How to Estimate Fovea Position When The Fovea Cannot Be Identified Visually Anymore?

Vincent Fournet, Aurelie Calabrese, Séverine Dours, Frédéric Matonti, Eric Castet, Pierre Kornprobst

How to Estimate Fovea Position When The Fovea Cannot Be Identified Visually Anymore?

Vincent Fournet, Aurélie Calabrèse, Séverine Dours, Frédéric Matonti, Eric Castet, Pierre Kornprobst

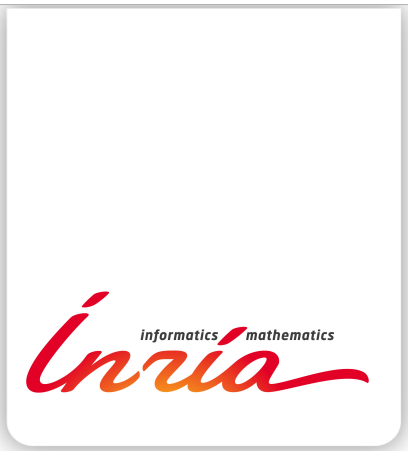
RESEARCH REPORT

N° 9419

September 2021

Project-Team Biovision

ISSN 0249-6399 ISRN INRIA/RR--9419--FR+ENG



How to Estimate Fovea Position When The Fovea Cannot Be Identified Visually Anymore?

Vincent Fournet*, Aurélie Calabrèse*, Séverine Dours†‡,

Frédéric Matonti§, Eric Castet†, Pierre Kornprobst*

Project-Team Biovision

Research Report n° 9419 — September 2021 — 45 pages

* Université Côte d'Azur, Inria, France
† Aix Marseille Univ, CNRS, LPC, Marseille, France
‡ Arc en Ciel
§ Centre Monticelli Paradis d'Ophtalmologie, Marseille, France

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Abstract:
In the presence of maculopathies, due to structural changes in the macula region, the fovea is usually located in pathological fundus images using normative anatomical measures (NAM). This simple method relies on two conditions: that images are acquired under standard testing conditions (primary head position and central fixation) and that the optic disk is visible entirely on the image. However, these two conditions are not always met in the case of maculopathies, en particulier lors de taches de fixations. Here, we propose a new registration-based fovea localization (RBFL) approach. The spatial relationship between fovea location and vessel characteristics (density and direction) is learned from 840 annotated healthy fundus images and then used to predict the precise fovea location in new images. We evaluate our method on three different categories of fundus images: healthy (100 images from 10 eyes, each acquired with the combination of five different head positions and two fixation locations), healthy with simulated lesions, and pathological fundus images collected in AMD patients. Compared to NAM, RBFL reduced the mean fovea localization error by 59% in normal images, from 2.85°of visual angle (SD 2.33) to 1.16°(SD 0.86), and the median error by 53%, from 1.93°to 0.89°. In cases of right-left head tilt, the mean error is reduced by 76%, from 5.23°(SD 1.95) to 1.28°(SD 0.9). With simulated lesions of 400 deg^2, the proposed RBFL method still outperforms NAM with a 10% mean error decrease, from 2.85°(SD 2.33) to 2.54°(SD 1.9). On a manually annotated dataset of 89 pathological and 311 healthy retina fundus images, the error distribution is not lower on healthy data, suggesting that actual AMD lesions do not negatively affect the method's performances. The vascular structure provides enough information to precisely locate the fovea in fundus images in a way that is robust to head tilt, eccentric fixation location, missing vessels, and real macular lesions.

Key-words:
Maculopathies, fovea, fundus images, microperimetry

Comment Estimer La Position De La Fovéa Quand La Fovéa Ne Peut Pas Etre Identifée Visuellement?

Résumé :

En présence de maculopathies, dues à des modifications structurelles de la région de la macula, la fovéa est généralement localisée sur les images pathologiques du fond d'œil à l'aide de mesures anatomiques normatives (NAM). Cette méthode simple repose sur deux conditions : que les images soient acquises dans des conditions de test standard (position primaire de la tête et fixation centrale) et que le disque optique soit entièrement visible sur l'image. Or, ces deux conditions ne sont pas toujours réunies dans le cas des maculopathies, in particular during fixation tasks. Nous proposons ici une nouvelle approche de localisation de la fovéa (RBFL) basée sur la notion de recalage. La relation spatiale entre l'emplacement de la fovéa et les caractéristiques des vaisseaux (densité et direction) est apprise à partir de 840 images de fonds d'œil sains annotées, puis utilisée pour prédire l'emplacement précis de la fovéa dans de nouvelles images. Nous évaluons notre méthode sur trois catégories différentes d'images du fond d'œil : des images saines (100 images provenant de 10 yeux, chacune acquise avec la combinaison de cinq positions différentes de la tête et de deux emplacements de fixation), des images saines avec des lésions simulées, et des images pathologiques du fond d'œil recueillies chez des patients atteints de DMLA. Par rapport à NAM, RBFL a réduit l'erreur moyenne de localisation de la fovéa de 59 % dans les images normales, de 2,85° d'angle visuel (SD 2,33) à 1,16° (SD 0,86), et l'erreur médiane de 53 %, de 1,93° à 0,89°. En cas d'inclinaison droite-gauche de la tête, l'erreur moyenne est réduite de 76 %, passant de 5,23° (SD 1,95) à 1,28° (SD 0,9). Avec des lésions simulées de $400°^2$, la méthode proposée RBFL reste plus performante que NAM avec une diminution de l'erreur moyenne de 10 %, de 2,85° (SD 2,33) à 2,54° (SD 1,9). Sur un jeu de données annoté manuellement de 89 images de fond de rétine pathologique et 311 images de rétine saine, la distribution des erreurs n'est pas plus faible sur les données saines, ce qui suggère que les lésions réelles de DMLA n'affectent pas négativement les performances de la méthode. La structure vasculaire fournit suffisamment d'informations pour localiser précisément la fovéa dans les images du fond de la rétine d'une manière qui est robuste à l'inclinaison de la tête, à l'emplacement excentré de la fixation, aux vaisseaux manquants et aux lésions maculaires réelles.

Mots-clés :

Maculopathies, fovéa, images du fond d'œil, microperimétrie.

Contents

1 Introduction **6**
1.1 Context . . . 6
1.2 Image Processing Techniques . . . 7
1.3 Challenges and Goal . . . 9
1.4 Paper Outline . . . 9

2 Registration-Based Fovea Localization Method (RBFL) **10**
2.1 General Principle . . . 10
2.2 Statistical Representation . . . 11
2.2.1 Average Density ($\bar{V}$) . . . 11
2.2.2 Average Directions ($\bar{D}$) . . . 13
2.3 Registation Criteria . . . 15
2.3.1 Energy Definition . . . 15
2.3.2 Definition of the Region-Based Weight ($\lambda_{\mathcal{T}}$) . . . 18
2.4 Algorithmic Details . . . 19
2.4.1 Multiscale . . . 19
2.4.2 Optimization Method, Parameters Initialization, Bounds and Order . . . 19
2.4.3 Enhancement of the Density . . . 21

3 Results **22**
3.1 Testing Set: Fundus Images on Healthy Subjects with Different Head Positions and Fixations . . . 23
3.2 Global Performance . . . 23
3.3 Effect of Fixation Location and Head Position . . . 26
3.4 Effect of Axial Length/Myopia on the Method . . . 28
3.5 Relationship Between the Transformation Parameters and Head Position and Fixation . . . 28
3.6 Simulated Maculopathies . . . 29
3.7 Results on Real AMD Images . . . 31

4 Discussion **32**
4.1 Are There Other Anatomical Landmarks and Characteristics? . . . 32
4.2 Structure Tensor Registration is Not Efficient . . . 32
4.3 Could Deep Learning Methods Be Useful in Our Problem? . . . 33
4.3.1 What About Solving The Problem in 3D? . . . 33
4.4 A Method for Optic Disk Detection? . . . 34

5 Conclusion **35**

6 Acknowledgements **36**

A Appendix **37**
A.1 Parameters Setting and Computational Performances . . . 37
A.2 The Notion of Tensors for Representing Direction Distributions . . . 37
A.3 SA-UNet and Its Application to Our Case . . . 39

In the presence of maculopathies, due to structural changes in the macula region, the fovea is usually located in pathological fundus images using normative anatomical measures (NAM). This simple method relies on two conditions: that images are acquired under standard testing conditions (primary head position and central fixation) and that the optic disk is visible entirely on the image. However, these two conditions are not always met in the case of maculopathies, en particulier lors de taches de fixations. Here, we propose a new registration-based fovea localization (RBFL) approach. The spatial relationship between fovea location and vessel characteristics (density and direction) is learned from 840 annotated healthy fundus images and then used to predict the precise fovea location in new images. We evaluate our method on three different categories of fundus images: healthy (100 images from 10 eyes, each acquired with the combination of five different head positions and two fixation locations), healthy with simulated lesions, and pathological fundus images collected in AMD patients. Compared to NAM, RBFL reduced the mean fovea localization error by 59% in normal images, from 2.85°of visual angle (SD 2.33) to 1.16°(SD 0.86), and the median error by 53%, from 1.93°to 0.89°. In cases of right-left head tilt, the mean error is reduced by 76%, from 5.23°(SD 1.95) to 1.28°(SD 0.9). With simulated lesions of 400 deg^2, the proposed RBFL method still outperforms NAM with a 10% mean error decrease, from 2.85°(SD 2.33) to 2.54°(SD 1.9). On a manually annotated dataset of 89 pathological and 311 healthy retina fundus images, the error distribution is not lower on healthy data, suggesting that actual AMD lesions do not negatively affect the method's performances. The vascular structure provides enough information to precisely locate the fovea in fundus images in a way that is robust to head tilt, eccentric fixation location, missing vessels, and real macular lesions.

1 Introduction

1.1 Context

Retinal images and maculopathies Fundus images are photographs of the retina (i.e., the back of the eye) that allow to visualize the main structures of the retina at the macroscopic level. For normal eyes, the main visible structures are: the optic disk (i.e., the head of the optic nerve, located in the temporal area), the macula (i.e., the central region of the retina centered onto the fovea) and the vascular branches (composed of veins and arteries branching out from the optic disk) (Fig. 1A). While the vessel trajectories are different in every individual eye (Mutlu and Leopold, 1964; Semerád and Drahanský, 2020), they obey some identifiable patterns. For example, the optic disk area is systematically the region with the highest vessel density, while the macular area is much more empty.

Among the several devices that integrate a fundus camera, microperimeters are powerful ophthalmic tools that allow to take color fundus images while also mapping the visual field and measuring fixation stability. Therefore, these devices are of great interest when dealing with patients suffering from retinal defects, and especially central field loss (CFL) lead by maculopathies.

Maculopathies, such as Age-related Macular Degeneration (AMD), are pathologies of the retina that lead to photoreceptors death in the macular region. This structural change is visible on the fundus image, which shows abnormal retinal pigmentation on the foveal region. Thus, the position of the fovea cannot be identified visually anymore, as opposed to healthy eyes (Fig. 1B).

Furthermore, since patients with CFL lose the ability to fixate with their fovea, it is impossible to locate it with a simple and quick fixation exam (Tarita-Nistor et al., 2011, 2017). Therefore, in the presence of maculopathies, the foveal region becomes impossible to locate with simple, straightforward techniques. This is especially problematic for patients with CFL, for whom identifying the exact location of the fovea is crucial to establish an accurate functional diagnosis.

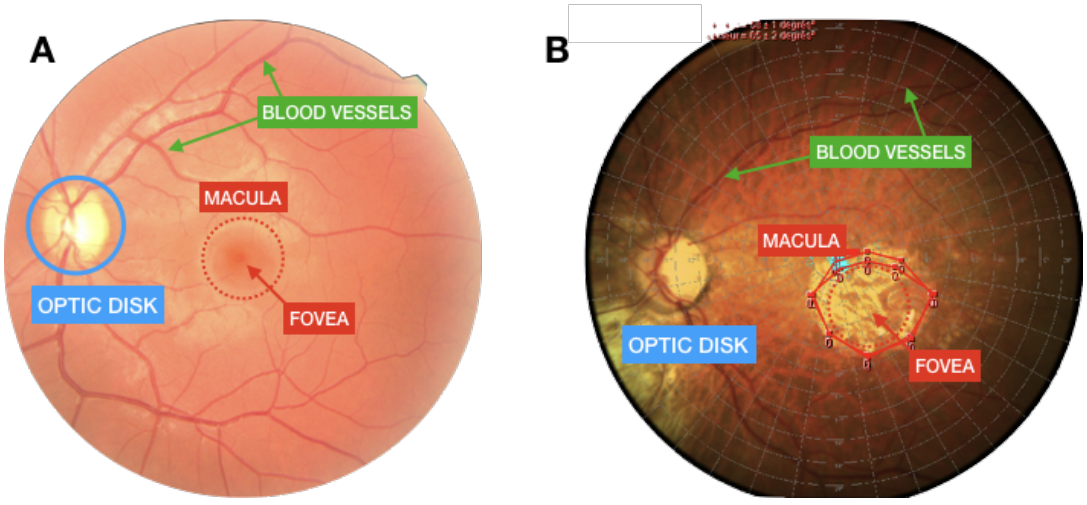


Figure 1: Annotated fundus images showing: the optic disk, the macula, the fovea and the vascular structure of the retina.

Simplest solution: normative anatomical measures (NAM) To estimate the position of the fovea on pathological fundus images, the most frequently used method is to rely on

normative anatomical measures (Gomes et al., 2009; Calabrèse et al., 2011; Ahuja et al., 2013; Vullings and Verghese, 2021). According to the literature, the mean foveal distance temporal to the middle of the optic disk is 15.5 ± 1.1 degrees horizontally and –1.5 ± 0.9 degrees vertically (Fig. 2A) (Rohrschneider, 2004; Timberlake et al., 2005; Reinhard et al., 2007; Tarita-Nistor et al., 2008). However using such normative measures presents several limitations. First, these measures represent a population average while the actual position of the fovea can vary dramatically from one individual to the next (Nair et al., 2021). Second, these normative measures were estimated under "standard" testing conditions (i.e. primary position of the head and central fixation stimulus). In the presence of a central scotoma however, fixating on a specific target requires to use eccentric vision. In the case of a large scotoma, the eccentricity required to fixate is so large that individuals may have to tilt their head and/or gaze to fixate. Therefore, they won't be able to maintain their head in primary position, which will have a significant impact on the relative position of the different anatomical structures of the eye (i.e., fovea, optic disk and blood vessels). In addition to eccentric vision, abnormal ocular torsion, which has been observed in a number of strabismus conditions (Kang et al., 2020), can also amplify this phenomenon. For instance, the distance between the optic disk and the fovea are highly dependent on the eye and head position, as discussed in Rohrschneider (2004). This dependency relationship is illustrated in Fig. 2 where the same healthy eye performed a fixation task under different conditions. In Fig. 2A, the eye was tested under "standard" conditions, with the head in primary position while fixating at a central cross. In this case, the macula is located at the center of the image and normative estimates of the fovea position matches perfectly its actual position (represented by the fixation cross). In Fig. 2B however, the subject tilted her head to the right to mimic a pathological behavior. This rotation causes greater variation in the vertical distance between the optic disk and the fovea. In this case, the normative estimates do not allow to locate the fovea accurately. Finally in Fig. 2C, the head tilt was also combined with the presentation of an eccentric fixation cross (5 ° to the right). Such shift does not allow to capture the optic disk entirely on the image, making this approach, based on the optic disk position, useless. Despite these strong limitations, this method remains currently the best one available to estimate the position of the fovea on color fundus images where the morphology of the macular region has severely changed.

1.2 Image Processing Techniques

The field of image processing has been broadly active to develop new methods to estimate the location of the fovea in normal color fundus images. Two main types of image processing methods have been developed: anatomical structures-based methods and deep learning methods.

The first one is based on manually built features, derived from the visible anatomical structures of the human retina. The second one exploits the feature extraction power of convolutional neural networks, implicitly extracting and modeling the visible anatomical structures.

Anatomical structures-based methods The main visible anatomical structures in a fundus image of a healthy subject are the optic disk, the blood vessels and the macula. These structures can be used as landmarks to locate the macular region, and consequently the fovea.

While the simplest method is to detect the fovea as the center of the darkest circular region on the whole fundus image (Sinthanayothin et al., 1999; Singh et al., 2008), most approaches work in two stages: first, (i) the estimation of a region of interest (ROI) that most likely contains the fovea, followed by (ii) the precise localization of the fovea within this ROI, using color intensity information. These more complex approaches use either blood vessels or optic disk information to estimate an accurate ROI. With the blood vessels-related approach, the ROI is estimated by

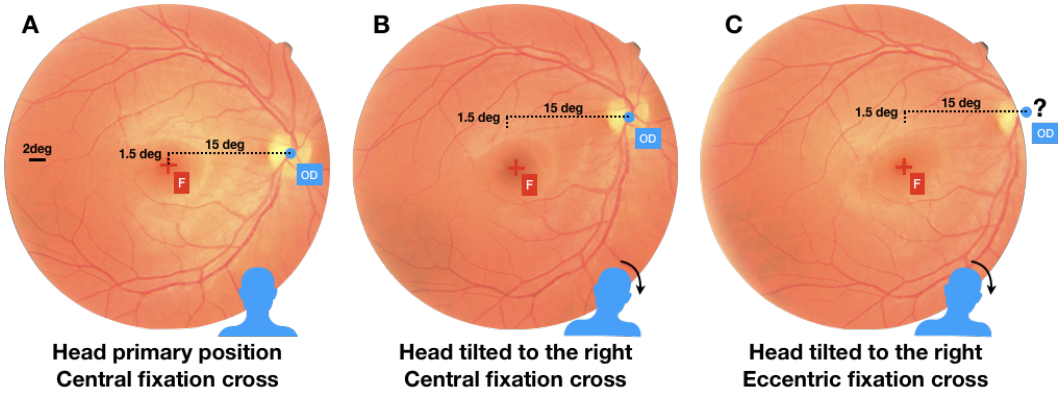


Figure 2: In some cases, measures described by Rohrschneider (2004) may not be used efficiently to locate the fovea, either because relative positions are not accurate (Fig. 2B). In other extreme conditions, the optic nerve may not be entirely visible on the image (Fig. 2C). Another side effect may be that the image quality (e.g., contrast) may not be optimal because the acquisition process does not operate in the normal conditions.

detecting the main vessels, fitting a parabola to their trajectory and defining the ROI at a given distance of the parabolas's vertex along the parabola's main axis (Li and Chutatape, 2004). With the optic disk-based approach, the ROI is estimated by detecting the center of the optic disk and considering a specific optic disk-fovea distance (Sagar et al., 2007; Sekhar et al., 2008).

Despite their efficiency with healthy fundus images, these methods, that rely on color intensity information within the macular area, cannot be applied as is to cases where the morphology of the macula has severely changed. However, a similar approach could be considered suitable if developed using only anatomical features that are not comprised within the macular area (i.e. the blood vessels and optic disk).

Deep learning methods As opposed to the anatomical structures methods that work on a case-by-case basis, currently developed deep learning methods constitute a powerful alternative as they build relevant and representative features from large amounts of data. Their application to computer vision in biomedical images have raised much attention lately, especially in the context of fundus images : amongst other tasks, they have been used for lesion, vessel, optic disk and fovea detection (Changlu et al., 2021; Tan et al., 2017; Al-Bander et al., 2018; An et al., 2020). Refer to see Li et al. (2021) for a review.

Overall, two different deep learning-based approaches can be considered : one considers fovea localization as a segmentation task, the other as a regression task.

In the segmentation approach, each pixel of the fundus image is classified by a convolutional neural network (CNN) into either "fovea" or "non-fovea" categories. Single CNNs have been applied successfully to segment simultaneously the fovea and optic disk (Kamble et al., 2020) or the fovea, the optic disk and the blood vessels (Tan et al., 2017), implicitly learning a prior on their relative positions. Following the same logic, a more hierarchical "coarse-to-fine" approach was also developed, using a first single CNN to extract two ROIs containing respectively the fovea and the optic disk, then followed by a second CNN to locate them precisely inside these ROIs (Al-Bander et al., 2018).

One can instead treat fovea localization as a regression task. In this case, the ouput of the neural network represents the coordinates of the fovea on the image (Xie et al., 2021).

In cases that interest us, (i.e., fundus images with non-visible fovea), no local feature makes the fovea region distinct from the rest of the fundus. Therefore, the segmentation approach is not well-suited. The regression approach however, seems more appropriate since the regression network is built to make predictions using all the features it extracts, even the ones relative to retinal regions away from the fovea. Yet, this method has always been trained and used with healthy fundus images, where the macular region could be used as a significant feature by the network. Applying deep learning to our problem would either require data from eyes with maculopathies, where the fovea annotation would be prone to errors, or creating a dataset with simulated lesions hiding the fovea.

1.3 Challenges and Goal

Despite the existence of efficient methods for fovea localization, to our knowledge none of them have been specifically designed and thoroughly tested on data where the fovea is not visible, and in particular on AMD data. Indeed, AMD images are often altered in ways that make these methods unusable : the macula is no longer visible, the optic disk may be hidden, and some vessels are hidden by the lesions. AMD images can also have very different appearances, or have a low image contrast.

These alterations have two consequences to take into account when specifically designing for AMD images. First, they make it harder to find robust landmarks to base the fovea localization on. Second, the macular lesions make precisely annotating the fovea in AMD images prone to errors, as the fovea cannot be visually located. Therefore, precisely annotated data sets of AMD images cannot realistically exist.

The solution we choose is to use the blood vessels, which seem to be the most robust landmarks to these observed image alterations: contrary to the macula or the optic disk, vessels are visible in nearly every fundus image, even in AMD cases. As the vascular structure does not contain any direct visual information about the fovea location, interestingly, we can use healthy retinas where the position of the fovea can be clearly identified in the fundus images, serving us as a precise ground truth.

However, locating the fovea from the vessels of a single fundus image remains challenging, as the retinal blood vessels appear with a large range of variability (Fig. 3). The first part of this variability concerns the natural, inter-individual differences. Two images from two different eyes can have very dissimilar looking vascular patterns, with no common landmark (Fig. 3A,B): these patterns are indeed unique to each individual, so much so that they are commonly used as a biometric characteristic (Semerád and Drahanský, 2020). Furthermore, as noted above (Fig. 2), a non-primary head position or an eccentric fixation during the acquisition will change the apparent orientation and position of the vessels. The second part of this variability comes, in our case, from maculopathies which will furthermore modify the vascular structure of a retina by destroying part of the vessels and deforming the retina (Fig. 3C,D).

The goal of the present work is thus to propose a new method to predict the fovea position, based only on the vessels, and insensitive to the observed variability between the vascular structures in the fundus images. To tackle this variability problem, we will build a statistical representation of the expected vascular structure in a fundus image, from a large data set, which will then be used to locate the fovea in new fundus images.

1.4 Paper Outline

This paper is organized as follows. In Sec. 2, we present the mathematical description of our registration-based fovea prediction (RBFL) method. In Sec. 3 we will analyse the performance

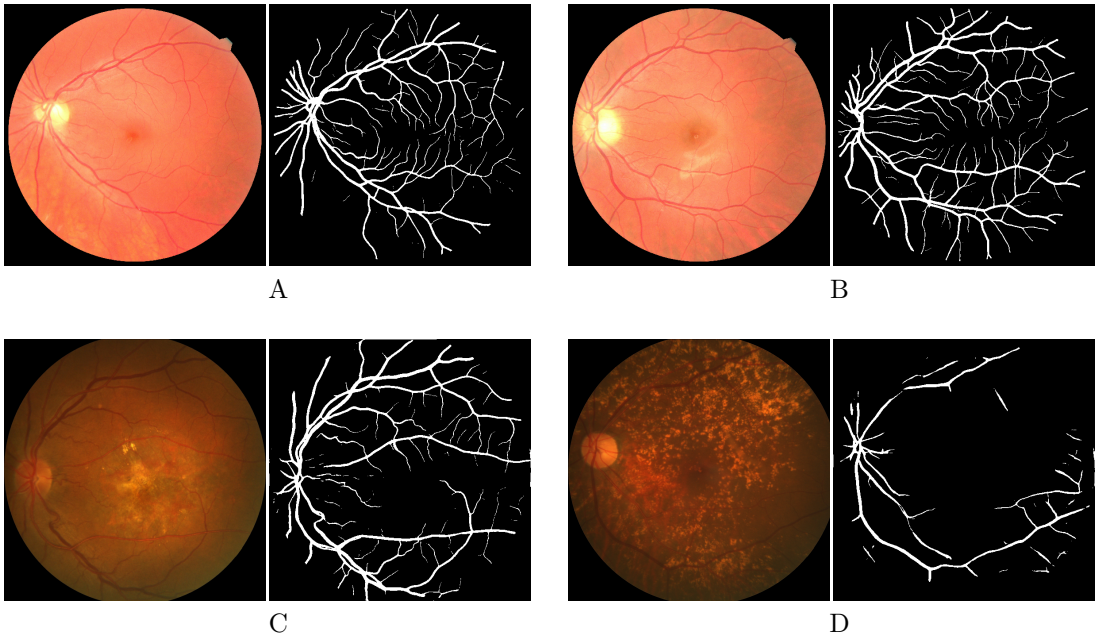


Figure 3: Illustration of the variability of visible vessels encountered in fundus images. Fundus images of four subjects, and the respective visible vessels, are represented. Fig. 3A and Fig. 3B are well-illuminated images from healthy subjects, and show the natural inter-individual variance in vascular structures. Fig. 3C,D come from AMD patients: the lesions hide part of the vessels.

of our approach on both healthy retinas (where all vessels are visible) and simulated and real pathological situations (where some vessels are partially hidden or not detected). In Sec. 4, we notably discuss other methods and ideas considered for this problem. In Sec. 5, we summarize our main findings and highlight the most interesting avenue for further study.

2 Registration-Based Fovea Localization Method (RBFL)

2.1 General Principle

Figure 4 illustrates the general principle of the proposed approach. Our goal is to predict the position of the fovea from vessels information only as commented in section Sec. 1.3. Given the high variability of vessel maps, predicting the fovea position from a single map appears to be a very challenging problem since there is *a priori* no robust landmark that could be used.

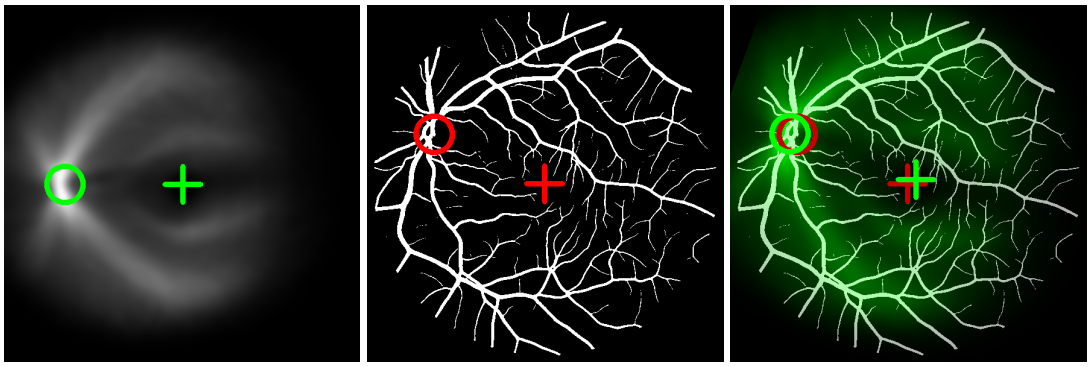
To overcome this difficulty, given a vessel map v, we propose to compare it with statistical representations of an ensemble of vessel maps, on which the fovea position is known. By doing so, we can predict a fovea position for the vessel map we consider. More precisely, by comparison we mean aligning the vessel map with a statistical representation of vessels, i.e., solve a registration problem. Typically, let us assume that an average density of vessels $\bar{V}$ is known, together with the corresponding reference fovea position (Fig. 4A). Given a vessel map v (Fig. 4B), we want to find the best transformation $\mathcal{T}$ which minimizes the difference between the vessel map v and the transformed average density (Fig. 4C), i.e., solve the following problem:

$$\inf_{\mathcal{T}} E(\mathcal{T}), \tag{1}$$

where

$$E(\mathcal{T}) = E_V(v, \mathcal{T}(\bar{V})).$$

Thanks to this registration, the prediction of the fovea position in v is defined by the average fovea position in $\bar{V}$ (Fig. 4C).



A B C

Figure 4: Illustration of the registration-based fovea localization method considering only vessel spatial distribution. Fig. 4A Vessel density map obtained a set of realigned vessel maps, with the reference fovea and optic disk position marked by a cross and a circle respectively, Fig. 4B Vessel map v in which the fovea has to be located. Here we assume that the ground truth is available allowing error estimates. Red cross indicates the true position of the fovea. Fig. 4C Vessel density map in green color superimposed on the vessels map after registration. The fovea position of the vessel density map (green cross) serves as an estimate for the fovea position in v. When ground truth is available (red cross), error can be estimated (distance between green and red crosses).

As shown later, our method will use another characterization of the vessels, namely their directions. Our method will consist in solving (1) with

$$E(\mathcal{T}) = E_V(v, \mathcal{T}(\bar{V})) + \eta E_D(d, \mathcal{T}(\bar{D})), \tag{2}$$

where η is a parameter to balance the two terms, d represents the vessels directions in v, and $\bar{D}$ the average directions.

2.2 Statistical Representation

The statistical representation will be built on a part of the REFUGE data set (Fu et al., 2019). The REFUGE data set contains 1200 eye fundus images with a manually annotated fovea location. Out of these, 120 are from eye diagnosed with glaucoma. We keep 840 non-glaucoma images as a training data set. For the rest of this paper, we refer to this training data set as D_{train}.

2.2.1 Average Density ($\bar{V}$)

The first characterisation consists in computing an average vessel density map $\bar{V}$ showing, for each position (x, y), the likelihood to have a vessel passing through (x, y).

For one image: Let us first consider the case of one image. Given a retina fundus image u, we estimate a vessel map v, using the retinal vessel segmentation network SA-UNet (Changlu et al., 2021). This is a modified version of U-Net, a type of convolutinnal neural networks designed for biomedical image segmentation, with added dropout blocks and a spatial attention module (see Appendix, Sec. A.3 for more details). We used the weights provided by the authors, pre-trained on the DRIVE (Staal et al., 2004) dataset. It outputs a binary map indicating the presence or not of a vessel at each location (x, y).

For a set of images: Next, given a set of images u_i from the training set D_{train}, we can estimate an average density as follows (Fig. 5). Vessels maps v_i are first estimated and then aligned: For each u_i, since we know the optic disk position (x_i^{OD}, y_i^{OD}) and the fovea position (x_i^F, y_i^F), we compute the transformation that maps these two positions to a common reference position for both the optic disk $(x_{\text{ref}}^{OD}, y_{\text{ref}}^{OD})$ and the fovea $(x_{\text{ref}}^F, y_{\text{ref}}^F)$. More precisely, for each image, the problem consists in computing the exact similarity transformation (the simplest class of transformation for the perfect alignment of two pairs of points), denoted by $\mathcal{T}_i$, such that:

$$\mathcal{T}_i(x_i^{OD}, y_i^{OD}) = (x_{\text{ref}}^{OD}, y_{\text{ref}}^{OD}) \quad \text{and} \quad \mathcal{T}_i(x_i^F, y_i^F) = (x_{\text{ref}}^F, y_{\text{ref}}^F). \tag{3}$$

After applying the transformations to each vessel map, we compute a first density estimate $\bar{V}_0$ by averaging the warped vessel maps:

$$\bar{V}_0 = \frac{1}{|D_{\text{train}}|} \sum_{u_i \in D_{\text{train}}} \mathcal{T}_i(v_i),$$

where $|D_{\text{train}}|$ is the cardinal of D_{train}. The final average density $\bar{V}$ is then obtained after applying to $\bar{V}_0$ a smoothing operator $\mathcal{S}$ (see Remark 1):

$$\bar{V} = \mathcal{S}(\bar{V}_0). \tag{4}$$

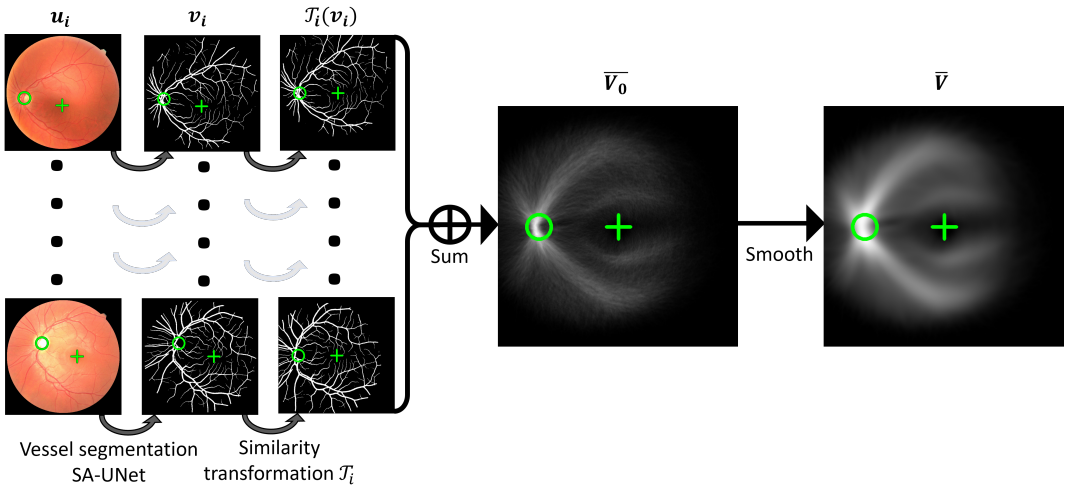


Figure 5: Overview of vessel map estimation: Vessels maps from the training set are realigned thanks to the optic disk and fovea positions to produce the average map which is then smoothed.

Remark 1 ***About the smoothing operator $\mathcal{S}$ in*** (4). *Two options were considered. A first option has been to perform an isotropic smoothing by a convolution with a Gaussian kernel k_σ : $\bar{V}(x) = (k_\sigma * \bar{V}_0)(x)$. The smoothing level is controlled by the parameter σ. The classical problem with this method is that it blurs everything and does not preserves the structures. A better option used here is to apply an anisotropic smoothing method to better preserve the bundle of vessels. To do so we used a partial differential equation (PDE) approach where the smoothing operator takes into account an estimate of the direction of vessels to smooth in that direction only. This is done by considering the notion of oriented Laplacians (Tschumperle and Deriche, 2005). Introducing the notion of time evolution ($\bar{V} = \bar{V}(t, x, y)$) and starting from the initial condition $\bar{V}(t = 0, x, y) = \bar{V}_0(x, y)$, the PDE to solve is:*

$$\frac{\partial \bar{V}}{\partial t} = trace(\mathbf{TH}), \tag{5}$$

where $\mathbf{H}$ represents the Hessian of $\bar{V}$, and $\mathbf{T}$ the diffusion tensor field which defines in which direction to smooth. In our case, this diffusion tensor field will be estimated from the vessel maps (see more details in the following section, Remark 2). In this case, the time plays the role of the smoothing level.

2.2.2 Average Directions ($\bar{D}$)

The second characterisation consists in computing an average direction map $\bar{D}$ showing, for each position (x, y), the most likely direction of a vessel passing through (x, y). To build this map, the idea is again to exploit the training set coupled with the notion of structure tensors, which is a classical notion in the domain of anisotropic diffusion (Weickert, 1998; Aubert and Kornprobst, 2006).

The intuition behind the idea of tensors for representing a direction distribution is explained in the Appendix (Sec. A.2). In a nutshell, given a vector $w = (w_x, w_y)$, we can define the following tensor:

$$\mathbf{T}[w] = ww^T = \begin{pmatrix} {w_x}^2 & w_x\, w_y \\ w_x\, w_y & {w_y}^2 \end{pmatrix},$$

which is positive semidefinite. Its eigenvalues are $\lambda_1 = |w|^2$, and $\lambda_2 = 0$ and there exists an orthonormal basis of eigenvectors e_1 parallel to w and e_2 orthogonal to w. In a way, $\mathbf{T}[w]$ corresponds to a matrix representation of w where the notion of orientation has disappeared (since e_i and $-e_i$ are both eigenvectors). As a whole, this representation brings a key advantage when we need to sum directions (e.g., to compute an average direction) since, instead of averaging vectors which is an ill-defined problem, one can average their associated tensors, and from this average tensors one can estimate the average direction.

For one image: Let us first consider the case of one image u. Given its vessel map v, let us define its associated structure tensor field $\mathbf{s}$ containing the local vessel direction information. To diminish eventual problems due to noise, v is first convolved with a Gaussian kernel k_σ:

$$v_\sigma = k_\sigma * v. \tag{6}$$

Then, we need to estimate vectors that are aligned with the vessels. To do so, we use the gradient ∇v_σ which is, by definition, locally orthogonal to the vessels, and just consider the orthogonal vectors:

$$t = (\nabla v_\sigma)^\perp. \tag{7}$$

This local information is then averaged by convolving componentwise $\mathbf{T}[t] = tt^T$ with a Gaussian kernel k_ρ. The result is the structure tensor field $\mathbf{s}$:

$$\mathbf{s} = k_\rho * \mathbf{T}[t]. \tag{8}$$

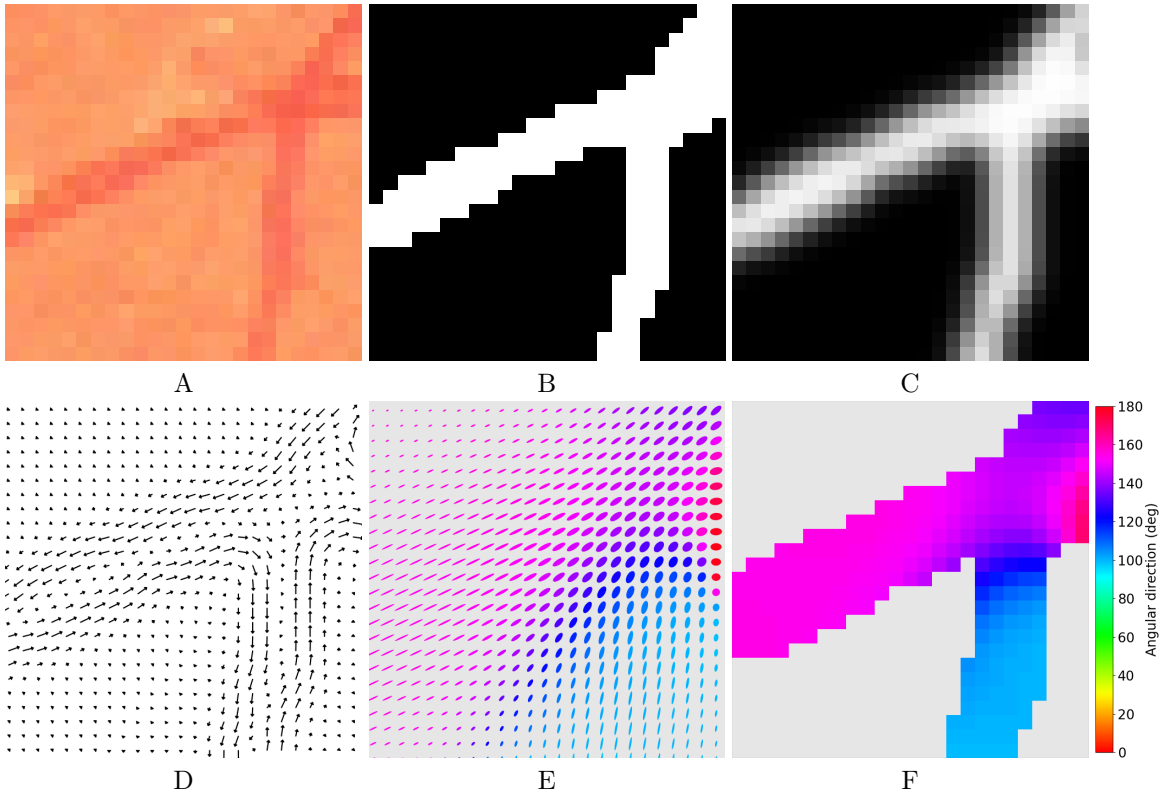


Figure 6: Illustration of the different steps of direction map estimation. Only a small patch of each image or map is represented. Fig. 6A: Input fundus image u. Fig. 6B: Vessel map v. Fig. 6C: Smoothed vessel map v_σ Fig. 6D: Vector field orthogonal to the gradient of v_σ, $t = (\nabla v_\sigma)^\perp$. Each vector is represented as an oriented arrow. Fig. 6E: Structure tensor field $\mathbf{s}$.. Fig. 6F: Vessel direction map d.

Finally, from $\mathbf{s}$ we derive the direction of the vessels at each position (x, y), denoted by $d(x, y)$. This direction at (x, y) is given by the direction of $e_1(x, y)$, the eigenvector of largest eigenvalue of the tensor $\mathbf{s}(x, y)$. Once normalized, this eigenvector can be rewritten as:

$$e_1(x, y) = \begin{pmatrix} e_{1,x} \\ e_{1,y} \end{pmatrix} = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix},$$

with $\alpha \in [0, \pi[$. So we can define $d(x, y)$ by:

$$d(x, y) = \arccos(e_{1,x}). \tag{9}$$

By construction, the values of $\bar{D}$ are between 0 and π, thus giving the vessel directions without any information of orientation.

The whole process is illustrated on a small fundus image patch in Fig. 6. The final vessel directions are only represented within the vessels in Fig. 6F, as outside them the computed directions can neither be interpreted as vessel directions, nor used by this method as we will see further (Sec. 2.3.1).

The properties of the structure tensor field are illustrated in Fig. 7. Each structure tensor is represented as an ellipse whose axis directions and sizes represent its eigenvectors and eigenvalues. The positive semi-definite matrix $\mathbf{s}(x, y)$ has orthonormal eigenvectors (e_1, e_2) whose corresponding eigenvalues are $\lambda_1 \geqslant \lambda_2 \geqslant 0$. They describe the average contrast in the eigendirections within a neighborhood of size $O(\rho)$. The vector e_1 indicates the orientation minimizing the gray-value fluctuations, i.e., the direction of the vessels. Locally constant areas, in this case non-vessel areas (Fig. 7F), are characterized by $\lambda_1 \approx \lambda_2 \approx 0$. Structures with multi-directional variations, e.g. within the optic disk (Fig. 7E), crossings (Fig. 7F) and bifurcations (Fig. 7G) are characterized by $\lambda_1 \approx \lambda_2$. Linelike structure, such as linear portions of vessels (Fig. 7F) are characterised by $\lambda_1 \gg \lambda_2 \approx 0$. Fig. 7B shows the first map we derive from this tensor field: the saliency map ξ, which we will introduce later (Eq. (13)),

For a set of images: Next, given a set of images u_i from the training set D_{train}, we can estimate an average direction map as follows. For each fundus image u_i from the training set D_{train}, we start from the realigned vessel maps, as defined for the average density image estimation. Let us denote them by $\tilde{v}_i = \mathcal{T}_i(v_i)$, where $\mathcal{T}_i$ is defined as in (3).

For each $\tilde{v}_i$, we estimate is associated structure tensor $\mathbf{s}_i$ as defined by (6)–(8). The next step consists in accumulating direction information across the training set, by simply summing the tensor fields pixel-wise:

$$\forall (x, y), \quad \bar{\mathbf{S}}(x, y) = \sum_{u_i \in D_{\text{train}}} \mathbf{s}_i(x, y). \tag{10}$$

Result is shown in Fig. 8A. Finally, from $\bar{\mathbf{S}}$ we derive the average direction of vessels at each position (x, y) as in (9). Result is shown in Fig. 8B.

Remark 2 ***Complementary information related to Remark*** 1. *Coming back to the oriented Lapacian method described in* (5)*, we can now state that the tensor field* $\mathbf{T}$ *used to smooth the density is* $\mathbf{T} = \bar{\mathbf{S}}$ *as defined in* (10)*. This means that the oriented Laplacian uses the full information about direction distribution, not just the main direction.*

2.3 Registation Criteria

2.3.1 Energy Definition

Let us consider a fundus image u for which the fovea position needs to be predicted. After estimating the associated vessels and directions (v and d), and given the average density and directions $\bar{V}$ and $\bar{D}$ (as defined in Sections 2.2.1 and 2.2.2), the problem is to minimize the energy (Eq. (1)) with respect to the transformation $\mathcal{T}$, where $E(\mathcal{T})$ is defined by:

$$E(\mathcal{T}) = E_V(v, \mathcal{T}(\bar{V})) + \eta E_D(d, \mathcal{T}(\bar{D})).$$

In this section, we describe the type of transformation of $\mathcal{T}$ we look for, and define the two terms E_V and E_D of this equation.

Class of transformation: Here we assume that $\mathcal{T}$ is a similarity transformation, accounting for translation, rotation and uniform scaling. We decompose $\mathcal{T}$ into four parameters:

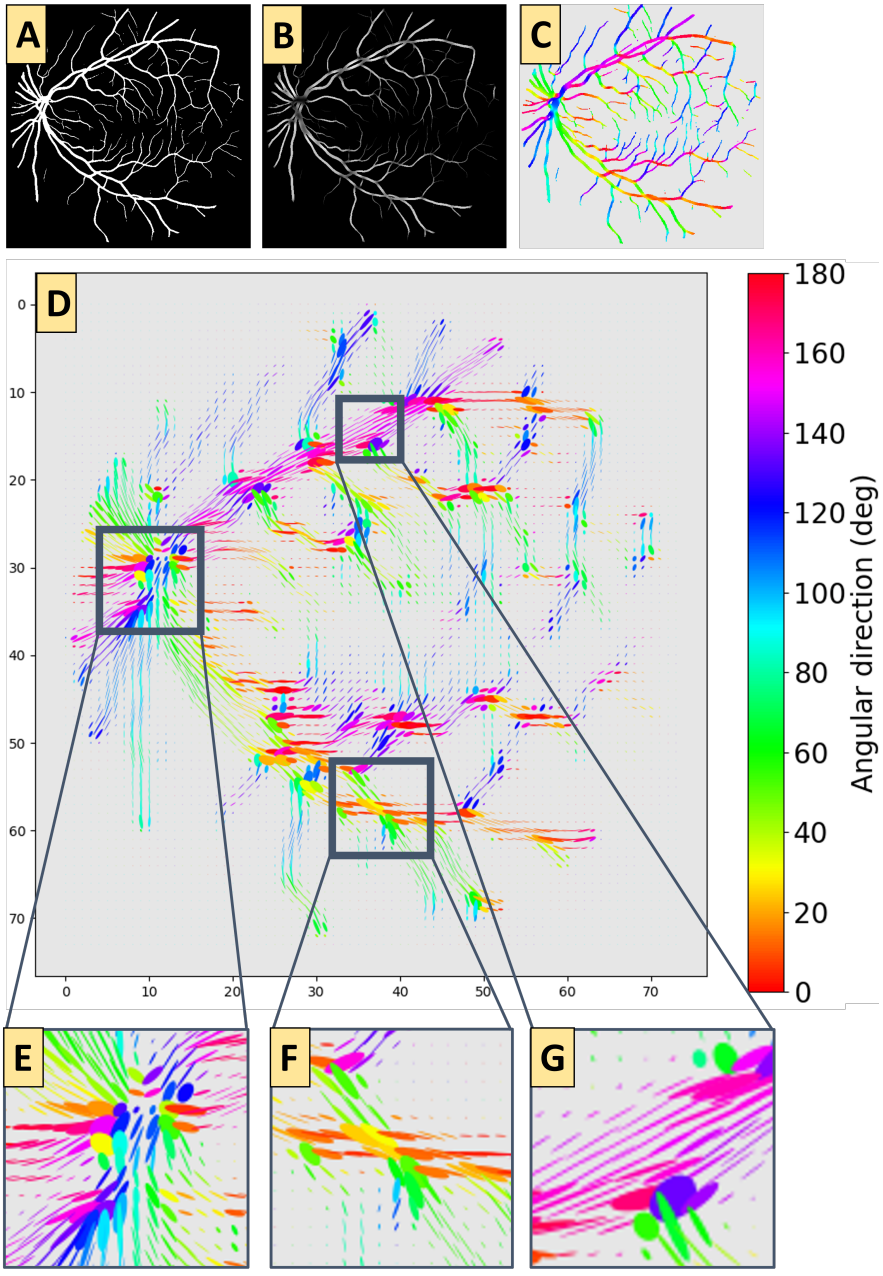
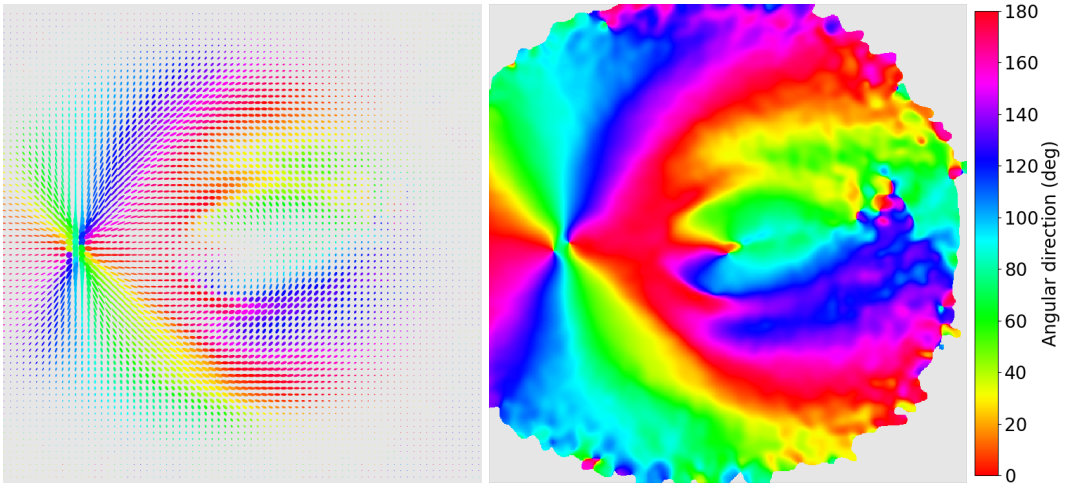


Figure 7: Illustration of the structure tensor field $\mathbf{s}$ for a single fundus image. Fig. 7A: Input vessel segmentation image v. Fig. 7B: The saliency map ξ. Fig. 7C: The vessel direction map d. Fig. 7D: The tensor field $\mathbf{s}$. Fig. 7E,F,G focus on specific parts of the tensor field : the optic disk (Fig. 7E), and straight vessels, bifurcations and crossings (Fig. 7F,G).



A B

Figure 8: Results of the average vessel direction map estimation. Fig. 8A The average structure tensor field $\bar{\mathbf{S}}$. Fig. 8B The resulting average vessel direction map $\bar{D}$.

- $t_x(\mathcal{T})$: translation along the horizontal axis,
- $t_y(\mathcal{T})$: translation along the vertical axis,
- $\theta(\mathcal{T})$: rotation around the reference fovea location $(x^F_{\text{ref}}, y^F_{\text{ref}})$,
- $s(\mathcal{T})$: uniform scaling around the reference fovea location $(x^F_{\text{ref}}, y^F_{\text{ref}})$.

Density term: The term E_V must penalize an incorrect alignment of $\bar{V}$ w.r.t. v. If they are correctly aligned, areas of high density in $\mathcal{T}(\bar{V})$ should correspond to areas containing vessels in v, and conversely areas of low density in $\mathcal{T}(\bar{V})$ should mostly correspond to empty areas in v. To ensure this, we choose a weighted mean squared error:

$$E_V(\mathcal{T}) = \sum_{x,y} \frac{\lambda_{\mathcal{T}}(x,y)}{C_{\mathcal{T}}} \left(v(x,y) - \mathcal{T}(\bar{V})(x,y)\right)^2, \tag{11}$$

where $\lambda_{\mathcal{T}}(x,y)$ is a weight used to give more importance to specific retinal regions (see Sec. 2.3.2 for its definition), $C_{\mathcal{T}}$ is a normalization coefficient ($C_{\mathcal{T}} = \sum_{x,y} \lambda_{\mathcal{T}}(x,y)$). Note this normalization coefficient is required since $\sum_{x,y} \lambda_{\mathcal{T}}(x,y)$ depends on $\mathcal{T}$ by the definition of $\lambda_{\mathcal{T}}$.

Direction term: The term E_D must penalize an incorrect alignment of $\bar{D}$ w.r.t. d. They are correctly aligned at a point (x,y) if the directions $d(x,y)$ and $\mathcal{T}(\bar{D}) + \theta_{\mathcal{T}}$ are close, modulo π. Note that the reason we need to add $\theta_{\mathcal{T}}$ is illustrated in Fig. 9. This is simply because we apply rotations to functions representing orientations, so that the values themselves also need to account also for the rotation applied.

So, based on directions only, one could define the direction term E_D by:

$$E_D(\mathcal{T}) = \sum_{x,y} \frac{\lambda_{\mathcal{T}}(x,y)}{C_{\mathcal{T}}} \sin^2\left(d(x,y) - (\mathcal{T}(\bar{D})(x,y) + \theta(\mathcal{T}))\right). \tag{12}$$

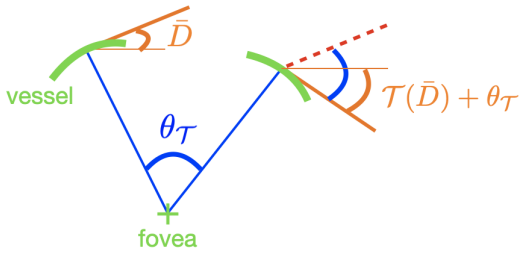


Figure 9: Illustration of the effect of a similarity transform onto the average direction map.

where $\frac{\lambda_{\mathcal{T}}}{C_{\mathcal{T}}}$ is the same weight used in the density term, and defined in 2.3.2. Since function $\sin^2$ is π-periodic, this energy is minimal if $d(x,y) \cong \mathcal{T}(\bar{D}) + \theta_{\mathcal{T}}$ modulo π, and maximal if $d(x,y) \cong \mathcal{T}(\bar{D}) + \theta_{\mathcal{T}} + \frac{\pi}{2}$ modulo π.

However, expression (12) poses problems in situation where, e.g., there are no vessels in v at some position (x,y). In that case, v is locally constant, and there is not interpretation of d as a vessel direction. More generally, a direction $d(x,y)$ has no interpretation in areas with multi-directional variations of v, which includes vessel crossings or bifurcations. To account for these cases, we propose to add an additional weight based on the saliency map ξ:

$$\xi(x,y) = v(x,y)(\lambda_1(x,y) - \lambda_2(x,y)) \tag{13}$$

where $\lambda_1(x,y)$ and $\lambda_2(x,y)$ are respectively the higher and lower eigenvalue of the structure tensor $\mathbf{s}(x,y)$. This saliency map is minimal and close to 0 outside the vessels in v, or when $\lambda_1(x,y) \approx \lambda_2(x,y)$, i.e. in areas of multi-directional variations. It is maximal when $\lambda_1(x,y) \gg \lambda_2(x,y) \approx 0$, along linear portions of vessels, i.e. where we actually want to compare directions. As a consequence, the direction term E_D is modified as:

$$E_D(\mathcal{T}) = \sum_{x,y} \frac{\lambda_{\mathcal{T}}(x,y)}{C_{\mathcal{T}}} \frac{\xi(x,y)}{\max(\xi)} \sin^2\left(d(x,y) - (\mathcal{T}(\bar{D})(x,y) + \theta(\mathcal{T}))\right). \tag{14}$$

2.3.2 Definition of the Region-Based Weight ($\lambda_{\mathcal{T}}$)

The region-based weight $\lambda_{\mathcal{T}}$, which appears in both E_V (11) and E_D (14) is used to bring more or less relative importance to three regions:

(i) The first region is related to an anatomical property. The macular region surrounding the fovea is poor in terms of visible vessels: macular vessels are very dense but have a very small diameter, making them invisible in fundus photographs. This is thus a consistent observation across vessel images, and moreover, we are precisely in precisely predicting the fovea position, which belongs to this region. Therefore it is particularly important to give to this region $\bar{\omega}$ a higher weight w_{macula}.

(ii) The second region is related to the field of view of the images acquired. Given a retina fundus image u, one can define its domain of definition Ω; similarly, when average density and directions were estimated, one can define a reference domain of definition $\bar{\Omega}$. We choose $\bar{\Omega}$ as the smallest domain containing at least $\nu = 50\%$ of the domains $\mathcal{T}_i(\Omega_i)$, for all $u_i \in D_{\text{train}}$, where $\mathcal{T}_i$ is the transformation used to warp v_i onto $\bar{V}$ (Eq. (3)) and Ω_i is the domain of u_i.

The idea is that, when estimating the energy E, it only makes sense to consider pixels which belong to both Ω and $\mathcal{T}(\bar{\Omega})$, i.e., where data is available. In these regions outside the domain of definition, a weight w_{outside} will be assigned. Note that, for algorithmic reasons, this weight will depend on the current registration scale (see Sec. 2.4.1 for details).

(iii) The rest is given a general weight w_{general}.

This region-based information is then combined to define the region-based weight $\lambda_{\mathcal{T}}$ as illustrated in Fig. 10.

2.4 Algorithmic Details

2.4.1 Multiscale

In order to speed up the computations and make the registration more robust to high-frequency local minima, we derive our method to work in a multi-scale manner. To do so, we build Gaussian image pyramids of the input vessel map v and of $\bar{D}$:

$$\{v^{(m)}\}_{m=0,\dots,M-1}, \quad \text{and} \quad \{\bar{V}^{(m)}\}_{m=0,\dots,M-1},$$

so that the image dimensions (s_x, s_y) are reduced by two between each scale, where M is the number of scales, $m = 0$ being the finer scale (i.e., corresponds to the original size). Then the method consists in solving successively the optimization problem (15) across scales, i.e.,

$$\inf_{\mathcal{T}^{(m)}} E(\mathcal{T}^{(m)}), \tag{15}$$

from $m = N - 1$ to $m = 0$, where $\mathcal{T}^{(m)}$ is the transformation estimated at scale m. Given an estimation $\mathcal{T}^{(m)}$ at a scale $m > 0$, we use it to have an initial estimation the problem at scale $m - 1$. More precisely, given the estimation at a scale $m > 0$:

$$[t_x(\mathcal{T}^{(m)}), t_y(\mathcal{T}^{(m)}), \theta(\mathcal{T}^{(m)}), s(\mathcal{T}^{(m)})],$$

we use as an initial estimate for scale $m - 1$:

$$|2t_x(\mathcal{T}^{(m)}), 2t_y(\mathcal{T}^{(m)}), \theta(\mathcal{T}^{(m)}), s(\mathcal{T}^{(m)})].$$

We noted in Sec. 2.3.2, that in the region-based weight $\lambda_{\mathcal{T}}$, the outer value w_{outside}, i.e. the ponderation outside the domains of definition of the fundus image and average vessels representations, depends on the registration scale. We chose to make the value of w_{outside} decrease as the scale increases. Indeed, we noticed that a higher w_{outside} lead to a decreased rate of very large fovea prediction error, always outputing solutions not far away to the ground truth. However, we can achieve an overall greater precision when using a null value of w_{outside}. Therefore, we use a higher w_{outside} at the lowest scales, ensuring a good initialization for the final scale, at which point we can set w_{outside} to 0 to get the best precision possible.

2.4.2 Optimization Method, Parameters Initialization, Bounds and Order

To perform the optimization, Powell's method (Powell, 1964) is used. This method is classicaly used in multi-modal registration, as it is simple to implement and does not require computing derivatives. In a nutshell, this method consist in sequentially optimizing the function along 1D search-directions: in our case, we initially optimize over each parameter of the similarity

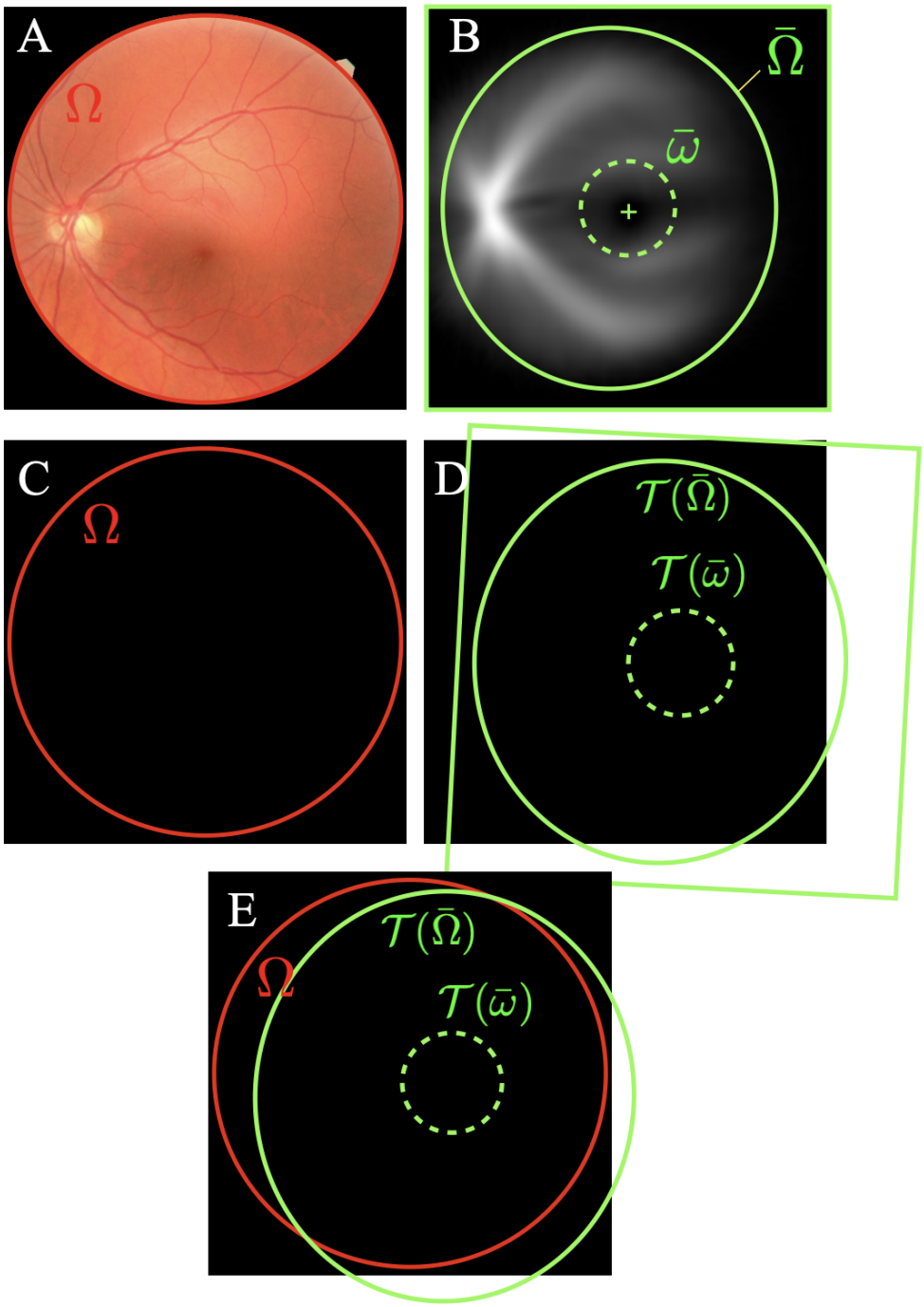


Figure 10: Illustration of the regions given importance in the region-based weight $\lambda_{\mathcal{T}}$, for a given transformation $\mathcal{T}$. The region-based weight is higher within the reference macular region $\mathcal{T}(\bar{\omega})$, and lower outside the intersection of the image domains $\mathcal{T}\left(\bar{\Omega}\right)$ and Ω.

transformation $\mathcal{T}$ presented in 2.3.1, one after the other. Note that the algorithm introduces new search-directions as combinations of these parameters.

Though it is a simple method, it requires choosing the initialization and bounds for the transformation parameters. It also requires to chose an initial order of parameters, in which the optimization will be performed. A poor choice of initialization, bounds or order can result in the optimization getting stuck in irrelevant local minima.

The chosen initialization and bounds are given in Tab. 1, and we detail thereafter how they were chosen. We also explain in which order the optimizations were performed.

Initialization. Starting at the coarser scale ($m = M - 1$, Sec. 2.4.1), we need to define an initial condition for the four parameters of the transform $\mathcal{T}^{(M-1)}$. Without human input, we did not find any simple and robust way to know whether the retina appears rotated, translated, or at a different scale from that of an average eye with a central fixation and primary head position. Therefore, to define the initial condition, here we simply assumed that the eye is of a standard size, and in a standard position with a central fixation, i.e.,

$$\mathcal{T}^{(M-1)} = (0, 0, 0, 1).$$

Bounds. Defining bounds is necessary to avoid Powell's method to stop in local minimia which would be not relevant. Bounds are also natural because the type of images we consider are constrained by the imaging acquisition itself. Indeed, the variability of transformations in retinal fundus images come from the head position and fixation position of the observer. So to estimate plausible bounds for the transformation parameters, we have used the data from the our testing set D_{test} for which we know that a variety of positions (for both head and fixation) were tested. Again, this is only to define some range of variations for the parameter. In practice, given all images u_i from D_{test} for which ground truth is known (i.e., fovea and optic disk position), we first estimated the transforms $\mathcal{T}_i$ such that:

$$\mathcal{T}_i((x^F_{\text{ref}}, y^F_{\text{ref}})) = (x^F_i, y^F_i) \quad \mathcal{T}_i((x^{OD}_{\text{ref}}, y^{OD}_{\text{ref}})) = (x^{OD}_i, y^{OD}_i).$$

Then, for each parameter $p_i \in \{t_{x\,i}, t_{y\,i}, \theta_i, s_i\}$, we select its bounds p_{low} and p_{high} such that:

$$p_{\text{low}} = \min_{u_i \in D_{\text{test}}} \beta p_i, \quad \text{and} \quad p_{\text{high}} = \max_{u_i \in D_{\text{test}}} p_i/\beta, \tag{16}$$

where β is a multiplication factor in $[0, 1]$ to let some margin for the bounds.

Order. Another decision to make when applying the Powell's method is the order in which we consider the different parameters during the optimization. In practice, there is no strict rule and this depends on the nature of the problem. Here the order has been determined by testing different solutions. The best option we found was to optimize with respect to transformation parameters in the following order:

$$t_y(\mathcal{T}) \rightarrow t_x(\mathcal{T}) \rightarrow \theta(\mathcal{T}) \rightarrow s(\mathcal{T}).$$

2.4.3 Enhancement of the Density

The average density $\bar{V}$ takes continuous values in $[0, 1]$ (see result in Fig. 5) while vessel maps v are binary in {0,1}. Since both are compared in the term E_V (11), we found some improvement when an enhancement was applied to $\bar{V}$. Our interpretation is that enhancing $\bar{V}$ makes it closer to a binary image and thus easier to compare with v.

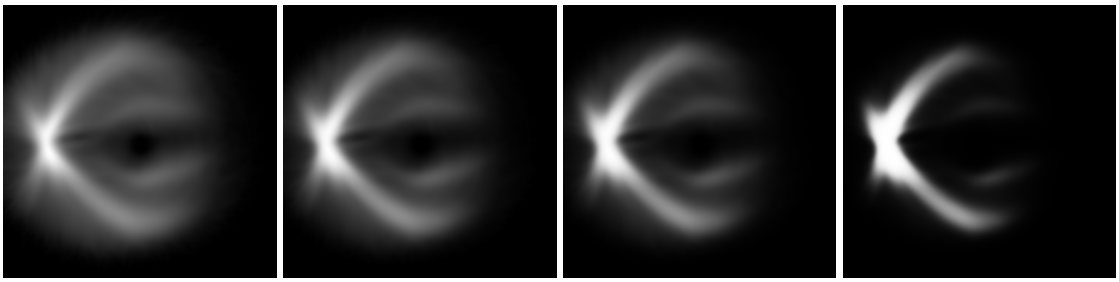
In practice, in the term E_V we use the enhanced version of $\bar{V}$, denoted by $\bar{V}_{\text{sig}}$, defined by applying a sigmoid function and a normalization:

$$\bar{V}_{\text{sig}} = \frac{\text{sig}(\bar{V}) - \min(\text{sig}(\bar{V}))}{\max(\text{sig}(\bar{V})) - \min(\text{sig}(\bar{V}))}, \tag{17}$$

with:

$$\text{sig}(s) = \frac{1}{1 + e^{\frac{-(2s-1)}{\mu}}}. \tag{18}$$

μ is a parameter to control the strength of the enhancement. The effect of this parameter on the average density is shown in Fig. 11.



(A) Initial $\bar{V}$ (B) $\mu = 0.45$ (C) $\mu = 0.3$ (D) $\mu = 0.15$

Figure 11: Effect of the sigmoid enhancement on the original average density $\bar{V}$, depending on the value of the sigmoid's parameter μ. Fig. 11A: The initial average vessel density map $\bar{V}$. Fig. 11B,C,D: The sigmoid-enhanced average vessel density maps, with respective sigmoid parameters $\mu_B = 0.45$, $\mu_C = 0.3$ and $\mu_D = 0.15$. In our method, we use $\mu = 0.3$, which gave the best results.

3 Results

In this section we analyse the performance or our approach (RBFL). Appendix A.1 presents the parameters we used to compute these results, and the computational performances of the method. Section 3.1 presents the testing set collected and used to evaluate the performances of the method. Sections 3.2–3.5 cover the main part of the analysis done with our dataset made of healthy retinas.[1] In Sec. 3.6, we show how our approach performs in the presence of scotomas, thanks to a parametric study where scotoma's size and shapes will hide partially vessels information. In Sec. 3.7, we propose a final result on a different data set, composed of real fundus images of retinas with AMD. Our approach will be evaluated with respect to ground truth. We will also compare the performances of the proposed RBFL to those of the classical solution using normative anatomical measures (NAM, Sec. 1.1, Fig. 2), i.e., locating the fovea 1.5 degrees inferior and 15.5 degrees temporal relatively to the manually annotated optic disk position.

[1] As a particular fixation location of head position has a different effect on the visible vessels depending on whether we consider a right or left eye, we only present the results for the right eyes. Note that the results for left eyes are similar.

3.1 Testing Set: Fundus Images on Healthy Subjects with Different Head Positions and Fixations

We collected[2] a dataset of 198 fundus images from 21 healthy eyes, coupled to fixation examinations for precise fovea annotation, with different head positions and fixation locations. 19 normally sighted subjects (11 females) were recruited to collect data on the microperimeter MP-3 (Nidek Inc.). Their ages range from 19 to 40. Two subjects had the two eyes tested, the rest were tested on a randomly selected eye.

The following protocol was adopted to collect the data for each eye :

(i) An off-center fixation location is randomly chosen among those presented in figure 12 (F_{Left}, F_{Right}, F_{Up} or F_{Down}).

(ii) The subject performs five fixation examinations with this fixation location with each of the five head positions presented in figure Fig. 13 ($H_{Primary}$, H_{Left}, H_{Right}, $H_{Frontward}$ and $H_{Backward}$). Every fixation examination is followed by a fundus photograph.

(iii) The subject performs a fixation examination with central fixation location $F_{Central}$ and primary head position $H_{Primary}$.

(iv) If the subject does not feel too inconvenienced by the bright flash of the six photographs taken, he is proposed to perform the four remaining fixation examinations and fundus photographs with central fixation $F_{Central}$ and head positions H_{Left}, H_{Right}, $H_{Frontward}$ and $H_{Backward}$.

(v) The axial length of the eye is measured with a non-contact optical device for ocular biometry (IOLMaster, Zeiss Instruments). The subjective refraction of the eye is measured and the visual acuity is tested.

The fixation location corresponds to the fovea location on each image. Furthermore, the optic disks centers are manually annotated. For the rest of this paper, we refer to this testing data set as D_{test}.

3.2 Global Performance

We present the general performances the RBFL method on the testing set D_{test}, compared to the errors of the NAM method. Errors are given in degrees of visual angle, as it is a more anatomically relevant unit. This can easily be computed because we know the pixel to degree ratio r of the microperimeter MP3: a pixel is equivalent to $r = 0.07904$ degrees of visual angle.

Figure 14A shows the distribution of the fovea localization errors of both RBFL and NAM, on the whole testing set, as violin plots. From NAM, our method reduced the mean fovea localization error by 59%: from 2.85 (SD 2.33) degrees for NAM, to 1.16 (SD 0.86) degrees for RBFL. It also reduces the median error by 54%, from 1.93 degrees for NAM to 0.89 degrees for RBFL.

From its construction, we expect NAM to perform at its best when considering only eyes in central fixation and primary head position. Figure 14B shows the same distributions of errors for both methods, on the central fixation-primary head position subset of D_{test}. On this data, RBFL still outperforms NAM: it reduces the mean error by 48%, from 1.32 (SD 0.76) for NAM to 0.69 (SD 0.34) for RBFL, and the median error by 57%, from 1.28 for NAM to 0.55 for RBFL.

Another way to view these results is to consider the cumulative distributions. Figure 14C shows the cumulative distributions of fovea localization errors for both RBFL and NAM, on the

[2]The data was collected by Séverine Dours at Clinique Monticelli, Marseille. Courtesy Frédéric Matonti.

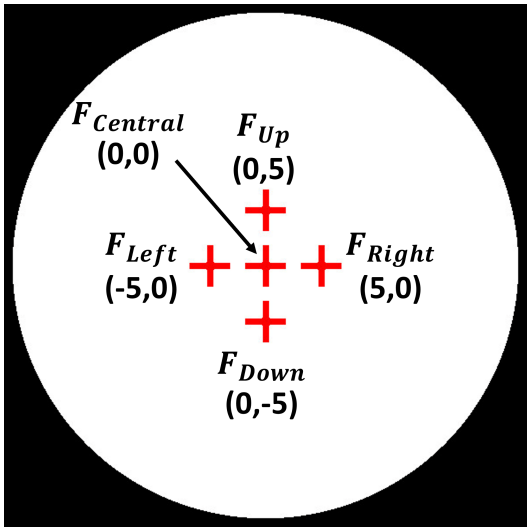


Figure 12: Illustration of the five fixation points presented to the subjects. The white part represents the field of view of the fundus camera used, and the red crosses the fixation points. The fixation points can be in the center of the camera's field of view ($F_{Central} = (0, 0)$), or five degrees of visual angle away from it, in the horizontal or vertical direction.2

Primary position ($H_{Primary}$)

Left position (H_{Left})

Right position (H_{Right})

Backward position ($H_{Backward}$)

Frontward position ($H_{Frontward}$)

Figure 13: Illustration of the five head positions taken by the subjects.

whole testing set D_{test}, and Fig. 14D shows it on the central fixation-primary head position subset of D_{test}. We choose a fovea localization success threshold of 2 degrees, i.e. a prediction is considered successful if its error is below 2 degrees. On the whole data set, NAM has a fovea localization success rate of 52%, and RBFL improves it to 84%. On the central fixation-primary head position subset, NAM has a 73% success rate and RBFL brings it to 100%.

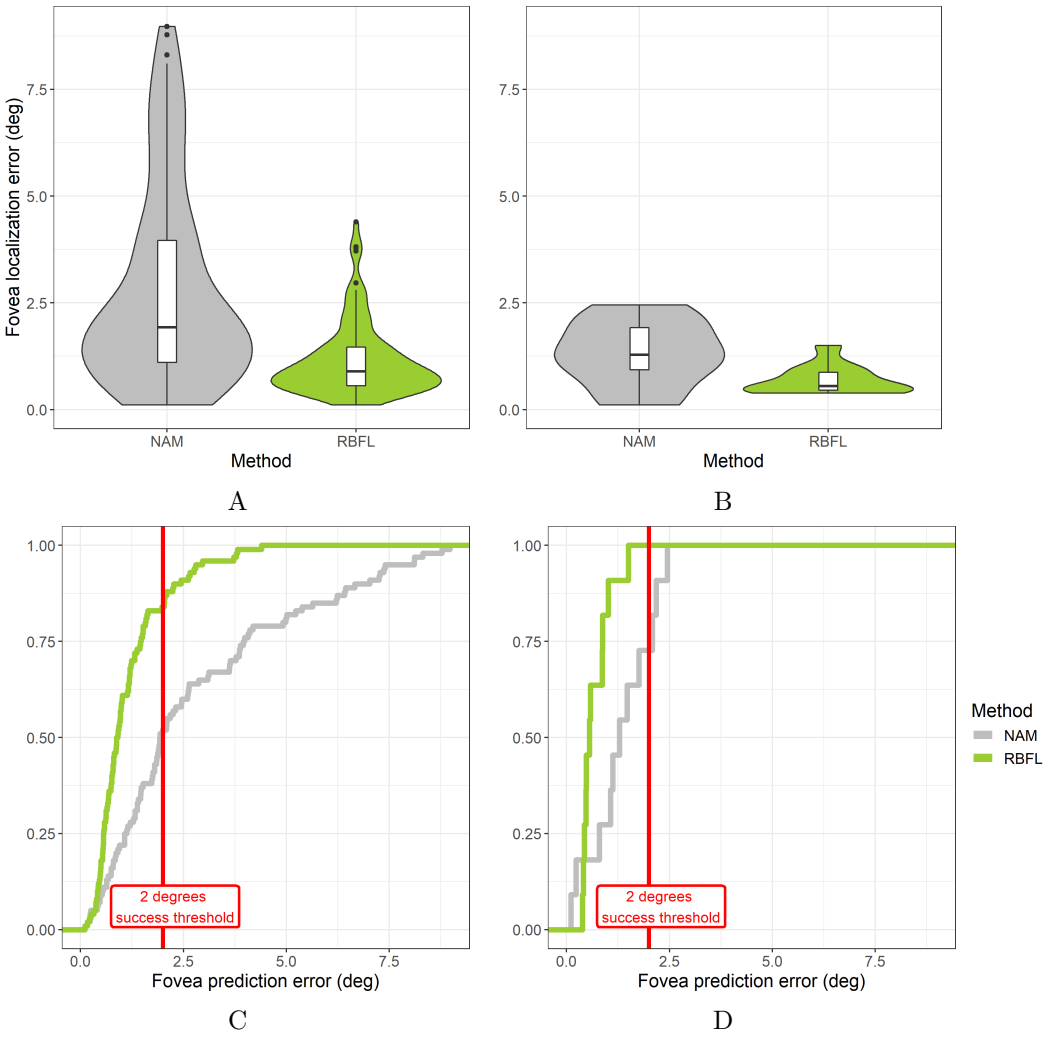


Figure 14: Fovea localization errors of the proposed RBFL method, compared to the standard NAM method. Fig. 14A: Error distributions as violin plots, on the whole testing set D_{test}. Fig. 14B: Error distributions as violin plots, on the central fixation-primary head position subset of D_{test}. Fig. 14C: Cumulative error distributions, on the whole testing set D_{test}. Fig. 14C: Cumulative error distributions, on the central fixation-primary head position subset of D_{test}.

3.3 Effect of Fixation Location and Head Position

In Fig. 14, we have shown that the performance with eyes in central fixation and primary head position is notably better than the overall performance considering all cases. In Fig. 15 we propose to investigate more thoroughly how all configurations of fixation locations and head positions affect the performance of NAM and RBFL methods.

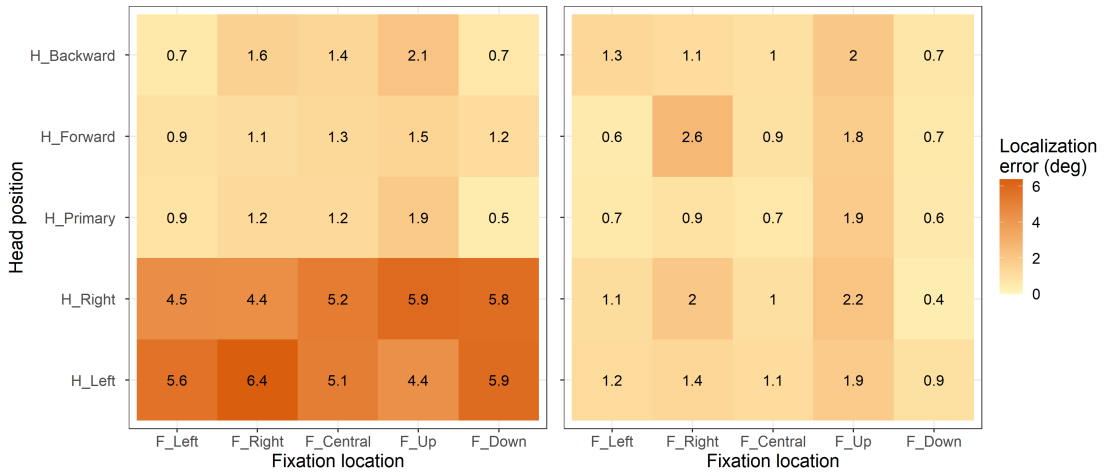
For the NAM method (Fig. 15A), we can notice a relatively worse performance for the left and right head positions. This is coherent with the fact that these head positions induce a vertical shift of the optic disk position which serves as a reference in NAM.

Concerning the RBFL method (Fig. 15B), it mostly has relatively weaker performances for two fixation locations: fixations up and right, which are the two fixations with the highest mean errors, 1.94 and 1.66 respectively, while the rest have lower mean errors (0.97 for left, 0.92 for central and 0.67 for down fixation). A potential explanation is illustrated in Fig. 16. We observe that in these two eye fixation configurations the region with high vessel density corresponding to the optic disk is partially or completely hidden, suggesting that this is an important information used by our method.

Concerning the head position, we cannot draw any conclusion about an effect. With a Wilcoxon test, we found no statistically significant ($p < 0.05$) difference between the error distributions of the different head positions. Therefore, the RBFL method seems robust to different head positions. Fig. 17 shows a visualization of the results of RBFL on the same retina for the five different head positions.

In effect, this robustness to head position leads to the largest improvement over NAM: this method assumes that the fovea-optic disk angle is constant, but as shown by Fig. 17, a right or left orientation changes this angle. This leads NAM to a mean error of 5.23 (SD 1.95) degrees for these two positions combined, against 1.28 (SD 0.90) for RBFL, representing an error decrease of 76%.

In the other three head positions, namely backward, forward and primary, RBFL represents a lesser improvement over NAM: from a mean error of 1.32 (SD 0.71) degrees for NAM to 1.08 (SD 0.82) degrees for RBFL, i.e. an error decrease of 18%.



A (NAM) B (RBFL)

Figure 15: Mean fovea localization error for each fixation location-head position pair, for the NAM method (Fig. 15A) and the RBFL method (Fig. 15B).

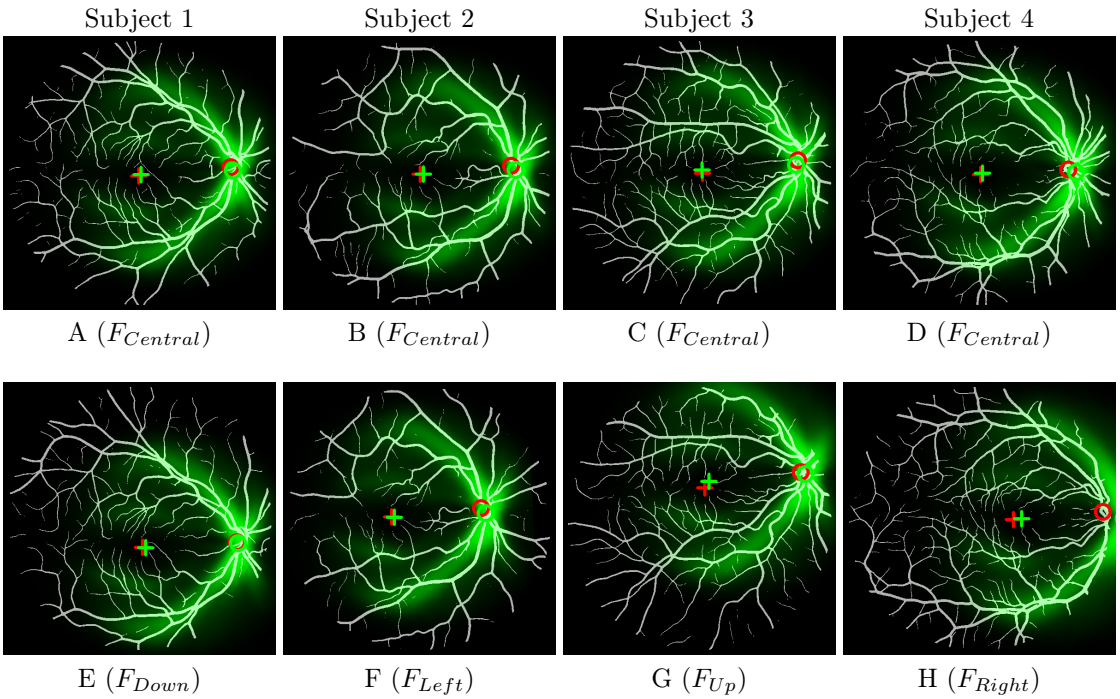


Figure 16: Illustration of the effect of fixation on the visible vessels and the RBFL results. All images shown here correspond to a primary position of the head. Each column is a different subject. Eye fixations are indicated below each image. The first row always corresponds to a central fixation $F_{Central}$. In the second row, different fixations are shown.

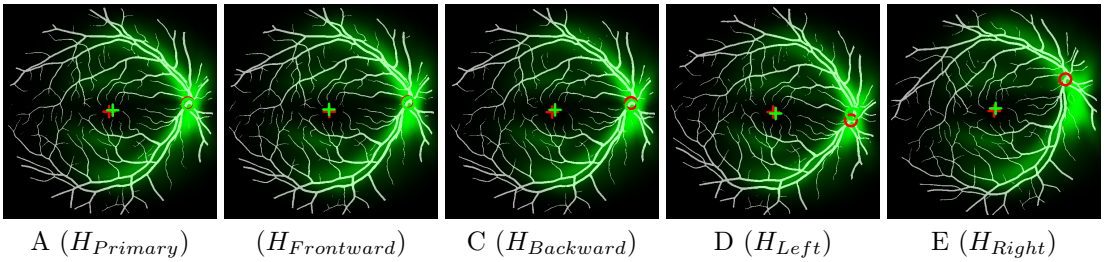


Figure 17: Illustration of the effect of head position on the visible vessels and the RBFL results. All the images are from the same subject, with central fixation, and correspond to different head positions.

3.4 Effect of Axial Length/Myopia on the Method

A higher degree myopia, or a higher eye axial length, can cause a distorted retinal shape. The geometry of the eye is changed, and so might be the geometry of the apparent vessels. Therefore, we may wonder if a higher axial length is correlated to a higher fovea localization error with our method.

We know the refractive error and axial length of each eye in the test set D_{test}. Fig. 18 shows the fovea localization error are as a function of the axial length, for the eyes with central fixation and main head position.

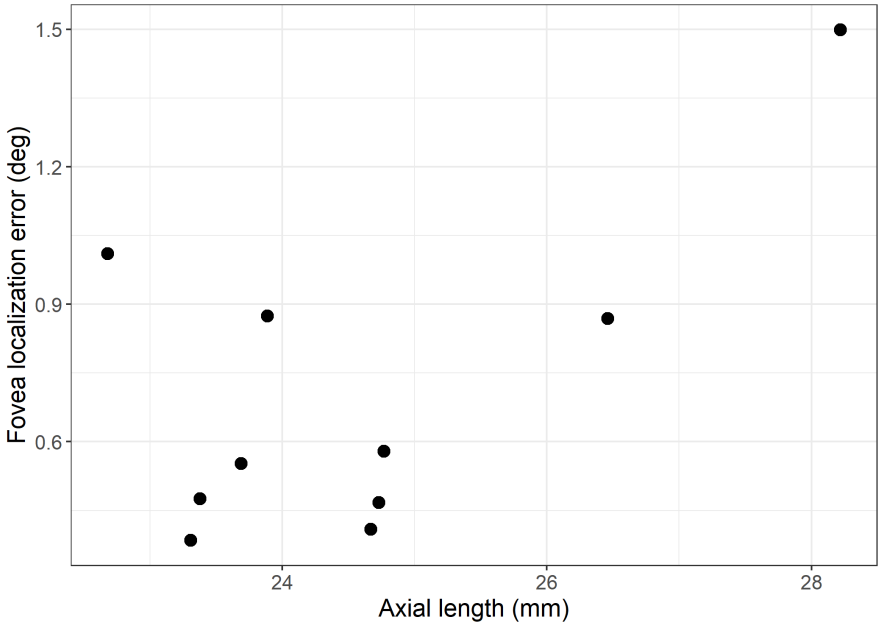


Figure 18: Effect of the axial length on the performances of the method.

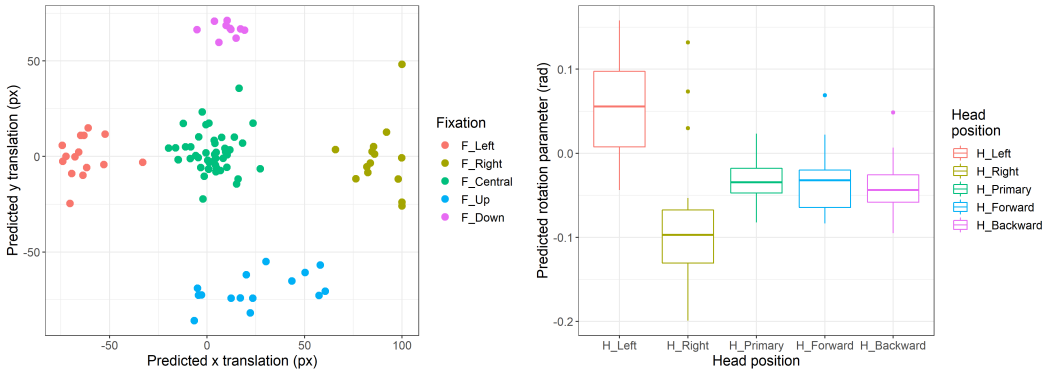
We do not see any effect. However, ten eyes is too small of a sample size to draw any conclusion, and differences in error may just be due to other individual variability. Further investigation with additional data could help us understand whether higher axial lengths affect the performances of our method.

3.5 Relationship Between the Transformation Parameters and Head Position and Fixation

In Sec. 3.3 we have shown that a non-standard head position and fixation location can affect the performances our method. Reversely, we find that the output transformation of our method can give us information about the head position or fixation location of an input image. Figure 19 shows the relationship between the estimated translation or rotation parameters, and the fixation location or head position respectively.

Fig. 19A shows both the predicted translation parameters (t_x, t_y), and the fixation location. We can clearly identify five clusters, corresponding to the five possible fixation locations. On the other hand, head position is correlated to the predicted rotation θ. Fig. 19B shows the predicted rotation parameter as a function of head position. The predicted rotations in primary, forward

or backward head positions are similar, with respective means of -0.036 (SD 0.029), -0.034 (SD 0.038) and -0.042 (SD 0.034). In left and right positions however we obtain significantly different results, with respective means of 0.055 (SD 0.057) and -0.085 (SD 0.082).



A B

Figure 19: Relationship between the estimated transformation parameters and the head orientation or fixation location. Fig. 19A Predicted translation parameters and fixation location. Fig. 19B Predicted rotation parameter as a function of head position.

These results are not surprising: the origin of the scaling and rotation in $\mathcal{T}$ are the reference fovea $(x^F_{\text{ref}}, y^F_{\text{ref}})$, so the predicted fovea $\mathcal{T}(x^F_{\text{ref}}, y^F_{\text{ref}})$ will only depend on the translation part of $\mathcal{T}$. Furthermore, we can see in Fig. 17 that left and right head positions will significantly change the fovea-optic disk angle, which is θ up to a constant, while this angle remains visually constant in other positions.

3.6 Simulated Maculopathies

In order to test the robustness of the method to AMD-like lesions, we introduce simulated scotomas in our data. These simulated scotomas are black masks applied over the foveas (x^F_i, y^F_i) of the vessel maps v_i. We use three shapes and five sizes of masks, which are illustrated in Fig. 20. The three shapes are: a circle, a horizontal ellipse and a vertical ellipse (the ellipses have a major axis twice as large as the minor axis). The sizes are taken as 20, 50, 100, 200 and 400 deg^2 of surface area. We apply the masks with each combination of size and shape to each vessel map $v_i \in D_{\text{test}}$, and test the method on this new augmented data set.

Fig. 21 presents the results we obtain. As one could expect, the mean error globally increases with the mask size, for every shape, and the maximal error is obtained with a size of 400 deg^2. However interestingly, the mean error decreases for every mask shape until a size of 50 deg^2. At these sizes, only the very small vessels around the macula are masked (see Fig. 20). This might indicate that these vessels are detrimental to the RBFL method, and introduce meaningless noise.

Even at the largest mask size of 400 deg^2, and with the most challenging shape at this size which is the vertical ellipse, the method still performes better than the NAM method (which only relies on the optic disk and in not affected in any way by macular lesions). Indeed, in this case the RBFL method has a mean error of 2.54 (SD 1.9), which is 11% less than the mean error of NAM (2.85 (SD 2.33)).

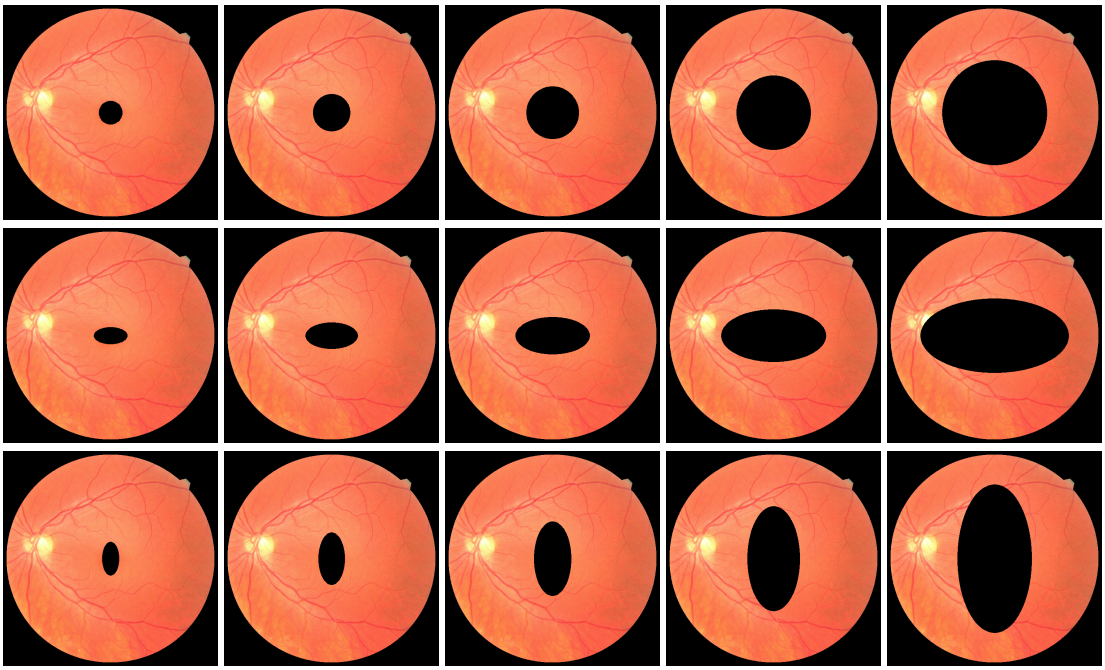


Figure 20: Illustration of the different sizes and shapes of masks used to hide the vessels, to simulate a macular lesion. Each row corresponds to one of the different shapes used, and each column to a different size: 20, 50, 100, 200 and 400 degrees^2 respectively.

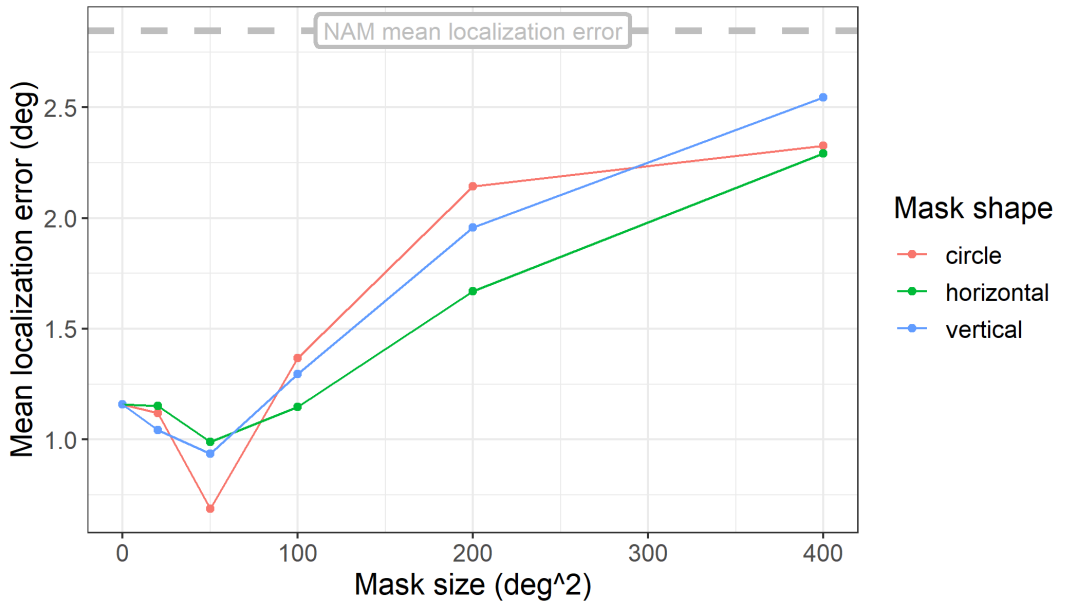


Figure 21: Mean localization error of the proposed RBFL method as a function of lesion size, for each lesion shape. Grey dotted line represents the mean error of the NAM method, which does not depend on mask size or shape.

3.7 Results on Real AMD Images

While the method gives promising results on images with simulated macular lesions, one might wonder if it can be applied to real AMD data. To test this, we use the ADAM data set (Fu et al., 2020). It contains 400 eye fundus images with a manually annotated fovea location. 89 images are from belong to AMD patients, and 311 from non-AMD patients. Though the annotations on the AMD images have to be considered carefully, as the fovea often not visible because of the lesions, we can get an idea of whether the method is applicable in a real case. For a comparision, we also compute results on the 311 non-AMD images.

As the field of view to acquire these images is different from that in D_{test}, and we do not know the correspondence in degrees of visual angle, the results will be given in pixels.

Fig. 22 presents the error distributions on both of these data sets. The mean localization error on AMD images is 35.8 (SD 32.6) pixels, which we can compare to the 59.6 (SD 59.0) on the healthy images. A Wilcoxon test rejects the null-hypothesis that the error distribution on the non-AMD data is lower than the error distribution on AMD data ($p - value = 0.000642$). Our current results therefore suggest that real AMD lesions do not alter the method's performances in a negative way.

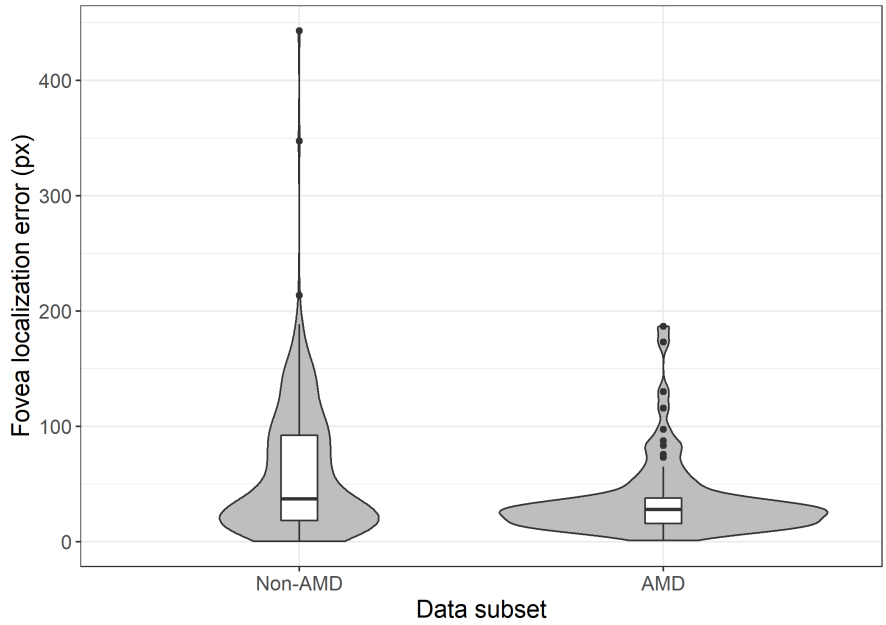


Figure 22: Fovea localization error distribution on healthy retinas, compared to the results on real AMD cases.

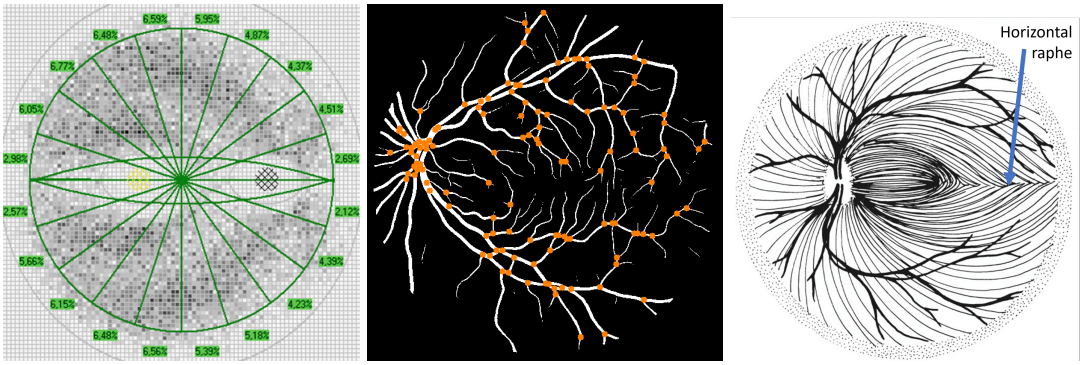
4 Discussion

4.1 Are There Other Anatomical Landmarks and Characteristics?

One main challenge in our problem was to deal with the strong variability of retinal fundus images. This variability concerns both the raw intensity distributions of fundus images (and even more with impaired retinas), and the vessels patterns. In this paper, we chose to circumvent this problem by using a statistical representation of vessels. Another option would be to find robust anatomical landmarks. Two options were considered.

The first has been to use vessel bifurcations and crossings (Fig. 23A–B). Inspired by the clear pattern of bifurcations and crossings obtained across 670 images (Fig. 23A), an idea could be to identify an axis of symmetry between the upper and lower parts. Unfortunately, this idea is difficult to apply in practice because the number of bifurcation and crossing points is low in a single image, and because the upper and lower retinal hemispheres are not always visible in the same proportion due to, e.g., head position (Fig. 23B).

The second option has been to consider the notion of raphe (Jansonius and Schiefer, 2020), which is defined by the horizontal meridian separating the inferior and superior retinal nerve fibers (Fig. 23C). Interestingly, this structure represents a robust landmark but it is only defined at a microscopic level (retinal nerve fiber bundle trajectories are detected using OCT imaging), and not at the macroscopic level that we have to consider.



A B C

Figure 23: Alternative Anatomical Landmarks and Characteristics. Fig. 23A: Distribution of bifurcations and crossings (Courtesy Semerád and Drahanský (2020)). The yellow mark represents the fovea, and the black mark the optic disk. Fig. 23B: Vessel bifurcations and crossings for one vessel map. Each orange dot represents a crossing or bifurcation. Fig. 23C: Illustration of the horizontal raphe (Courtesy Pineles and Balcer (2019)): it is the watershed between the inferior and superior retinal nerve bundles.

4.2 Structure Tensor Registration is Not Efficient

When defining the energy to minimize E (1), our first attempt has been to have a first term for average density, and a second term for structure tensors as a whole instead of simply the directions. For this second term, the problem was to define how to properly registers tensor fields, i.e., define a proper metric. To do so, we used the geodesic distance (Faugeras et al.,

2007), defined for two symmetric definite positive matrices $\mathbf{s}_1$ and $\mathbf{s}_2$ as:

$$d_{Geodesic}(\mathbf{s}_1, \mathbf{s}_2) = \sqrt{\frac{1}{2}\left(\log^2\left(\mathbf{s}_1^{-\frac{1}{2}}\mathbf{s}_2\mathbf{s}_1^{-\frac{1}{2}}\right)\right)}. \quad (19)$$

This method was not chosen in the end for two main reasons: (i) The geodesic distance is not easy to interpret and has singularities for null tensors which happen in our case, (ii) It is computationally heavy and the improvements with respect to using the average density alone were not significant (+3000% of computational time for hardly any improvement). For these reasons, we chose to keep only the information from the main direction of tensors (i.e., vessel directions), so that a simpler notion of distance can be used, the additional computational cost is relatively low (+194%), but also the performance gain w.r.t. average density only is significant (-33% of mean error).

4.3 Could Deep Learning Methods Be Useful in Our Problem?

Given the common trend about deep learning methods, including for retina image processing (Sec. 1.2), a natural question was to investigate if they could be useful in our case to predict fovea position. To do so, we started from the framework we used to segment vessels, i.e., SA-UNet (Changlu et al., 2021). The interest of this network is that it has been pre-trained, and learnt to encode the important features from fundus images, in order to find the vessels, with the decoder.

As fovea localization should in our case be treated as a regression task (Sec. 1.2), we adapted the network to perform regression instead of segmentation. To do so, we thus proposed a new network design which would preserve the pre-trained encoder, and replaced the decoder by fully connected layers. We could then train the network and learn the weights of these new layers, while freezing the weights of the encoder (see Appendix, Sec. A.3 for more details).

For this training, the difficulty is that we need to train on images where there is no visual information allowing the fovea to be directly detected, otherwise the network could simply "see" the location of the fovea, like a human would. Since fovea position was known in our training set, a natural solution has been to use inpainting to fill an area around the fovea. Interestingly, despite several attempts using state-of-the-art algorithm for inpainting, we found that the predicted fovea position was in fact dependent on the area inpainted. In other words, our network was learning some visual patterns of the inpainted areas, even if, visually, nothing could be noticed. This surprising result is yet another good illustration of the biaises that could come from statistical learning. For this reason, no relevant and unbiased result could be produced with a deep learning approach. But more generally, for the problem we have to solve and the final application to real images, this methodology appears inadapted because of two contradicting requirements: having both a data set in which the fovea is hidden by AMD lesions, while knowing the ground truth location of the fovea.

4.3.1 What About Solving The Problem in 3D?

Another solution that we have investigated is to solve the problem in 3D, i.e., given modeling the retina as an ellipsoid, use the fact that the fovea can be found from to the location of the projection of the ellipsoid's vertex (Fig. 24). Indeed, the geometry of the retina is a well documented subject. It can be approximated by an oblate ellipsoid, with the ellispoid vertex consistently located 0.5mm nasally and 0.2mm inferiorly to the fovea (Atchison et al., 2005). Furthermore, the lengths of this ellipsoid's axes can be inferred from the refractive index of the eye.

So our problem amounts to estimate an ellispsoid model of the retina from retina fundus images. In theory, reconstructing 3D scene from a set of images of 2D images of the scene is possible, and this refer to large body of litterature in three-dimensional (stereo) vision (Faugeras, 1993; Moons et al., 2009). Such methods have been applied in the context of three dimensional reconstruction of the retina given a set of multiple views from slightly different angles of the same retina (Martinez-Perez and Espinosa-Romero, 2012; Hernandez-Matas et al., 2020). In theory, we could then also apply them to our D_{test} data set, since we have both multiple views of the same retina and the refractive index of the eyes. In particular, given the prior of an ellipsoid retina shape and size, and multiple views of the same retina, the method proposed by Hernandez-Matas et al. (2020) can be used to infer the location and orientation of this ellipsoid in real-world coordinates, as well as the orientation and location of the camera in real-world coordinates, for each image. This is enough information to infer the location of the fovea on the images.

This may be an interesting solution to further investigate but one should keep in mind several potential difficulties and limitations: (i) even if, theoretically, stereo works with two images, when dealing with real cases one needs much more images and it is difficult to predict the number of images that will be necessary to reach a precise 3D reconstruction, (ii) the quality of the images may be a problem, especially when applying this approach in real cases, (iii) the vessels we will reconstruct and that will serve to find the ellipsoid parameter have different depths and this may have an impact on the precision of the parameters estimation, (iv) last but not least, this kind of method require a strong expertise.

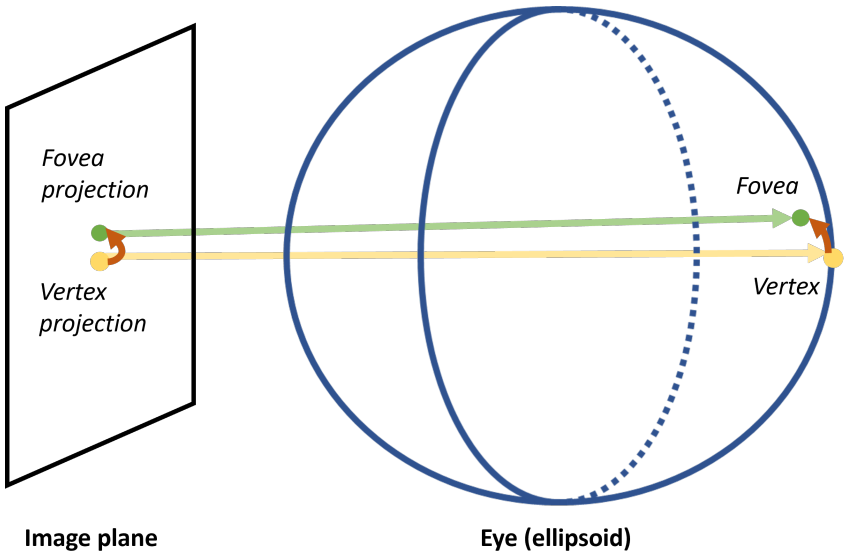


Figure 24: Illustration of the geometry of the retina. The retina is very well approximated by an ellipsoid, with the fovea at a known distance from its vertex (Atchison et al., 2005). If we knew the position and orientation of this ellipsoid relatively to the image plane, we could locate the fovea on the image.

4.4 A Method for Optic Disk Detection?

We used this registration-based method to estimate the location of the fovea in fundus images, using the location of the reference fovea $(x^F_{\text{ref}}, y^F_{\text{ref}})$ of the average vessel density $\bar{V}$ and directions

$\bar{D}$. However, by construction, we also know the location of the reference optic $(x_{\text{ref}}^{OD}, y_{\text{ref}}^{OD})$ on these maps. Therefore, this method can be used as is to locate simultaneously the fovea and the optic disk.

From our experiences on the test set D_{test}, the proposed method can be used to locate the optic disk with a mean localization error of 2.4 (SD 2.3) degrees of visual angle. Fig. 25 shows the optic disk localization error distribution on the D_{test} data set.

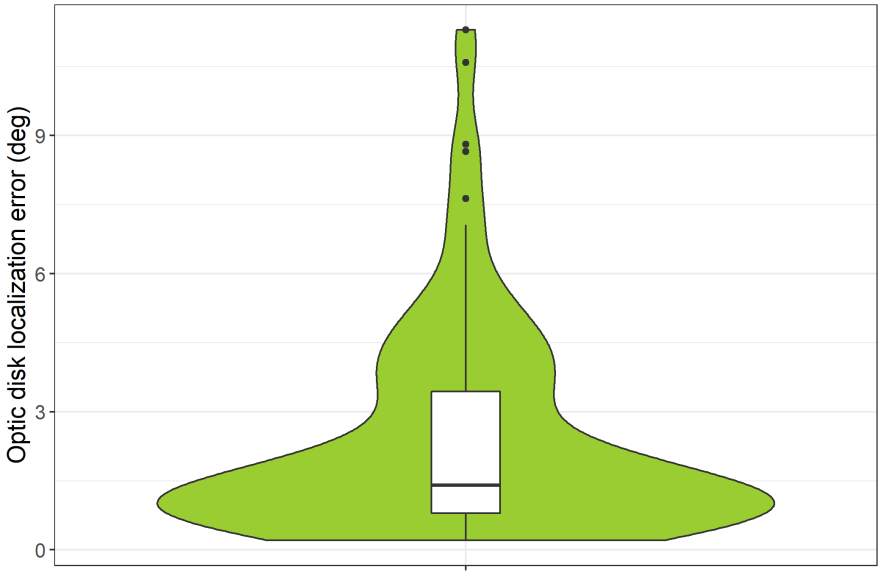


Figure 25: Optic disk localization error distribution.

Further tweaking of the method's parameters, such as adding a higher weight around the reference optic disk location $\mathcal{T}\left((x_{\text{ref}}^{OD}, y_{\text{ref}}^{OD})\right)$ in the weights $\lambda_{\mathcal{T}}$, could surely help making optic disk localization more precise. Because our method locates the optic disk from the vessels alone, without any color information, it has an interesting advantage compared to other, traditional segmentation approach. Indeed, our method can be used to locate the optic disk even when it is partially or totally out of the field of view of the image.

5 Conclusion

In the presence of maculopathies, locating the fovea on fundus images is a crucial but difficult task. While normative anatomical measures based on the optic disk location have been extensively used, we find that their precision is in practice very sensitive to the subject's position and fixation during the image acquisition.

To solve this problem, the proposed Registration-Based Fovea Localization (RBFL) method is solely based on the vascular structure. Indeed, vessels seem to be the most robust landmark to base the localization on: unlike the optic disk they remain visible when patients tilt their head or have eccentric fixation. Furthermore, the main vessels remain visible in poorly illuminated images or images with maculopathies.

In essence, the method exploits the statistical regularity of the retinal vascular structures. By computing the average vessel density and directions over a large data set, in which the reference fovea location is known, we have shown how tightly connected the density and directions of the vessels are to their anatomical location on the fundus. Thus we can simply locate the fovea in a new fundus image by performing the registration of the average vessel density and directions onto the vessels of this image.

Our tests indicate that in practice, the vessels indeed hold enough information to precisely locate the fovea in a single image. The $RBFL$ method addresses the aforementioned drawbacks of the normative anatomical measures, achieving all-around better performances, especially in cases of head tilt. Large simulated lesions hiding the vessels will alter the localization performances of $RBFL$, but it remains more precise than the alternative normative anatomical measures. Further testing with more realistic simulated image alterations could be done and bring better understanding of the robustness and fail-cases of this method.

Small inaccuracies were in particular identified when the areas of higher vessel density around the optic disk fall outside of the fundus images. Algorithmic improvements can surely be made to increase the overall precision of the method. One way could be to explore more thoroughly the parameter initialization and bounds of the registration algorithm, which are important factors. One could use the observation that the estimated registration parameters are related to the head position and fixation location. Indeed, a prior (obtained by human input, or by a simpler algorithm) on head position of fixation location could then be used to set more appropriate initialization and bounds in each specific case.

6 Acknowledgements

In this work, we have investigated several possible solutions and we are very thankful to several colleagues who provided their support to help us progress in these differents tracks. We would like to express our very great appreciation to Dr David Tschumperlé who shared his expertise to find the best inpainting approach. We thank Dr Frédéric Precioso for his valuable and constructive suggestions concerning the machine learning approach. Special thanks also to Dr Théodore Papadopoulo who gave us some very nice introductory course about 3D reconstruction approach.

A Appendix

A.1 Parameters Setting and Computational Performances

Parameters used in our paper are summarised in Tab. 1.

Although this was not a priority in our work, we describe in this section the computational performance of our multiscale approach. The hardware environment set up for the experiment was as follows: an Intel(R) Core(TM) i7-4720HQ 2.60GHz CPU and 8 GB of memory. Implementation was made using Python 3.7. We ran our approach on images of size $(592, 592)$ (see Appendix, Sec. A.1, for parameters settings). Average runtime is 6.60 seconds/image total: 2.05 seconds/image for the segmentation, and 4.55 seconds/image for the registration.

A.2 The Notion of Tensors for Representing Direction Distributions

Let's briefly explain why do we need tensors. Let's assume that we have a set of vectors $w_i = (w_i^x, w_i^y)$ from which we would like to know how they are distributed, e.g., are they aligned or evenly distributed in terms of direction? Of course one could think about looking at the average vector

$$\sum_i w_i,$$

but this simple computation is not useful. To convince yourself, just think about a set of opposite vectors whose average would be zero. Now, let $d(\theta)$ be the vector $(\cos\theta, \sin\theta)$. An elementary calculation shows that the function $F(\theta) = (d\,(\theta) \cdot w_i)^2$ is maximal if d is parallel to w_i, and is minimal if d is orthogonal to w_i. We can also remark that maximizing (respectively minimizing) $F(\theta)$ is equivalent to maximizing (respectively minimizing) the quadratic form $d^t\, w_i w_i^t\, d$. The matrix

$$w_i w_i^t = \begin{pmatrix} {w_i^x}^2 & w_i^x w_i^y \\ w_i^x w_i^y & {w_i^y}^2 \end{pmatrix}, \tag{20}$$

is positive semidefinite, its eigenvalues are $\lambda_1 = |w_i|^2$, and $\lambda_2 = 0$ and there exists an orthonormal basis of eigenvectors ω_1 parallel to w_i and ω_2 orthogonal to w_i. Applying the same reasoning for the set of all vectors $\{w_i\}$, the idea is thus to minimize the sum of the quadratic forms, i.e.,

$$\sum_i d^t\, w_i w_i^t\, d = d^t \mathbf{T}\, d,$$

where the tensor $\mathbf{T}$ is defined by:

$$\mathbf{T} = \sum_i w_i w_i^t, \tag{21}$$

which is positive semidefinite. Interestingly, its eigenelements will thus present the distribution of directions. This tensor is usually represented by an ellipse so that (Fig. 26):

- the major axis defined by the eigenvectors e_1 corresponding to the highest eigenvalue λ_1,
- the minor axis defined by the other orthogonal eigenvector e_2.
- axis lengths are proportional to eigenvalues (λ_1, λ_2).

Description	Parameter	Value	Section(s) or Equation(s)
Image properties			
Image dimensions (pixels)	(s_x, s_y)	(592,592)	Sec. 2.4.1
Pixel/degree ratio in D_{test}	r	0.07904	Sec. 3.2
Energy	$E = E_V + \eta E_D$		Eq. (2)
Region-based weight	λ		Sec. 2.3.2
Macular region weight	w_{macula}	10	Sec. 2.3.2
General weight	w_{general}	1	Sec. 2.3.2
Outer weight	$\{w^m_{\text{outside}}\}_{m=0,\ldots,M-1}$	$\{0, \frac{1}{3}, \frac{2}{3}, 1\}$	Sec. 2.3.2
Density term weight	η	2	
Vessel direction estimation			
Image Gaussian	σ	1	Eq. (6)
Tensor field Gaussian	ρ	5	Eq. (6)
Average density enhancement			
Sigmoid parameter	μ	0.3	Eq. (18)
Registration algorithm			
Transformation parameters			
Vertical translation	t_y		Sec. 2.3.1
lower bound (at scale m)		$-100/2^m$	Sec. 2.4.2
upper bound (at scale m)		$100/2^m$	Sec. 2.4.2
initial value		0	Sec. 2.4.2
Horizontal translation	t_x		Sec. 2.3.1
lower bound (at scale m)		$-100/2^m$	Sec. 2.4.2
upper bound (at scale m)		$100/2^m$	Sec. 2.4.2
initial value		0	Sec. 2.4.2
Rotation	θ		Sec. 2.3.1
lower bound		-0.7	Sec. 2.4.2
upper bound		0.7	Sec. 2.4.2
initial value		0	Sec. 2.4.2
Uniform scaling	s		Sec. 2.3.1
lower bound		0.85	Sec. 2.4.2
upper bound		1.15	Sec. 2.4.2
initial value		1	Sec. 2.4.2
Bounds margin factor	β	0.9	Eq. (16)
Number of scales	M	4	Sec. 2.4.1

Table 1: Parameters of the method RBFL.

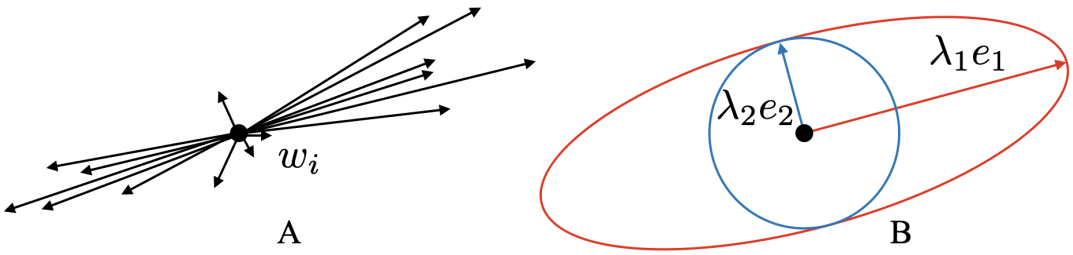


Figure 26: Illustration of tensor interpretation. Fig. 26A represents a set of vectors w_i for which a tensor can capture the distribution shown by an ellipse in Fig. 26B.

A.3 SA-UNet and Its Application to Our Case

As highlighted in Sec. 1.2, deep learning networks have shown great promise as a tool to locate the fovea in healthy fundus images. The question remains: can we do the same in images with maculopathies? As we discussed, we could try to use a deep regression network, which seem more appropriate to our case than a segmentation network. Nevertheless, some problems remain to solve.

A standard fovea localization network trained on healthy data would not provide any relevant results, as it would use visible macular features to make its prediction. However, having a ground truth fovea location in an image with macular lesion is impossible, making training on such data not an option either. Therefore, training on healthy data with artificially hidden foveas seems like the only potential remaining solution to this problem. Nonetheless, we are constrained in the way we can hide the foveas: once again, no direct distinguishable feature should be detectable by the network giving clues about the fovea mask. For example, if we hide the foveas under a large green circle, the network learns that the fovea is always somewhere under a large green circle. This would introduce bias in the network, making it worthless in real uses, on new or truly pathological data.

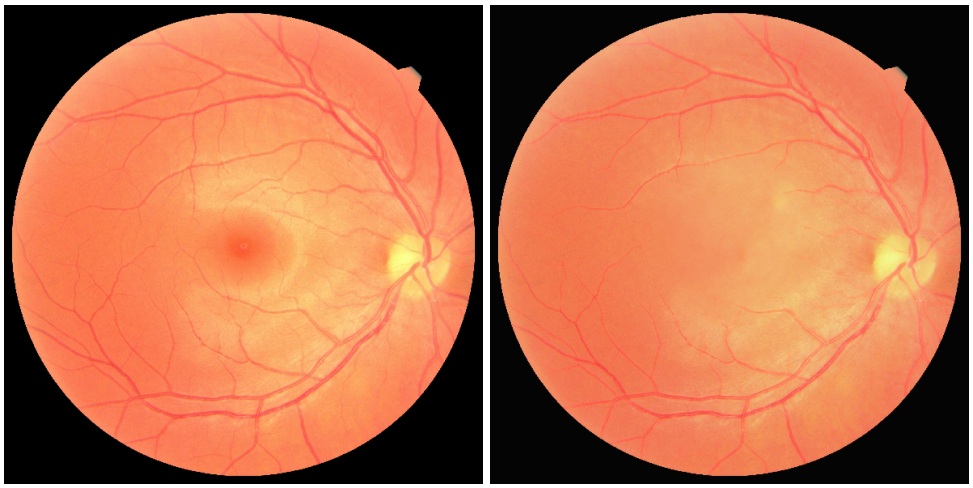
Taking these constraints into consideration, we proposed to use inpaint the macular region in healthy data. Inpainting is a classical process in image processing, aiming to fill a missing part of an image. In this case, we masked a circle containing the fovea (the circle has a random offset so the fovea is not necessarly at the center of the mask), which was inpainted as illustrated in Fig. 27. We tried two types of inpainting methods : a patch-based method, and a more complex approach decomposing each image in a low frequency part and a texture part with a rolling guidance filter (Zhang et al., 2014), then inpainting the low frequency part using a partial differential equation approach and the textures with a patch-based approach.

The network architecture is composed of the pre-trained encoder of SA-UNet (Changlu et al., 2021) (retinal vessel segmentation network), with an added regression head instead of the segmentation decoder. This regression head would be trained on healthy data from D_{train}, with an inpainted macula. Intuitively, this could solve our problem:

(i) We the inpainting, the fovea should be effectively hidden. Figure 27 shows the results of the inpainting method tested: the local information has been removed, and visually one cannot directly locate the fovea anymore.

(ii) The network architecture and inpainting should make the approach unable to detect and use the location of the fovea from features specific to the inpainted region. Indeed, the fovea is hidden in a way such there is seemingly no texture distinction between the inpainted macula and the rest of the fundus. To make it even harder for the network to use features

of the inapinted regions to locate the fovea, we randomize the inpainting region location, so the fovea location inside it is random (e.g. the fovea is not systematically at the center if the inpainted region). Additionally, the regression head can only use the encoded features from a vessel segmentation network. This encoder will not be trained, so *a priori* it should not encode so much features relevant to the inpainted fovea, but instead mostly encode features related to the vessels. In particular, the encoder will be unable to learn new features, so the ability of the network to use these inpainting-related features should be limited.



A B

Figure 27: Illustration of the inpainting of tha macula. Fig. 27A: the original fundus image. Fig. 27B: the inpainted fundus image.

Figure 28 presents the architecture of the network. The input of the network is a color fundus image u, and its output is the predicted fovea location (x_i^F, y_i^F).

The network was trained using a mean squared error loss function, the ADAM optimizer with a learning rate of 0.001, for 50 epochs, with a batch size of 4. Random rotation and flip were used for data augmentation. We now present the results obtained by this approach. We computed them on our test dataset D_{test}, where we inpainted the maculas, with masks centered on the fovea.

The mean localization error obtained with the network is 2.43 degrees of visual angle, making it 15% less than the normative anatomical measures (NAM) method (with 2.85 mean error), but 100% more than the proposed RBFL method (with a 1.16 mean error).

Figure 29A shows the error distribution of this deep learning approach against those of RBFL and NAM. Figure 29B shows the mean error for each head position-fixation location pair. This plot can be compared to those in Sec. 3.2.

We can see that this method struggles in the up and down fixations. This may be explained by the fact that these fixations do not appear in the training dataset: the fixation variability is a right-left variability, not up-down. As this is a statistical model, a bias or lack of diversity in the data leads to this bias in the method.

To test whether the method detects and uses specific features of the inpainted area, we can simply shift the inpainting region by +20 pixels along both the x and y axes, and see how the prediction of the network change. If we do so, the network's prediction are on average shifted +10.8 pixels along the x axis and +3 pixels along the y axis. This suggests that the

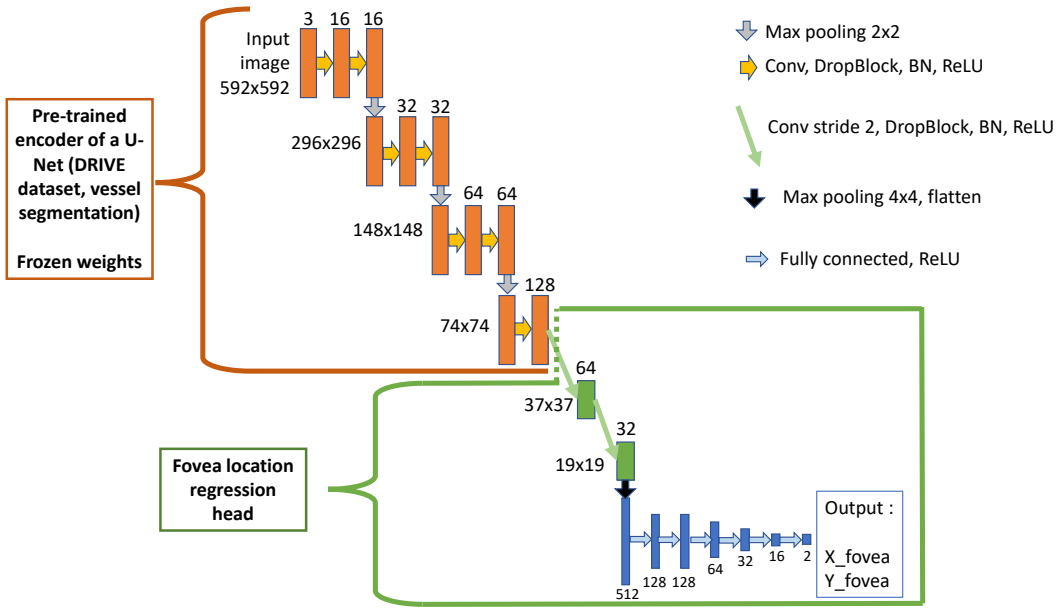


Figure 28: The modified SA-Unet architecture.

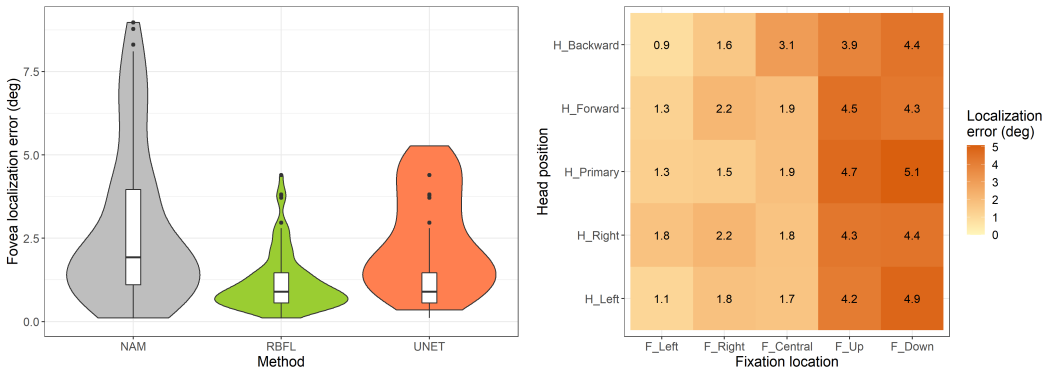


Figure 29: Performances of the modified UNet for fovea localization. Fig. 29A: Fovea localization error distribution as violin plots for the NAM, RBFL and UNet methods. Fig. 29B: Mean fovea localization error of the UNet method for each head position-fixation location pair.

network indeed can detect and use this inpainting information. By inpainting the images, we have introduced bias and information about the ground truth location.

On top of the worse performances of this method compared to the proposed RBFL method, two problems have made this deep learning solution impractical. First, we were not able to hide the macula without introducing information usable by the network. Second, our training data set is biased and lacks diversity about eccentric fixation locations. This therefore shows that applying deep learning to fovea localization in pathological fundus images is a very delicate subject, and deep learning might not be adapted to this type of data. A solution to go further could be to build and use a more diverse training data set, and to apply a neural network onto vessel segmentation maps v, eliminating any introduced bias by forcing the network to use only vessel-related information.

References

Ahuja, A. K., Yeoh, J., Dorn, J. D., Caspi, A., Wuyyuru, V., McMahon, M. J., Humayun, M. S., Greenberg, R. J., daCruz, L., and Group, A. I. S. (2013). Factors Affecting Perceptual Threshold in Argus II Retinal Prosthesis Subjects. *Translational Vision Science & Technology*, 2(4):1–1.

Al-Bander, B., Al-Nuaimy, W., Williams, B. M., and Zheng, Y. (2018). Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. *Biomedical Signal Processing and Control*, 40:91–101.

An, C., Wang, Y., Zhang, J., Bartsch, D.-U. G., and Freeman, W. R. (2020). Fovea localization neural network for multimodal retinal imaging. *Proceedings SPIE 11511, Applications of Machine Learning 2020*, pages 196–202.

Atchison, D., Pritchard, N., Schmid, K., Scott, D., Jones, C., and Pope, J. (2005). Shape of the retinal surface in emmetropia and myopia. *Investigative ophthalmology & visual science*, 46:2698–707.

Aubert, G. and Kornprobst, P. (2006). *Mathematical problems in image processing: partial differential equations and the calculus of variations (Second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag.

Calabrèse, A., Bernard, J.-B., Hoffart, L., Faure, G., Barouch, F., Conrath, J., and Castet, E. (2011). Wet versus dry age-related macular degeneration in patients with central field loss: different effects on maximum reading speed. *Invest Ophthalmol Vis Sci*, 52(5):2417–24.

Changlu, G., Marton, S., Yugen, Y., Wenle, W., Buer, C., and Changqi, F. (2021). Sa-unet: Spatial attention u-net for retinal vessel segmentation. *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 1236–1242.

Faugeras, O. (1993). *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press.

Faugeras, O., Lenglet, C., Papadopoulo, T., and Deriche, R. (2007). Non rigid registration of diffusion tensor images. Technical Report 6104, INRIA.

Fu, H., Li, F., Orlando, J. I., Bogunović, H., Sun, X., Liao, J., Xu, Y., Zhang, S., and Zhang, X. (2019). Refuge: Retinal fundus glaucoma challenge. IEEE Dataport.

Fu, H., Li, F., Orlando, J. I., Bogunović, H., Sun, X., Liao, J., Xu, Y., Zhang, S., and Zhang, X. (2020). Adam: Automatic detection challenge on age-related macular degeneration. IEEE Dataport.

Gomes, N. L., Greenstein, V. C., Carlson, J. N., Tsang, S. H., Smith, R. T., Carr, R. E., Hood, D. C., and Chang, S. (2009). A Comparison of Fundus Autofluorescence and Retinal Structure in Patients with Stargardt Disease. *Investigative Ophthalmology & Visual Science*, 50(8):3953–3959.

Hernandez-Matas, C., Zabulis, X., and Argyros, A. A. (2020). Rempe: Registration of retinal images through eye modelling and pose estimation. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3362–3373.

Jansonius, N. M. and Schiefer, U. (2020). Anatomical Location of the Raphe and Extended Raphe in the Human Retina: Implications for Assessment of the Optic Nerve with OCT. *Translational Vision Science & Technology*, 9(11):3–3.

Kamble, R., Samanta, P., and Singhal, N. (2020). Optic disc, cup and fovea detection from retinal images using u-net++ with efficientnet encoder. In Fu, H., Garvin, M. K., MacGillivray, T., Xu, Y., and Zheng, Y., editors, *Ophthalmic Medical Image Analysis*, page 93–103, Cham. Springer International Publishing.

Kang, H., Lee, S. J., Shin, H. J., and Lee, A. G. (2020). Measuring ocular torsion and its variations using different nonmydriatic fundus photographic methods. *PLOS ONE*, 15(12):1–11.

Li, H. and Chutatape, O. (2004). Automated feature extraction in color retinal images by a model based approach. *IEEE transactions on bio-medical engineering*, 51:246–54.

Li, T., Bo, W., Hu, C., Kang, H., Liu, H., Wang, K., and Fu, H. (2021). Applications of deep learning in fundus images: A review. *Medical Image Analysis*, 69:101971.

Martinez-Perez, M. and Espinosa-Romero, A. (2012). Three-dimensional reconstruction of blood vessels extracted from retinal fundus images. *Optics express*, 20:11451–65.

Moons, T., Gool, L., and Vergauwen, M. (2009). *3D Reconstruction from Multiple Images: Part 1 - Principles*. Now Publishers Inc.

Mutlu, F. and Leopold, I. H. (1964). The Structure of Human Retinal Vascular System. *Archives of Ophthalmology*, 71(1):93–101.

Nair, A. A., Liebenthal, R., Sood, S., Hom, G. L., Ohlhausen, M. E., Conti, T. F., Valentim, C. C. S., Ishikawa, H., Wollstein, G., Schuman, J. S., Singh, R. P., and Modi, Y. S. (2021). Determining the Location of the Fovea Centralis Via En-Face SLO and Cross-Sectional OCT Imaging in Patients Without Retinal Pathology. *Translational Vision Science & Technology*, 10(2):25–25.

Pineles, S. L. and Balcer, L. J. (2019). Visual loss: Optic neuropathies. In Liu, G. T., Volpe, N. J., and Galetta, S. L., editors, *Liu, Volpe, and Galetta's Neuro-Ophthalmology (Third Edition)*, pages 101–196. Elsevier, third edition edition.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2):155–162.

Reinhard, J., Messias, A., Dietz, K., Mackeben, M., Lakmann, R., Scholl, H. P., Apfelstedt-Sylla, E., Weber, B. H., Seeliger, M. W., Zrenner, E., and Trauzettel-Klosinski, S. (2007). Quantifying fixation in patients with stargardt disease. *Vision Res*, 47(15):2076–85.

Rohrschneider, K. (2004). Determination of the location of the fovea on the fundus. *Investigative ophthalmology & visual science*, 45:3257–8.

Sagar, A., S, B., and Chandrasekaran, V. (2007). Automatic detection of anatomical structures in digital fundus retinal images. In *Proceedings of IAPR Conference on Machine Vision Applications, MVA 2007*, pages 483–486.

Sekhar, S., Al-Nuaimy, W., and Nandi, A. (2008). Automated localisation of optic disk and in retinal fundus images. *European Signal Processing Conference*.

Semerád, L. and Drahanský, M. (2020). *Handbook of Vascular Biometrics*, chapter 11. Springer International Publishing.

Singh, J., Joshi, G. D., and Sivaswamy, J. (2008). Appearance-based object detection in colour retinal images. In *2008 15th IEEE International Conference on Image Processing*, pages 1432–1435.

Sinthanayothin, C., Boyce, J., Cook, H., and Williamson, T. (1999). Automated localization of the optic disc, fovea and retinal blood vessels from digital color fundus images. *The British journal of ophthalmology*, 83:902–10.

Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., and Van Ginneken, B. (2004). Digital retinal image for vessel extraction (drive) database. *Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands.*

Tan, J. H., Acharya, U. R., Bhandary, S., Chua, K., and Sivaprasad, S. (2017). Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *Journal of Computational Science*, 20.

Tarita-Nistor, L., Brent, M. H., Steinbach, M. J., and González, E. G. (2011). Fixation stability during binocular viewing in patients with age-related macular degeneration. *Invest Ophthalmol Vis Sci*, 52(3):1887–93.

Tarita-Nistor, L., Gill, I., González, E. G., and Steinbach, M. J. (2017). Fixation stability recording: How long for eyes with central vision loss? *Optom Vis Sci*, 94(3):311–316.

Tarita-Nistor, L., Gonzalez, E. G., Markowitz, S. N., and Steinbach, M. J. (2008). Fixation characteristics of patients with macular degeneration recorded with the mp-1 microperimeter. *Retina*, 28(1):125–33.

Timberlake, G. T., Sharma, M. K., Grose, S. A., Gobert, D. V., Gauch, J. M., and Maino, J. H. (2005). Retinal location of the preferred retinal locus relative to the fovea in scanning laser ophthalmoscope images. *Optom Vis Sci*, 82(3):177–85.

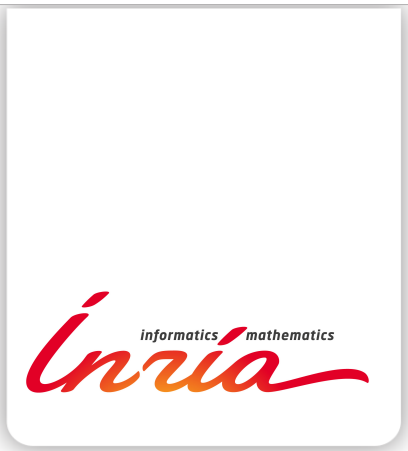
Tschumperle, D. and Deriche, R. (2005). Vector-valued image regularization with pdes: A common framework for different applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27:1–12.

Vullings, C. and Verghese, P. (2021). Mapping the binocular scotoma in macular degeneration. *J Vis*, 21(3):9.

Weickert, J. (1998). *Anisotropic Diffusion in Image Processing.* Teubner-Verlag, Stuttgart.

Xie, R., Liu, J., Cao, R., Qiu, C. S., Duan, J., Garibaldi, J., and Qiu, G. (2021). End-to-end fovea localisation in colour fundus images with a hierarchical deep regression network. *IEEE Transactions on Medical Imaging*, 40(1):116–128.

Zhang, Q., Shen, X., Xu, L., and Jia, J. (2014). Rolling guidance filter. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, page 815–830, Cham. Springer International Publishing.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399