marin, A. Aquino, M.E. Gegundez-Arias, J.M. Bravo, A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features, IEEE Trans. Med. Imaging 30 (1) (2011) 146–158.
[19] S. Roychowdhury, D.D. Koozekanani, K.K. Parhi, Blood vessel segmentation of fundus images by major vessel extraction and subimage classification, IEEE J. Biomed. Health Inform. 19 (3) (2015) 1118–1128.
[20] 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.
[21] N. Strisciuglio, G. Azzopardi, M. Vento, N. Petkov, Supervised vessel delineation in retinal fundus images with the automatic selection of B-COSFIRE filters, Mach. Vis. Appl. 27 (8) (2016) 1137–1149.
[22] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, L. Van Gool, Deep retinal image understanding, International Conference on Medical Image Computing and Computer-Assisted Intervention (2016) 140–148.

[23] A. Wu, Z. Xu, M. Gao, M. Buty, D.J. Mollura, Deep vessel tracking: a generalized probabilistic approach via deep learning, 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) (2016) 1363–1367.
[24] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, Advances in Neural Information Processing Systems 19 (2007) 801–808.
[25] M. Ranzato, On Learning Where to Look, arXiv preprint:1405.5488.
[26] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[27] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, Extracting and composing robust features with denoising autoencoders, Proceedings of the 25th International Conference on Machine Learning, ICML'08 (2008) 1096–1103.
[28] A. Coates, A. Ng, H. Lee, An analysis of single-layer networks in unsupervised feature learning, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics 15 (2011) 215–223.
[29] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 185–207.
[30] B. Cheung, E. Weiss, B.A. Olshausen, Emergence of Foveal Image Sampling from Learning to Attend in Visual Scenes, CoRR abs/1611.09430.
[31] G. Kovács, A. Hajdu, A self-calibrating approach for the segmentation of retinal vessels by template matching and contour reconstruction, Med. Image Anal. 29 (2016) 24–46.
[32] 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.
[33] B. Zhang, L. Zhang, L. Zhang, F. Karray, Retinal vessel extraction by matched filter with first-order derivative of Gaussian, Comput. Biol. Med. 40 (4) (2010) 438–445.
[34] E. Bekkers, R. Duits, T. Berendschot, B. ter Haar Romeny, A multi-orientation analysis approach to retinal vessel tracking, J. Math. Imaging Vis. 49 (3) (2014) 583–610.
[35] Y. Yin, M. Adel, S. Bourennane, Retinal vessel segmentation using a probabilistic tracking method, Pattern Recognit. 45 (4) (2012) 1235–1244.
[36] J. De, H. Li, L. Cheng, Tracing retinal vessel trees by transductive inference, BMC Bioinform. 15 (1) (2014) 20.
[37] S. Roychowdhury, D.D. Koozekanani, K.K. Parhi, Iterative vessel segmentation of fundus images, IEEE Trans. Biomed. Eng. 62 (7) (2015) 1738–1749.
[38] B.S.Y. Lam, Y. Gao, A.W.C. Liew, General retinal vessel segmentation using regularization-based multiconcavity modeling, IEEE Trans. Med. Imaging 29 (7) (2010) 1369–1381.
[39] B. Al-Diri, A. Hunter, D. Steel, An active contour model for segmenting and measuring retinal vessels, IEEE Trans. Med. Imaging 28 (9) (2009) 1488–1497.
[40] 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.
[41] Y. Zhao, J. Zhao, J. Yang, Y. Liu, Y. Zhao, Y. Zheng, L. Xia, Y. Wang, Saliency driven vasculature segmentation with infinite perimeter active contour model, Neurocomputing 259 (2017) 201–209.
[42] 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.
[43] 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) (2016) 16–27.
[44] 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.
[45] S. Aslani, H. Sarnel, A new supervised retinal vessel segmentation method based on robust hybrid features, Biomed. Signal Process. Control 30 (2016) 1–12.
[46] L.C. Rodrigues, M. Marengoni, Segmentation of optic disc and blood vessels in retinal images using wavelets, mathematical morphology and Hessian-based multi-scale filtering, Biomed. Signal Process. Control 36 (2017) 39–49.
[47] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[48] G.D. Joshi, J. Sivaswamy, Colour retinal image enhancement based on domain knowledge, 2008 Sixth Indian Conference on Computer Vision, Graphics Image Processing (2008) 591–598.
[49] M. Foracchia, E. Grisan, A. Ruggeri, Luminosity and contrast normalization in retinal images, Med. Image Anal. 9 (3) (2005) 179–190.
[50] H. Larochelle, G.E. Hinton, Learning to combine foveal glimpses with a third-order Boltzmann machine, Advances in Neural Information Processing Systems 23 (2010) 1243–1251.
[51] M.A. Ranzato, C. Poultney, S. Chopra, Y.L. Cun, Efficient learning of sparse representations with an energy-based model, Advances in Neural Information Processing Systems 19 (2007) 1137–1144.
[52] M.A. Ranzato, Y. Lan Boureau, Y.L. Cun, Sparse feature learning for deep belief networks, in: J.C. Platt, D. Koller, Y. Singer, S.T. Roweis (Eds.), Advances in Neural Information Processing Systems 20, 2008, pp. 1185–1192.
[53] A.J. Bell, T.J. Sejnowski, The "independent components" of natural scenes are edge filters, Vis. Res. 37 (23) (1997) 3327–3338.
[54] K. Gregor, Y. LeCun, Learning fast approximations of sparse coding, Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10 (2010) 399–406.
[55] Randomforest-Matlab, 2009 https://code.google.com/archive/p/randomforest-matlab/.
[56] A.D. 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.
[57] IOSTAR Dataset, 2017 http://www.retinacheck.org.
[58] RC-SLO Dataset, 2017 http://www.retinacheck.org.
[59] L. van der Maaten, Accelerating t-SNE using tree-based algorithms, J. Mach. Learn. Res. 15 (2014) 3221–3245.
[60] A.M. Mendonca, A. Campilho, Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction, IEEE Trans. Med. Imaging 25 (9) (2006) 1200–1213.
[61] M.E. Martinez-Perez, A.D. Hughes, S.A. Thom, A.A. Bharath, K.H. Parker, Segmentation of blood vessels from red-free and fluorescein retinal images, Med. Image Anal. 11 (1) (2007) 47–61.
[62] X. You, Q. Peng, Y. Yuan, Y. Cheung, J. Lei, Segmentation of retinal blood vessels using the radial projection and semi-supervised approach, Pattern Recognit. 44 (10) (2011) 2314–2324.
[63] M. Fraz, S. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A. Rudnicka, C. Owen, An approach to localize the retinal blood vessels using bit planes and centerline detection, Comput. Methods Progr. Biomed. 108 (2) (2012) 600–616.
[64] C.A. Lupascu, D. Tegolo, E. Trucco, FABC: retinal vessel segmentation using AdaBoost, IEEE Trans. Inf. Technol. Biomed. 14 (5) (2010) 1267–1274.
[65] A. Christodoulidis, T. Hurtut, H.B. Tahar, F. Cheriet, A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images, Comput. Med. Imaging Graph. 52 (2016) 28–43.
[66] A. Fawzi, M. Davies, P. Frossard, Dictionary learning for fast classification based on soft-thresholding, Int. J. Comput. Vis. 114 (2) (2015) 306–321.
[67] Z. Tu, X. Bai, Auto-context and its application to high-level vision tasks and 3D brain image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 32 (10) (2010) 1744–1757.
[68] C.L. Srinidhi, P. Rath, J. Sivaswamy, A vessel keypoint detector for junction classification, 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (2017) 882–885.
[69] S.B. Rangrej, J. Sivaswamy, Assistive lesion-emphasis system: an assistive system for fundus image readers, J. Med. Imaging 4 (2017) 4–16.
[70] G.D. Joshi, J. Sivaswamy, K. Karan, R. Prashanth, S.R. Krishnadas, Vessel bend-based cup segmentation in retinal images, 2010 20th International Conference on Pattern Recognition (2010) 2536–2539.
[71] K. Ram, G.D. Joshi, J. Sivaswamy, A successive clutter-rejection-based approach for early detection of diabetic retinopathy, IEEE Trans. Biomed. Eng. 58 (3) (2011) 664–673.
[72] A.R. Hassan, M.A. Haque, Epilepsy and seizure detection using statistical features in the complete ensemble empirical mode decomposition domain, TENCON 2015–2015 IEEE Region 10 Conference (2015) 1–6.
[73] A.R. Hassan, A. Subasi, Automatic identification of epileptic seizures from EEG signals using linear programming boosting, Comput. Methods Progr. Biomed. 136 (2016) 65–77.
[74] A.R. Hassan, S. Siuly, Y. Zhang, Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating, Comput. Methods Progr. Biomed. 137 (2016) 247–259.
[75] A.R. Hassan, M.A. Haque, Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating, Biocybern. Biomed. Eng. 36 (1) (2016) 256–266.
[76] A.R. Hassan, Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting, Biomed. Signal Process. Control 29 (2016) 22–30.
[77] A.R. Hassan, M.I.H. Bhuiyan, Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating, Biomed. Signal Process. Control 24 (2016) 1–10.
[78] A.R. Hassan, M.I.H. Bhuiyan, Automatic sleep scoring using statistical features in the EMD domain and ensemble methods, Biocybern. Biomed. Eng. 36 (1) (2016) 248–255.
[79] A.R. Hassan, M.I.H. Bhuiyan, Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting, Comput. Methods Progr. Biomed. 140 (2017) 201–210.
[80] A.R. Hassan, M.A. Haque, An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting, Neurocomputing 235 (2017) 122–130.
[81] K. Gopinath, J. Sivaswamy, Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement, IEEE J. Biomed. Health Informat. PP (99) (2018) 1.
[82] R. Mehta, J. Sivaswamy, M-net: a convolutional neural network for deep brain structure segmentation, 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (2017) 437–440.