

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/319517239>

# Retinal blood vessel segmentation in high resolution fundus photographs using automated feature parameter estimation

Conference Paper · October 2017

DOI: 10.1117/12.2283539

CITATIONS

2

READS

967

4 authors, including:



**José Ignacio Orlando**

National Scientific and Technical Research Council

41 PUBLICATIONS 548 CITATIONS

[SEE PROFILE](#)



**Mariana del Fresno**

National University of the Center of the Buenos Aires Province

59 PUBLICATIONS 430 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Automated retinal image analysis [View project](#)



IVUS Image analysis and quantification [View project](#)

# Retinal blood vessel segmentation in high resolution fundus photographs using automated feature parameter estimation

José Ignacio Orlando<sup>a,b,c</sup>, Marcos Fracchia<sup>c</sup>, Valeria del Río<sup>c</sup> and Mariana del Fresno<sup>b,c,d</sup>

<sup>a</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina;

<sup>b</sup>Pladema Institute, Gral. Pinto 399, 7000 Tandil, Argentina;

<sup>c</sup>Facultad de Ciencias Exactas, UNCPBA, Pinto 399, 7000 Tandil, Argentina;

<sup>d</sup>Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA), Buenos Aires, Argentina;

## ABSTRACT

Several ophthalmological and systemic diseases are manifested through pathological changes in the properties and the distribution of the retinal blood vessels. The characterization of such alterations requires the segmentation of the vasculature, which is a tedious and time-consuming task that is infeasible to be performed manually. Numerous attempts have been made to propose automated methods for segmenting the retinal vasculature from fundus photographs, although their application in real clinical scenarios is usually limited by their ability to deal with images taken at different resolutions. This is likely due to the large number of parameters that have to be properly calibrated according to each image scale. In this paper we propose to apply a novel strategy for automated feature parameter estimation, combined with a vessel segmentation method based on fully connected conditional random fields. The estimation model is learned by linear regression from structural properties of the images and known optimal configurations, that were previously obtained for low resolution data sets. Our experiments in high resolution images show that this approach is able to estimate appropriate configurations that are suitable for performing the segmentation task without requiring to re-engineer parameters. Furthermore, our combined approach reported state of the art performance on the benchmark data set HRF, as measured in terms of the F1-score and the Matthews correlation coefficient.

**Keywords:** Retinal vessel segmentation, Fundus imaging, Parameter estimation

## 1. INTRODUCTION

Fundus photographs are a cost effective, non-invasive medical imaging modality that is widely used by ophthalmologists for manually inspecting the retina.<sup>1</sup> It is currently the most used imaging technique for screening several ophthalmic diseases such as diabetic retinopathy<sup>2</sup> and glaucoma,<sup>3</sup> which are among the leading causes of avoidable blindness in the world.<sup>4</sup> Current systems for automated fundus image analysis usually require to segment the vasculature first,<sup>5</sup> as blood vessels aid in numerous applications. In particular, vessel segmentations are used for characterizing pathological changes associated with ophthalmic and systemic diseases,<sup>1</sup> for localizing other anatomical parts of the retina,<sup>6</sup> for detecting abnormalities such as red lesions,<sup>7,8</sup> and as landmarks for multimodal image registration.<sup>9</sup>

Automated blood vessel segmentation is a challenging task that has been widely explored in the literature.<sup>1</sup> In general, it is tackled by means of supervised or unsupervised methods. Unsupervised methods are based on pixel level features such as Gabor filters,<sup>10</sup> line detectors<sup>11</sup> or morphological operation,<sup>12</sup> among others.<sup>13</sup> These features are designed to characterize vascular pixels, and afterwards they are thresholded to retrieve a binary representation of the vasculature.<sup>13</sup> Supervised methods are built on top of these strategies, and are based first on training a classifier from annotated data and then, on categorizing image pixels using such a model.<sup>10,14,15</sup> Although most of the current methods are able to achieve high performance on standard low resolution data sets such as DRIVE<sup>16</sup> and STARE,<sup>17</sup> they usually fail when images are taken at higher resolutions. This is likely

---

Further author information: (Send correspondence to J.I.O.)

J.I.O.: E-mail: jiorlando@conicet.gov.ar, Telephone: +54 249 4439690

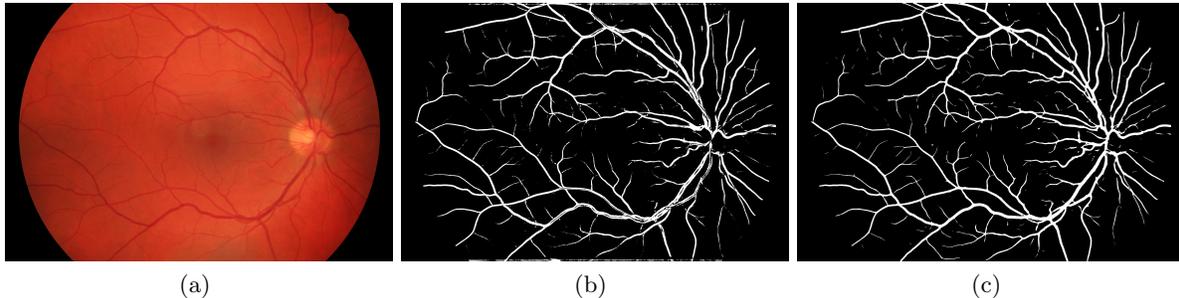


Figure 1: Retinal blood vessel segmentation in high resolution fundus images. (a) Fundus photograph. (b) Segmentation obtained using our method with parameters adapted using a scaling factor. (c) Segmentation obtained using our method with parameters estimated using linear regression.

because the existing features are highly parametrized and require an intensive calibration process to improve their original performance. However, it is extremely time consuming to tune these parameters using standard searching approaches such as grid search.<sup>18</sup> Furthermore, this process have to be repeated for every new data set with a different resolution.

One alternative to overcome this issue is to downscale the images to approximately fit the same resolution than those used for tuning parameters.<sup>3</sup> Nevertheless, this approach reduce the ability of the features to characterize thin structures, which is relevant in clinical applications.<sup>1</sup> Other strategies are based on modeling resolution changes and adjusting the feature parameters accordingly. In our previous work,<sup>19</sup> we proposed to apply a scaling factor, proportional to the change in the field of view (FOV) width, to automatically rescale those parameters. Albeit this approach is able to improve results on high resolution data sets, the resulting segmentations suffer from issues such as false negatives due to arteries central reflex (Figure 1), and are less accurate than the obtained on low resolution images. Vostatek *et al.*<sup>20</sup> have recently proposed a different strategy for predicting parameters, based on using linear regression. Such an approach is focused on correlating these values with the angular resolution of the images. A line is fitted to these points by minimizing the mean squared error, and its slope and intercept are subsequently used to automatically predict the parameters suitable for a new given resolution.

In this study we propose to take advantage of this recently published strategy by integrating it with our blood vessel segmentation method based on learning a fully connected conditional random field model.<sup>15,19</sup> In particular, we improve the original estimation strategy proposed by Vostatek *et al.*<sup>20</sup> by analyzing other structural parameters of the images that are also easily measurable. Moreover, we apply this estimator in combination with our supervised method, which incorporates shape priors within the learning process to better capture the interaction between vascular pixels.<sup>15,19</sup> Our hypothesis is that integrating this parameter estimator with our segmentation approach will result in better performance than merely using simple pixel classifiers such as Gaussian Mixture Models (GMMs).<sup>10,20</sup> We have evaluated this adaptive model on a benchmark data set of high resolution fundus images, HRF,<sup>18</sup> which is widely used in the literature for evaluating segmentation methods. Our results empirically shows that this hybrid approach significantly improve the original performance of the method, outperforming other existing strategies evaluated following a similar protocol.

The remainder of this paper is organized as follows. Section 2 explains our method, including details about the selected features, our segmentation approach and the parameter estimation strategy. Section 3 describes the data sets used in our experiments and the quantitative metrics applied for evaluation, while Section 4 presents the obtained results. Finally, Section 5 concludes the paper.

## 2. METHODS

A schematic representation of our method is depicted in Figure 2. Given different training sets of low resolution fundus images and their manual annotations, a grid search approach is followed to find the optimal configuration of feature parameters for each of their resolutions (Section 2.1.1). Subsequently, structural parameters such as

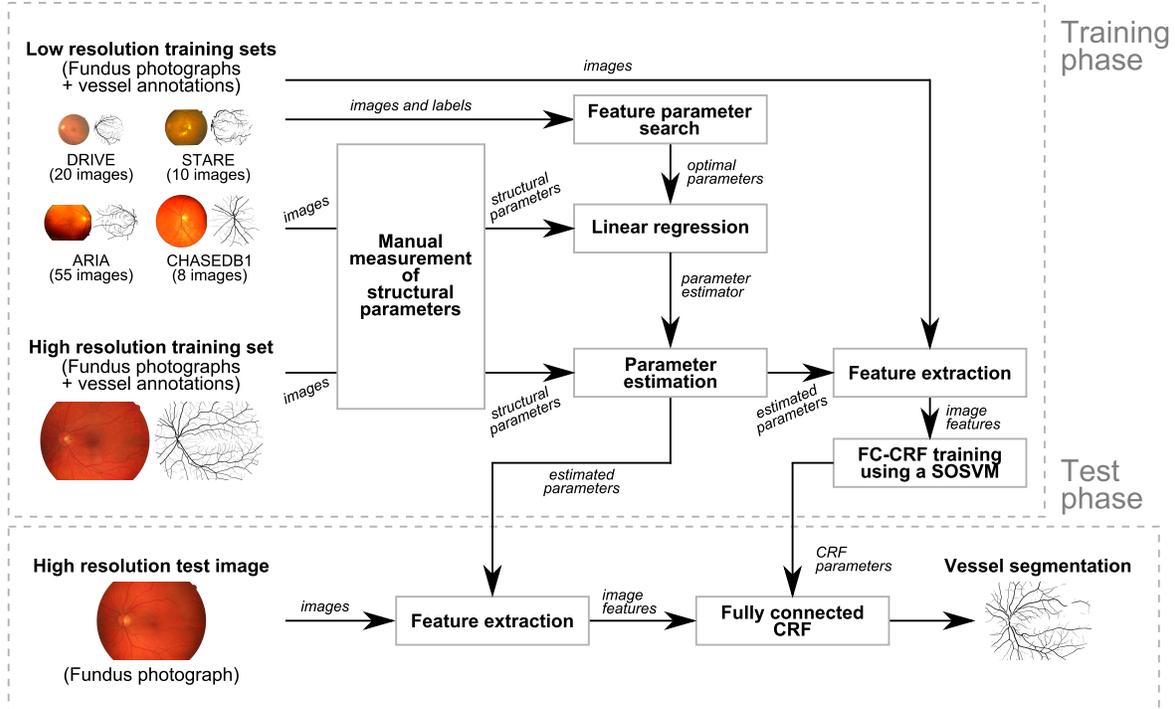


Figure 2: Schematic representation of our strategy for segmenting retinal vessels in high resolution images with automated parameter estimation.

the approximate diameter of the optic disc and the FOV, the calibre of the largest vessel and the ratio between the FOV diameter and the angle of aperture are measured from a subset of image examples. These values and the optimal parameters are afterwards used to fit an estimation line using linear regression. To learn the vessel segmentation model from other high resolution images, the structural parameters are taken from a subset of these other images, and they are used afterwards to automatically adjust the parameters of the selected features to this new resolution (Section 2.2). Finally, the features are extracted and used for training our supervised segmentation approach (Section 2.1.2), which is applied in test time to segment the vasculature on new images with unknown annotations.

## 2.1 Vessel segmentation approach

### 2.1.1 Feature extraction

As a proof of concept, we have retrieved a set of three different features that are widely applied in the literature for characterizing pixels belonging to blood vessels: the vessel enhancement technique based on mathematical morphology by reconstruction proposed by Zana and Klein,<sup>12</sup> the 2D Gabor wavelet by Soares *et al.*,<sup>10</sup> and the multiscale line detectors by Nguyen *et al.*<sup>11</sup> Nevertheless, this approach can be extended to other features different from those used in this paper. We will briefly describe them in the sequel to analyze their relevant parameters. The interested reader could refer to their original references and/or to a recently published review<sup>13</sup> for further information. Table 1 summarizes the parameters of these features.

The green color band is the one that exhibits the highest contrast between the retinal vasculature and the remaining anatomical parts of the fundus.<sup>13,21,22</sup> Hence, all the features are extracted from the inverted version of this specific channel. Furthermore, a larger angle of aperture is simulated as proposed by Soares *et al.*<sup>10</sup> to reduce potential artifacts in the borders of the FOV.

The feature based on morphological operations by reconstruction was originally introduced by Zana and Klein<sup>12</sup> for enhancing curvilinear structures on retinal angiographies and remote sensing imagery. It relies on a

series of morphological operations performed at different angles, using a linear structuring element of length  $l$ . By means of the application of openings, top-hats, a Laplacian of Gaussian, an additional opening, a closing and a last opening, linear connected elements whose curvature varies smoothly along a crest line are retrieved. We have used our own implementation of the feature, which has been made publicly available.<sup>19</sup> The design of its main parameter,  $l$ , was empirically observed to be extremely relevant to achieve a proper vessel enhancement. It was previously reported that the value of  $l$  is correlated with the caliber of the major vessel in the image.

The 2D Gabor wavelets are known by their intrinsic ability to capture oriented features, which become relevant for characterizing pixels belonging to the retinal vessels.<sup>10</sup> We have used the public implementation provided by Soares *et al.*<sup>10</sup> for extracting this feature. The main parameter of the method is the set of scales  $a$ , which is associated with the multiple calibres of the vessels in the image.

Finally, we also evaluated the application of line detectors, as proposed by Nguyen *et al.*<sup>11</sup> This approach is based on analyzing the response of the image to a line of length  $l \in \{1, \dots, W\}$  rotated at different angles.  $W$  corresponds to the largest length of the analyzed segments, and is a significant parameter of the method. As considering values of  $l$  too similar to each other in high resolution images would significantly increase the size of the feature vector with redundant information, we have restricted the number of potential scales to eight equidistant  $l$  values, spanning from 1 to  $W$ .

### 2.1.2 Fully connected CRF model for retinal vessel segmentation

The blood vessel segmentation task is tackled by means of a recently published method based on learning a Fully Connected Conditional Random Field (FC-CRF) model using a Structured Output Support Vector Machine (SOSVM). Such an approach has demonstrated to be effective enough for extracting the retinal vasculature,<sup>15,19</sup> and has been also applied in the context of other tasks such as automated glaucoma screening<sup>3</sup> or red lesion detection.<sup>8</sup> The interested reader could refer to the original reference for further details.

Formally, our purpose is to assign a labeling  $\mathbf{y} = \{y_i\}$  to every pixel  $i$  in the image  $I$ , with  $y_i \in \mathcal{L} = \{-1, 1\}$ , corresponding -1 to a non-vascular pixel and 1 to a vessel pixel. An estimated segmentation  $\mathbf{y}^*$  can be obtained by solving:

$$\mathbf{y}^* = \arg \min_{\mathbf{y} \in \mathcal{L}} E(\mathbf{y}|I) \quad (1)$$

where  $E(\mathbf{y}|I)$  is a Gibbs energy defined over the cliques of  $\mathcal{G}$  for a given labeling  $\mathbf{y}$  for  $I$ . This energy is given by:

$$E(\mathbf{y}|I) = \sum_i \psi_u(y_i, \mathbf{x}_i) + \sum_{(i,j) \in \mathcal{C}_{\mathcal{G}}} \psi_p(y_i, y_j, \mathbf{f}_i, \mathbf{f}_j) \quad (2)$$

where  $\mathbf{x}_i$  and  $\mathbf{f}_i$  are the unary and pairwise features, respectively. Unary potentials  $\psi_u$  define a log-likelihood over  $\mathbf{y}$ , and are obtained using a classifier.<sup>23</sup> On the contrary, pairwise potentials  $\psi_p$  define a similar distribution but over the interactions between pairs of pixels, as given by  $\mathcal{C}_{\mathcal{G}}$ . This set is defined by the graph connectivity rule: in our fully connected definition, all the pixels interact with each other.

Unary potentials are obtained as follows:

$$\psi_u(y_i, \mathbf{x}_i) = -\langle \mathbf{w}_{u_{y_i}}, \mathbf{x}_i \rangle - \mathbf{w}_{\beta_{y_i}} \beta \quad (3)$$

where  $\beta$  is a bias constant, and  $\mathbf{w}_{u_{y_i}}$  and  $\mathbf{w}_{\beta_{y_i}}$  are weight vectors for the features and the bias term, respectively, associated to the label  $y_i$ . The unary vector  $\mathbf{x}_i$  is given by an arbitrary combination of features extracted from the image (in this work, line detectors and 2D Gabor wavelets).

Pairwise potentials are restricted to be a linear combination of Gaussian kernels by the efficient inference approach by Krähenbühl and Koltun,<sup>23</sup> which is applied to minimize  $E(\mathbf{y}|I)$ . The pairwise energy is given by:

$$\psi_p(y_i, y_j, \mathbf{f}_i, \mathbf{f}_j) = \mu(y_i, y_j) \sum_{m=1}^M w_p^{(m)} k^{(m)}(f_i^{(m)}, f_j^{(m)}) \quad (4)$$

where each  $k^{(m)}$  is a fixed function over an arbitrary feature  $f^{(m)}$  (in this work, the response to the vessel enhancement method by Zana and Klein),  $w_p^{(m)}$  is a linear combination weight, and  $\mu(y_i, y_j)$  is a label compatibility function. The Gaussian kernels are used to quantify the similarity of  $f^{(m)}$  between neighboring pixels,

Table 1: Parameters to estimate.

Method	Reference	Identified parameters
<b>Mathematical morphology</b>	Zana and Klein, 2001 <sup>12</sup>	$l$
<b>2D Gabor Wavelets</b>	Soares <i>et al.</i> , 2006 <sup>10</sup>	$a$
<b>Line detectors</b>	Nguyen <i>et al.</i> , 2013 <sup>11</sup>	$W$
<b>Fully connected CRF</b>	Orlando <i>et al.</i> , 2014 <sup>15</sup> and 2017 <sup>19</sup>	$\theta_p$

while the compatibility function  $\mu$  penalizes similar pixels assigned to different labels, and is given by the Potts model  $\mu(y_i, y_j) = [y_i \neq y_j]$ .<sup>19</sup> The pairwise kernels have the following form:

$$k^{(m)}(f_i^{(m)}, f_j^{(m)}) = \exp\left(-\frac{|\mathbf{p}_i - \mathbf{p}_j|^2}{2\theta_p^2} - \frac{|f_i^{(m)} - f_j^{(m)}|^2}{2\theta_{(m)}^2}\right) \quad (5)$$

with  $\mathbf{p}_i$  and  $\mathbf{p}_j$  being the coordinate vectors of pixels  $i$  and  $j$ . Including the positions in the pairwise term allows to increase the effect of close pixel interactions over distant ones. The parameters  $\theta_p$  and  $\theta_{(m)}$  are used to control the degree of relevance of each kernel in the expression. Hence, if  $\theta_p$  increases, much longer interactions are taken into account. On the other hand, if  $\theta_p$  decreases, only local neighborhoods affect the result. Likewise, increasing or decreasing  $\theta_{(m)}$  will tolerate higher or lower differences on the pairwise feature. As in our previous work,  $\theta_{(m)}$  is fixed automatically as the median of a random sample of pairwise distances.<sup>19</sup> On the contrary,  $\theta_p$  must be properly adjusted as it strongly depends on the resolution of the images. We have evaluated if this parameter can be automatically determined by our estimation strategy, as described in Section 2.2.

The weights for both the unary and the pairwise potentials are learned using a SOSVM, as we have formerly proposed.<sup>15</sup> Further details about this supervised learning approach can be found in the reference.

## 2.2 Automated adjustment of feature parameters

Table 1 lists all the parameters to estimate: the length  $l$  of the structuring elements used by the Zana and Klein method, the scales  $a$  used to compute the 2D Gabor wavelets, the  $W_0$ ,  $W$  and *step* values for the line detectors, and the amplitude  $\theta_p$  of the fully connected CRF.

As previously mentioned, we used our version of the parameter estimation strategy proposed by Vostatek *et al.*<sup>20</sup> The original approach consists on first using low resolution images and their corresponding manual annotations to optimize each parameter by grid search, evaluating each configuration considering a performance measurement  $Q$ . Afterwards, a linear regression model must be learned from pairs of data points consisting on the optimal parameter value and its angular resolution. Our version of this approach introduces several modifications to the original pipeline. Vostatek *et al.* proposed to use the area under the ROC curve<sup>20</sup> as  $Q$ , to guide the optimization process. However, it has been previously demonstrated that this metric is affected by the degree of imbalance in the data.<sup>24</sup> Instead, we propose to use the area under the precision/recall curve to quantify the quality of the feature parameters, which is more appropriate to this type of problems.<sup>24,25</sup> The nature of the  $\theta_p$  parameter forces us to use a different optimization approach, as the FC-CRF have to be trained for a certain  $\theta_p$  value and then evaluated in terms of a binary segmentation metric. We performed this task as follows: given a value of  $\theta_p$ , we trained the FC-CRF on the training set and evaluated its contribution to improving the average F1-score (Section 3.2) in the validation set. This process was repeated for all the  $\theta_p$  values, and the parameter that reported the highest F1-score was taken as optimal.

Once the optimal parameters are found, different structural measurements are manually taken from 2 randomly sampled images of each subset in the set used for optimizing parameters. The average for each image pair is taken as a representative estimator of the other images in the subset. In particular, we have considered:

- **Largest vessel calibre** (in pixels): measured as the average length of 3 profiles manually drawn at different locations of the largest vessel.

Table 2: Data sets used in our experiments.

Data set	FOV (angle of aperture)	Resolution	Training set	Test set
<b>DRIVE</b>	45°	565 × 584	20 images	<i>Not used</i>
<b>STARE</b>	35°	605 × 700	10 images	<i>Not used</i>
<b>ARIA</b>	50°	768 × 576	55 images	<i>Not used</i>
<b>CHASEDB1</b>	30°	999 × 960	8 images	<i>Not used</i>
<b>HRF</b>	60°	3504 × 2336	15 images	30 images

- **Horizontal diameter of the optic disc** (in pixels): by measuring the length of a line horizontally drawn from the left to the right edge of the optic disc.
- **Width of the FOV** (in pixels): obtained automatically from the FOV binary masks.
- **Angular resolution**: taken as the ratio between the width of the FOV and the angle of aperture of the fundus camera.<sup>20</sup>

To analyze which of these structural measurements are more suitable for estimating each parameter, different lines are fitted to this data, and the coefficient of determination  $R^2$  of each linear regression model is used as an indicator of the overall model quality.<sup>26</sup> The structural measurement that resulted in the highest  $R^2$  value is taken as the optimal metric for a given model. Subsequently, this measure is manually taken from any new image, and the feature parameters are fixed according to the estimation provided by its corresponding model. Finally, F-tests were also performed to evaluate whether a linear regression model is suitable to perform the parameter estimation or not.<sup>26</sup>

### 3. EXPERIMENTAL SETUP

#### 3.1 Materials

As previously indicated in Section 2, our approach requires to be trained in two different ways. First, low resolution data sets are used to optimize feature parameters and to compute the corresponding estimators. Once these models are learned, a second training set, with approximately the same resolution than the test set, is needed to finally learn the fully connected CRF model.

To train our parameter estimators, we used the training sets of DRIVE<sup>16</sup> and CHASEDB1,<sup>27</sup> and two additional sets sampled from STARE<sup>17</sup> and ARIA.<sup>28,29</sup> Afterwards, the data set HRF<sup>18</sup> was applied for validating the complete pipeline. Table 2 summarizes the main characteristics of each database.

DRIVE<sup>16</sup> comprises 40 color fundus photographs (7 with pathologies), obtained from a diabetic retinopathy screening program in Netherlands. The set was originally divided into a training and a test set, each of them containing 20 images. However, we only used the training set in our experiments. STARE<sup>17</sup> contains 20 fundus images (10 of them with pathologies) commonly used to evaluate vessel segmentation algorithms. As the set is not divided into a training and a test set, we used the first 10 images to train the linear regression models. ARIA<sup>28,29</sup> is made up of three different groups of fundus images, 23 taken from patients with age related macular degeneration, 59 of patients with diabetes and 61 of healthy subjects. We built a training set from ARIA by extracting the first 8, 23 and 24 images from each subset, respectively. Finally, CHASEDB1<sup>27</sup> contains 28 fundus images of children, centered on the optic disc. This set is divided into a training and a test set, each of them containing 20 and 8 fundus photographs, respectively. We used these last 8 images for training our parameter estimator.

HRF<sup>18</sup> was used to validate our segmentation approach. It comprises 45 images, 15 of healthy subjects, 15 of patients with diabetic retinopathy and 15 of glaucomatous persons. As this data set was used for evaluating the full segmentation approach, we divided it as in<sup>19</sup> into a training and a test set. The training set is made up

of the first 5 images of each subset, while the remaining 30 images were used for test. Moreover, the training set was randomly split into a *training\** and a *validation* set, with the first one (10 images) used for learning the CRF model and the second one (5 images) for validating the regularization parameter  $C$ .

### 3.2 Evaluation metrics

Several metrics are used in the literature for evaluating blood vessel segmentation algorithms. In general, most of them are expressed in terms of sensitivity ( $Se$ , also known as recall,  $Re$ ), specificity ( $Sp$ ) and precision ( $Pr$ ), which are obtained as follows:

$$Se = Re = \frac{TP}{TP + FN} \quad (6)$$

$$Sp = \frac{TN}{TN + FP} \quad (7)$$

$$Pr = \frac{TP}{TP + FP} \quad (8)$$

Sensitivity quantifies the ability of the segmentation method to identify the vasculature, while precision measures how well the method is able to differentiate it with respect to other structures of the fundus. Similarly, specificity determines the capability of the method to properly distinguish the non-vascular structures.

As previously mentioned in Section 2.2, a metric  $Q$  is needed to guide the feature optimization procedure. All our experiments were performed using the area under the precision/recall curve to quantify feature’s performance. We choose this evaluation metric as it is appropriate to characterize features in imbalanced problems where the proportion of positive samples is smaller than the proportion of negative ones.<sup>24,25</sup>

To estimate the ability of our method to segment the vessels, we used  $Se$ ,  $Sp$  and  $Pr$ . Moreover, we included other global metrics such as the Matthews Correlation Coefficient, the  $F1$ -score and the  $G$ -mean, which are also robust under class imbalance.<sup>19</sup>

The Matthews Correlation Coefficient (MCC)<sup>22</sup> compares manual and automated segmentations, and is given by the equation:

$$MCC = \frac{TP/N - S \times P}{\sqrt{P \times S \times (1 - S) \times (1 - P)}} \quad (9)$$

where  $N = TP + TN + FP + FN$  is the total number of pixels in the image,  $S = (TP + FN)/N$  and  $P = (TP + FP)/N$ . It takes values between -1 and +1, where +1 indicates a perfect prediction, 0 is a random prediction and -1 a segmentation that is exactly the opposite than the true one.

The  $F1$ -score<sup>19</sup> is defined as the harmonic mean of the  $Pr$  and  $Re$ :

$$F1\text{-score} = \frac{2 \times Pr \times Re}{Pr + Re}. \quad (10)$$

Its maximum value, 1, corresponds to a perfect segmentation, while its minimum value, 0, corresponds to a completely wrong detection. This metric is equivalent to the Dice coefficient, which is also widely used for evaluating segmentation methods.

Finally, the  $G$ -mean has a similar behavior than the  $F1$ -score, although it is obtained as the geometric mean of the  $Se$  and  $Sp$ :

$$G\text{-mean} = \sqrt{Se \times Sp} \quad (11)$$

## 4. RESULTS

### 4.1 Parameter estimation

Table 4 presents the  $R^2$  values obtained using each structural measurement for fitting the parameter estimation models. For the Soares *et al.* feature we computed three different estimators, one for each of the three scales  $a = \{a_1, a_2, a_3\}$ . These values grow almost linearly with the image resolution, so the resulting model has a high  $R^2$  value. On the contrary, predicting  $\theta_p$  using linear regression appeared to be unfeasible, as the obtained  $R^2$

Table 3:  $R^2$  values obtained for each combination of feature parameter and structural measurements.  $p$ -values of the F-tests performed for each learned model are also included.

Parameter \ Measurement	$l$	$a_1$	$a_2$	$a_3$	$W$	$\theta_p$
Vessel calibre	0.941 $p \approx 6.1 \times 10^{-3}$	<b>0.953</b> $p \approx 4.4 \times 10^{-3}$	<b>0.985</b> $p \approx 7.5 \times 10^{-4}$	<b>0.974</b> $p \approx 1.8 \times 10^{-3}$	0.976 $p \approx 1.6 \times 10^{-3}$	0.003 $p \approx 0.936$
Optic disc diameter	<b>0.990</b> $p \approx 4.5 \times 10^{-4}$	0.945 $p \approx 5.6 \times 10^{-3}$	0.953 $p \approx 4.4 \times 10^{-3}$	0.914 $p \approx 1.1 \times 10^{-2}$	<b>0.985</b> $p \approx 7.6 \times 10^{-4}$	0.000 $p \approx 0.977$
FOV diameter	0.971 $p \approx 2.2 \times 10^{-3}$	0.813 $p \approx 3.6 \times 10^{-2}$	0.818 $p \approx 3.5 \times 10^{-2}$	0.734 $p \approx 6.4 \times 10^{-2}$	0.898 $p \approx 1.4 \times 10^{-2}$	0.016 $p \approx 0.838$
Angular resolution	0.973 $p \approx 1.9 \times 10^{-3}$	0.947 $p \approx 5.2 \times 10^{-3}$	0.935 $p \approx 7.2 \times 10^{-3}$	0.914 $p \approx 1.1 \times 10^{-2}$	0.982 $p \approx 1.1 \times 10^{-3}$	0.000 $p \approx 0.996$

values are close to 0. This means that the simplest possible model, which is the average of the samples used for learning the line, performs much better than the estimated model.

When analyzing each structural measurement separately, it is possible to see that the diameter of the optic disc allows to obtain the best estimations of  $W$  and  $l$ , while the calibre of the largest vessel is the best predictor for  $a$ . This results are complementary to those reported by Vostatek *et al.*, who only analyzed the number of pixels in the ground truth labeling and the angular resolution. The  $p$ -values reported by the F-tests performed for each learned model also support the idea that using lines to estimate feature parameters is valuable, although not for  $\theta_p$ .

Figure 3 illustrates the best parameter estimators for each feature and for  $\theta_p$ . Optimal values obtained by adjusting each parameter on the HRF training set are also included for comparison purposes, although they were not used to fit the model. It is possible to see that  $l$ ,  $a$  and  $W$  grows linearly with their corresponding structural measurement, which justify the usage of linear regression for fitting their values. On the contrary, the optimal  $\theta_p$  values of the low resolution data sets do not change linearly, a setting that explains why they cannot be properly approximated with linear regression.

## 4.2 Segmentation results

Segmentation results using our approach with automated feature parameter estimation are given in Table 4. As mentioned in Section 4.1, estimating  $\theta_p$  using a linear regression model is not feasible due to its non-linear behavior with respect to the selected structural measurements. Hence,  $\theta_p$  was fixed to the optimal value obtained according to the validation set sampled from the HRF training set. We also included other works in the literature that used the same evaluation protocol and/or training and test splits. Vostatek *et al.*<sup>20</sup> reported the performance obtained by evaluating the supervised method by Soares *et al.*<sup>10</sup> on HRF. Such an approach is based on learning a Gaussian Mixture Model classifier from the responses to the 2D Gabor wavelets. Vostatek *et al.* trained this method using a random sample of 15 images taken from HRF. It is unfeasible to perform an exact comparison as we have no certainty that the images used for testing are exactly the same than those used to evaluate our model. However, we also included these results to provide a general idea of the contribution of the fully connected CRF model with respect to using the original classifier. A series of Wilcoxon signed-ranks hypothesis tests were performed to compare the results obtained by our method with respect to our previous approach<sup>19</sup> and to those obtained by Odstrčilík *et al.*<sup>18</sup>

As seen in Table 4, our approach consistently perform better than our previous proposal based on scaling parameters using a compensation factor. The improvements obtained by adapting the feature parameters with our strategy and the selection of an optimal  $\theta_p$  value, as measured by all the considered quality metrics, are also statistically significant. In particular, this strategy is able to achieve consistently higher F1-score ( $p < 9.2 \times 10^{-7}$ ),  $G$ -mean ( $p < 9.2 \times 10^{-7}$ ) and MCC ( $p < 9.2 \times 10^{-7}$ ) values, which corresponds to a general improvement in the quality of the results. When decomposing these metrics in terms of their individual measurements, it is possible to see that the  $Se$  is significantly improved ( $p < 9.2 \times 10^{-7}$ ) by the estimation of the features, a setting that is related with a better ability to detect thin structures (Figure 4) and to overcome the issues of the original approach to deal with the bright central reflex in arteries (Figure 5). Moreover, larger  $Sp$  ( $p < 0.0044$ ) and  $Pr$  ( $p < 6 \times 10^{-6}$ ) values indicate a reduction in the number of false positive detections.

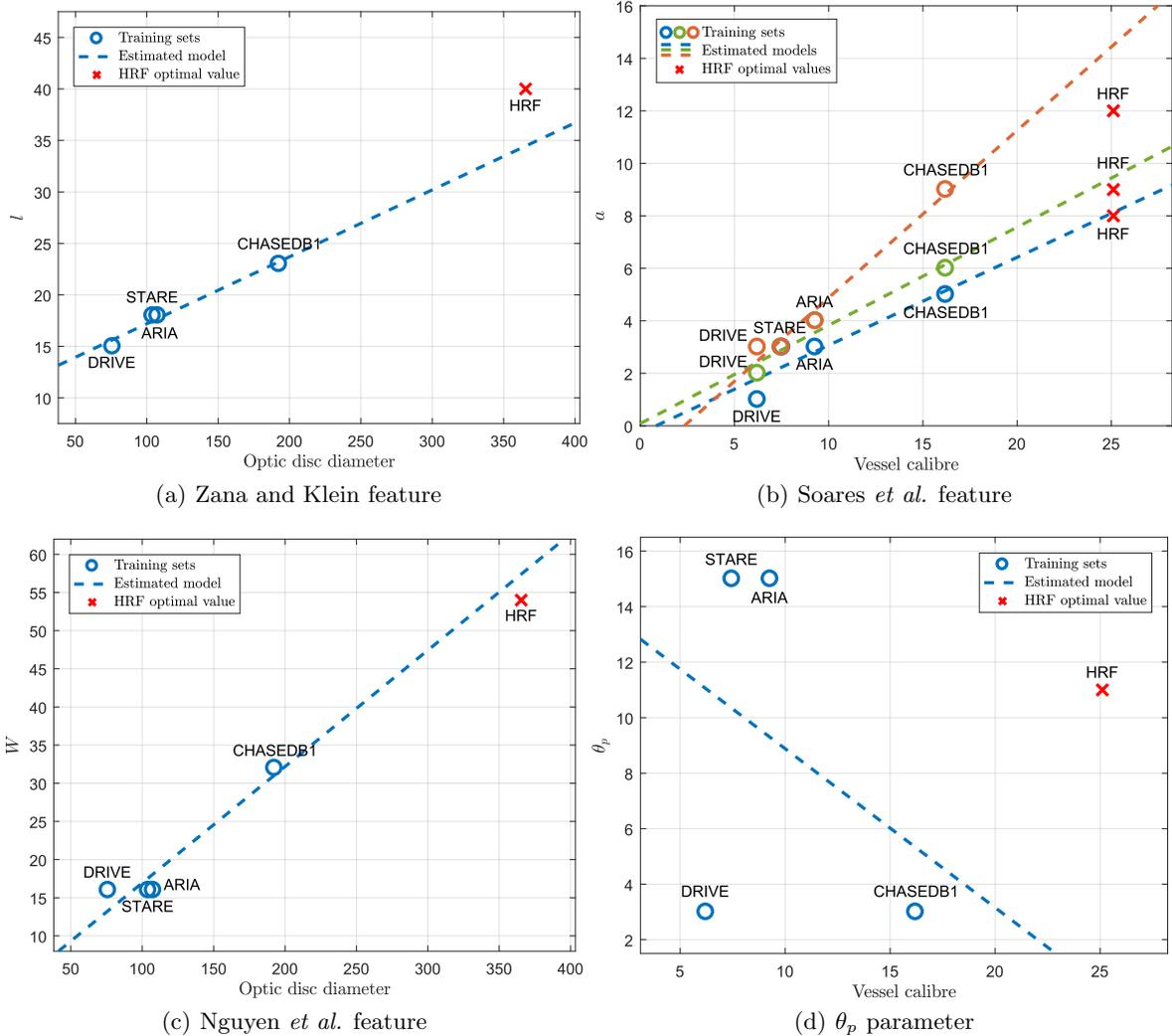


Figure 3: Best parameter estimators for each feature parameter. Optimal values on HRF are included only for comparison purposes, but were not used to fit the linear regression model.

Compared to other existing methods, it is worth noting that our approach achieved the highest average F1-score ( $p < 4.1 \times 10^{-6}$ ) and MCC ( $p < 1.9 \times 10^{-6}$ ) values. A higher average  $G$ -mean was obtained by the baseline method of Odstrčilík *et al.*<sup>18</sup> Yet, the difference is not statistically significant ( $p = 0.23$ ). Such an approach reported also a higher average  $Se$  value than our method, but it is not statistically significant ( $p = 0.09$ ). Furthermore, it is important to underline that the method by Odstrčilík *et al.* is based on matched filter responses that are recovered from filters calibrated for this specific data set. In our case, we used an automated parameter estimation approach that does not require such an intensive calibration. Moreover, our method achieve higher  $Sp$  ( $p < 0.006$ ) and  $Pr$  ( $p < 1.3 \times 10^{-5}$ ) values, which correspond to a reduction in the number of false positive detection.

## 5. DISCUSSION

In this paper we have presented an ensemble approach for blood vessel segmentation in high resolution images, based on automatically estimating feature parameters. In particular, we have integrated a novel strategy for parameter estimation using linear regression with a fully connected CRF model, which is known to achieve better

Table 4: Results obtained on HRF.

Methods	Se	Sp	Pr	F1	G-mean	MCC
Odstrcilik <i>et al.</i> , 2013 <sup>18</sup>	<b>0.7772</b>	0.9652	0.6950	0.7316	<b>0.8657</b>	0.7065
Vostatek <i>et al.</i> , 2017 (Soares) <sup>20</sup>	0.7340	<b>0.9800</b>	-	-	0.8481	-
Vostatek <i>et al.</i> , 2017 (Sofka) <sup>20</sup>	0.5830	0.9780	-	-	0.7550	-
Orlando <i>et al.</i> , 2017 <sup>19</sup>	0.7201	0.9713	0.7199	0.7168	0.8361	0.6900
<b>Our approach</b>	0.7669	0.9725	<b>0.7407</b>	<b>0.7503</b>	0.8636	<b>0.7267</b>

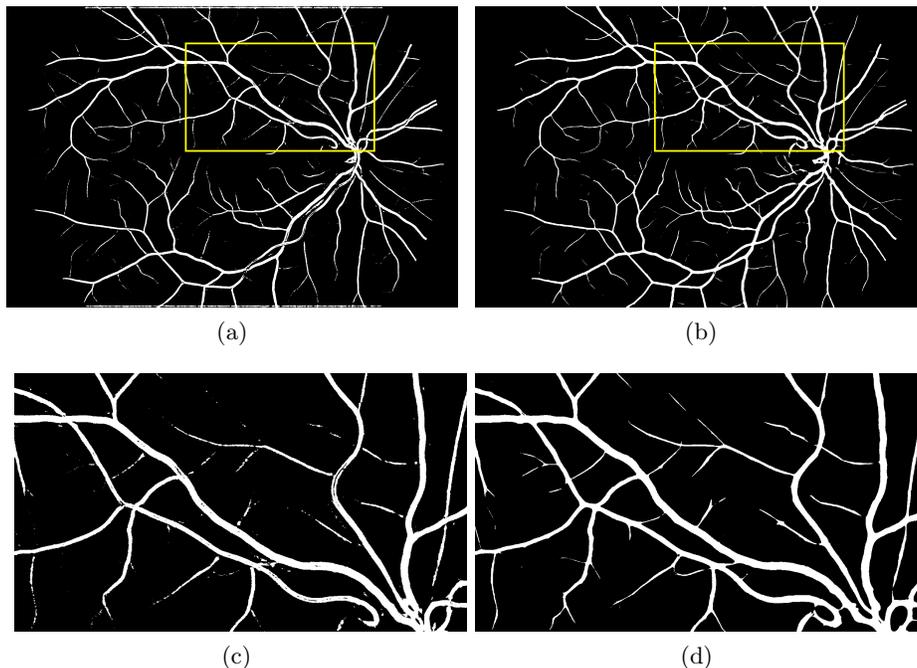


Figure 4: Detection of thin vessels, as seen on results obtained on image 10.h from HRF. (a, c) Results obtained using the  $\rho$  multiplier. (b, d) Results obtained using our approach.

results than other existing approaches. We have experimentally analyzed different structural measurements of the images and their potential usage as guidelines to automatically fit a regression line. Our results indicated that the optic disc diameter is suitable to estimate the parameters of the line detectors and the feature based on morphology by reconstruction, while the calibre of the major vessel is the best structural measurement to estimate the scales of the 2D Gabor filter. On the contrary, the experiments made to automatically adjust  $\theta_p$  indicated that this parameter does not scale linearly with respect to the resolution of the images (Figure 3(d)). When analyzing the optimal  $\theta_p$  values for each individual data set, we can see that the higher parameters were assigned to STARE and ARIA, while the lower values corresponded to DRIVE and CHASEDB1. STARE and ARIA are characterized by serious pathological cases in which large hemorrhages or exudates occur. On the contrary, images on DRIVE and CHASEDB1 correspond mostly to healthy patients. HRF also contains pathological images, although lesions are smaller. This might indicate that larger  $\theta_p$  values are more suitable to be used to segment images of patients with large pathologies.

When evaluating the segmentation method quantitatively, it was observed that integrating the parameter estimation approach improved all the evaluation metrics with respect to using the original scaling factor. Furthermore, the comparison to other works showed that our approach performed consistently better than other existing approaches that were evaluated using a similar training and test split.

In conclusion, it is possible to see that the estimation strategy applied in this context allows to obtain better

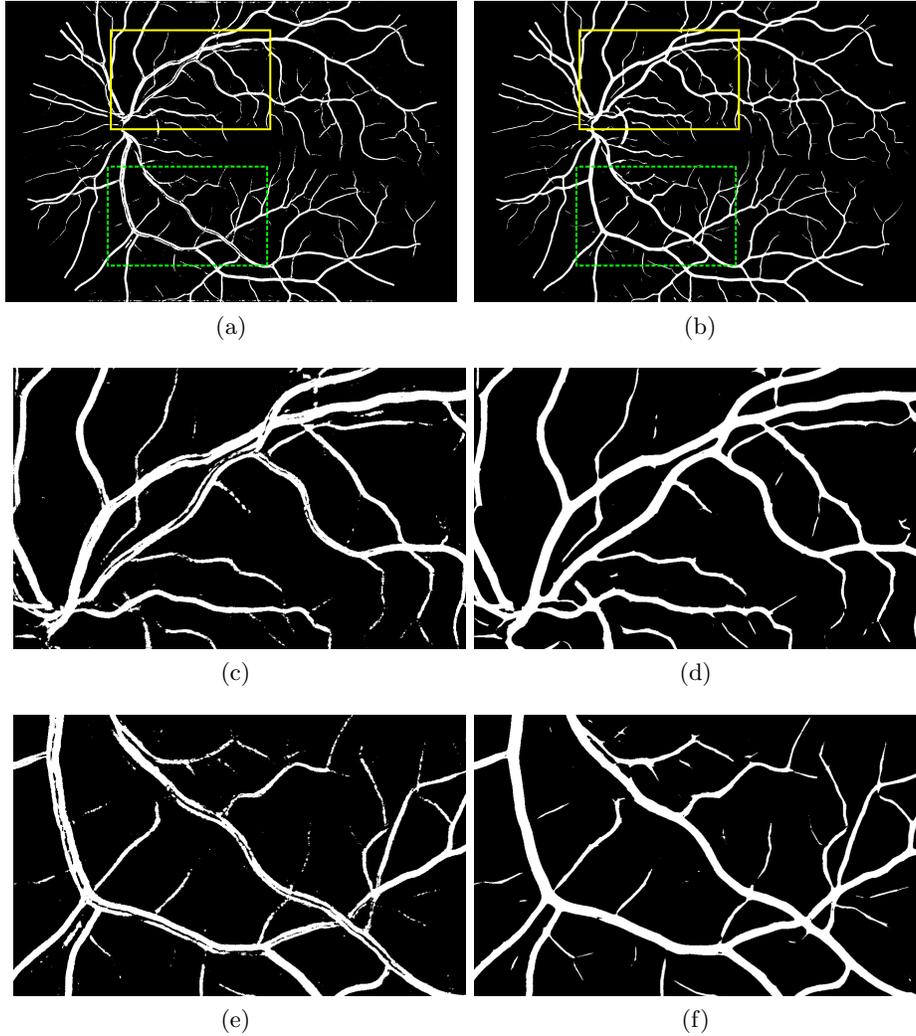


Figure 5: Improved segmentation of arteries with bright central reflex, as seen on image 12\_h from HRF. (a, c, e) Results obtained using the  $\rho$  multiplier. (b, d, f) Results obtained using our approach.

results in terms of overall quality measurements, with a consistent improvement in the detection of the thinner vessels and a more appropriate behavior under the presence of bright central reflex. This approach can be exploited not only for adjusting the parameters of hand crafted features but also to calibrate deep learning based methods, for instance, which are usually trained using patches whose size depends on the image resolution.<sup>30</sup> Segmentation masks and further implementation details are provided in <https://github.com/ignaciorlando/high-resolution-vessel-segmentation>.

### Acknowledgments

This work is partially funded by a NVIDIA Hardware Grant and ANPCyT PICT 2014-1730, PICT 2016-0116 and PICT start-up 2015-0006. J.I.O. is funded by a doctoral scholarship granted by CONICET. We would also like to thank Odstrčilík *et al.* for providing us with their segmentations.

## REFERENCES

- [1] Fraz, M. M. et al., “Blood vessel segmentation methodologies in retinal images—a survey,” *Computer Methods and Programs in Biomedicine* **108**(1), 407–433 (2012).
- [2] Mookiah, M. R. K., Acharya, U. R., Chua, C. K., Lim, C. M., Ng, E., and Laude, A., “Computer-aided diagnosis of diabetic retinopathy: A review,” *Computers in Biology and Medicine* **43**(12), 2136–2155 (2013).
- [3] Orlando, J. I., Prokofyeva, E., del Fresno, M., and Blaschko, M., “Convolutional neural network transfer for automated glaucoma identification,” in [*12th International Symposium on Medical Information Processing and Analysis*], 101600U–101600U, International Society for Optics and Photonics (2017).
- [4] Prokofyeva, E. and Zrenner, E., “Epidemiology of major eye diseases leading to blindness in Europe: A literature review,” *Ophthalmic Research* **47**(4), 171–188 (2012).
- [5] Abràmoff, M. D. et al., “Retinal imaging and image analysis,” *IEEE Reviews in Biomedical Engineering* **3**, 169–208 (2010).
- [6] Mendonca, A. M., Sousa, A., Mendonça, L., and Campilho, A., “Automatic localization of the optic disc by combining vascular and intensity information,” *Computerized medical imaging and graphics* **37**(5), 409–417 (2013).
- [7] Gupta, G., Ram, K., Kulasekaran, S., Joshi, N., Sivaprakasam, M., and Gandhi, R., “Detection of retinal hemorrhages in the presence of blood vessels,” in [*Proceedings of the Ophthalmic Medical Image Analysis First International Workshop, OMIA 2014, Held in Conjunction with MICCAI 2014*], Chen X, Garvin MK, L. J., ed., **1**, 105–112, Iowa Research Online (2014).
- [8] Orlando, J. I., Prokofyeva, E., del Fresno, M., and Blaschko, M. B., “Learning to detect red lesions in fundus photographs: An ensemble approach based on deep learning,” *arXiv preprint arXiv:1706.03008* (2017).
- [9] Zheng, Y., Daniel, E., Hunter, A. A., Xiao, R., Gao, J., Li, H., Maguire, M. G., Brainard, D. H., and Gee, J. C., “Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix,” *Medical image analysis* **18**(6), 903–913 (2014).
- [10] Soares, J. V. et al., “Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification,” *IEEE Transactions on Medical Imaging* **25**(9) (2006).
- [11] Nguyen, U. T. et al., “An effective retinal blood vessel segmentation method using multi-scale line detection,” *Pattern Recognition* **46**(3), 703–715 (2013).
- [12] Zana, F. and Klein, J.-C., “Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation,” *IEEE Transactions on Image Processing* **10**(7), 1010–1019 (2001).
- [13] Orlando, J. I. and del Fresno, M., “Reviewing preprocessing and feature extraction techniques for retinal blood vessel segmentation in fundus images,” *Mecánica Computacional XXXIII*(42), 2729–2743 (2014).
- [14] Sofka, M. and Stewart, C. V., “Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures,” *Medical Imaging, IEEE Transactions on* **25**(12), 1531–1546 (2006).
- [15] Orlando, J. I. and Blaschko, M., “Learning fully-connected CRFs for blood vessel segmentation in retinal images,” in [*Medical Imaging Computing and Computer Assisted Intervention*], Golland, P., Barillot, C., Hornegger, J., and Howe, R., eds., **8149**, 634–641, Springer (2014).
- [16] Niemeijer, M. et al., “Comparative study of retinal vessel segmentation methods on a new publicly available database,” in [*Medical Imaging 2004*], 648–656, International Society for Optics and Photonics (2004).
- [17] Hoover, A. et al., “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response,” *IEEE Transactions on Medical Imaging* **19**(3), 203–210 (2000).
- [18] Odstrčilík, J., Kolar, R., Budai, A., Hornegger, J., Jan, J., Gazarek, J., Kubena, T., Cernosek, P., Svoboda, O., and Angelopoulou, E., “Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database,” *IET Image Processing* **7**(4), 373–383 (2013).
- [19] Orlando, J. I., Prokofyeva, E., and Blaschko, M. B., “A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images,” *IEEE Transactions on Biomedical Engineering* **64**(1), 16–27 (2017).
- [20] Vostatek, P., Claridge, E., Uusitalo, H., Hauta-Kasari, M., Fält, P., and Lensu, L., “Performance comparison of publicly available retinal blood vessel segmentation methods,” *Computerized Medical Imaging and Graphics* **55**, 2–12 (2017).

- [21] Marín, D., Aquino, A., Gegúndez-Arias, M. E., and Bravo, J. M., “A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features,” *IEEE Transactions on Medical Imaging* **30**(1), 146–158 (2011).
- [22] Azzopardi, G., Strisciuglio, N., Vento, M., and Petkov, N., “Trainable cosfire filters for vessel delineation with application to retinal images,” *Medical image analysis* **19**(1), 46–57 (2015).
- [23] Krähenbühl, P. and Koltun, V., “Efficient inference in fully connected CRFs with Gaussian edge potentials,” in [*Advances in Neural Information Processing Systems*], 109–117 (2012).
- [24] Saito, T. and Rehmsmeier, M., “The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets,” *PloS one* **10**(3), e0118432 (2015).
- [25] Lo Vercio, L., Orlando, J. I., del Fresno, M., and Larrabide, I., “Assessment of image features for vessel wall segmentation in intravascular ultrasound images,” *International Journal of Computer Assisted Radiology and Surgery* **11**(8), 1397–1407 (2016).
- [26] Lomax, R. G. and Hahs-Vaughn, D. L., [*Statistical concepts: A second course*], Routledge (2013).
- [27] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., and Barman, S. A., “An ensemble classification-based approach applied to retinal blood vessel segmentation,” *IEEE Transactions on Biomedical Engineering* **59**(9), 2538–2548 (2012).
- [28] Zheng, Y., Hijazi, M. H. A., and Coenen, F., “Automated disease/no disease grading of age-related macular degeneration by an image mining approach,” *Investigative Ophthalmology & Visual Science* **53**(13), 8310–8318 (2012).
- [29] Fumero, F., Sigut, J., Alayón, S., González-Hernández, M., and González, M., “Interactive tool and database for optic disc and cup segmentation of stereo and monocular retinal fundus images,” in [*Short Papers Proceedings–WSCG 2015*], 91–97 (2015).
- [30] Liskowski, P. and Krawiec, K., “Segmenting retinal blood vessels with deep neural networks,” *IEEE Transactions on Medical Imaging* **35**(11), 2369–2380 (2016).