



Fovea localization in retinal images using spatial color histograms

Jose Sigut¹ · Omar Nuñez¹ · Francisco Fumero¹ · Silvia Alayon¹  · Tinguaro Diaz-Aleman²

Received: 2 December 2022 / Revised: 8 June 2023 / Accepted: 4 July 2023
© The Author(s) 2023

Abstract

The automatic location of the fovea is very useful for diagnosing retinal diseases. It is a complex problem for which different solutions have been proposed based on classical image processing and Deep Learning techniques. The method presented in this paper is based on histograms that combine spatial and color information in such a way that the spatial coordinates are incorporated into conventional color histograms as an additional dimension. The binarization of these histograms retains a considerable amount of relevant information from the original image, allowing them to be processed as if they were ordinary images. This approach to the problem results in a simple, fast and effective method for locating the fovea. Different experiments have been carried out with three popular sets of images: Messidor, REFUGE1 and DIARETDB1, and a comparison was made with other state-of-the-art techniques. Our results show that the proposed method, despite its simplicity, is capable of surpassing many of these techniques.

Keywords Retinal diseases diagnosis · Fovea localization · Spatial color histogram

✉ Silvia Alayon
salayon@ull.edu.es

Jose Sigut
jfsigut@ull.edu.es

Omar Nuñez
omar@isaatc.ull.es

Francisco Fumero
ffumerob@ull.edu.es

Tinguaro Diaz-Aleman
vtdac@hotmail.com

¹ Department of Computer Science and Systems Engineering, University of La Laguna (ULL), Av. Astrofísico Francisco Sánchez, s/n, La Laguna, 38200 Santa Cruz de Tenerife, Spain

² Department of Ophthalmology, Canary Islands University Hospital, 38320 Santa Cruz de Tenerife, Spain

1 Introduction

An eye fundus image is a color image that provides highly detailed information about the state of the retina and its fundamental structures: the macula and fovea, the optic disc, and vascularization. This work focuses on the fovea, which is the central area of the macula. The fovea contains the highest concentration of cones in the area and is the part of the retina that provides the greatest visual acuity. The cells in the fovea are especially vulnerable to chronic disease. Any damage to the fovea can cause injury, decreased vision, or even blindness. The severity of some retinal diseases, such as Diabetic Maculopathy, is directly related to the presence of lesions in the vicinity of the fovea [25]. As a result, this area is of great interest in the study of various pathologies, and there are numerous works in the literature that propose methods to locate it automatically, as will be detailed in Section 2.

The solution proposed in this paper to address the problem of locating the fovea relies on using spatial color histograms to distinguish the fovea in the image and determine the coordinates of its center. Conventional histograms discard spatial information, and thus cannot be used directly to locate objects in images. The incorporation of spatial information makes it possible to avoid this inconvenience. Section 3 explains in detail how these histograms are constructed and the advantages they offer. Although the combination of spatial and color information is not new [1, 6, 19, 29, 31, 37], the proposed approach is novel. On the one hand, the spatial information is incorporated directly into the histogram as an additional dimension alongside the color components. On the other hand, we are working with binarized histograms, meaning they can be processed as if they were ordinary binary images, which makes it possible to apply any image processing technique to them. This is a general procedure that can be extended to other problems, which is why we think that it is one of the main contributions of this work.

The other main contribution is the method itself for locating the fovea which, as will be demonstrated in the following sections, is simple, fast, and effective. The results obtained are quite competitive not only against other methods that use handcrafted features but also with respect to methods that use Deep Learning. Deep Learning methods are capable of automatically learning the best features to solve the problem showing promising results, but they present several disadvantages. A high number of images is required to train the networks adequately. This high number of images is not always available in a field like Medicine. The computational cost associated to the training process is also high.

Furthermore, as pointed out in [24], they can also exhibit robustness problems if the images to be tested differ to some extent from the training images. The fact that our method only needs a few samples to adjust some of its parameters is an advantage in this regard.

The proposed method is explained in detail in Section 3, including a summary of the algorithm, already published, for locating the optic disc, which is used as the starting point. Section 4 presents a comparison with other state-of-the-art techniques, evaluated using three sets of retinal images: Messidor, REFUGE1, and DIARETDB1. The results obtained are discussed in Section 5. The conclusions are presented in Section 6.

2 Related work

The methods that address the problem of locating the fovea can be divided into two categories [40]: handcrafted anatomical feature-based methods [3–5, 7, 10–12, 14, 22, 25, 30, 36, 38, 39, 42] and Deep Learning-based methods [2, 18, 20, 21, 27, 34, 40].

The methods based on handcrafted features use the known anatomical information of the fovea to locate it:

- 1 The fovea is a circular area (with an approximate diameter of 1.5 mm) and free of vessels.
- 2 It is darker than the surrounding tissue.
- 3 The distance between the center of the optic disc and the fovea is approximately 2.5 times its diameter.

Taking advantage of this information, these types of methods apply classical image processing techniques to solve the problem. To this end, an important preliminary step in many of these methods is to locate the optic disc and segment the vessels to give an approximate initial location of the fovea. Along these lines, [12] uses the knowledge of the estimated distance between the fovea and the optic disc, and the vascular tree, to apply thresholding techniques. The available anatomical information of the optic disc and the vascular tree is also used in [3] to detect the fovea by means of morphological operations. In [7], filters are proposed that can detect semi-elliptical convex shapes. In [11], the image is preprocessed using cropping, green channel extraction, contrast enhancement and application of mathematical closing, before unsupervised clustering algorithms are applied to it. In [42], a directional local contrast technique is proposed and the position constraint information between the fovea and optic disc is used. In [14], the axis of symmetry that separates the lower and upper regions of the retina located in the middle of the major vessel arcades is detected, and the ROI is calculated using morphological operations. In [5], a template-matching technique, information from the vessels and the circular Hough transform are used to automatically segment the optic disc; from there, they estimate the region of the macula, where the fovea is located, using various morphological operations. In [39] techniques are used that rely on known anatomical constraints on the relative locations of retinal structures, and mathematical morphology, to first locate and calculate the diameter of the optic disc, and then estimate the macular region. [30] proposes an assembly of several classifiers of different natures (edge detectors, based on entropy, on the Hough transform, and others), the combined response of which improves the results of locating the fovea. In [36], the optic disc is identified by regarding it as the area with the greatest intensity in the image. The vessels are located with a multilayer perceptron neural net, and the fovea is identified by analyzing typical characteristics of a fovea, for example, the darkest area in the neighborhood of the optic disc. In [25], a two-stage method is proposed: first, the image is pre-processed to remove the optic disc, and then the macula and fovea are located using the intensity property of a processed red-plane image. In [22], the fovea in fundus images is automatically detected by applying an adaptive Gaussian template in the vessel-free area of the image. [38] proposes a method that combines a new set of features with a minimum distance classifier to accurately locate the fovea. Blood vessel segmentation and fovea localization are done simultaneously based on the Convexity Shape Prior (CSP) algorithm in [10]. More recently, [4] presents a new methodology to simultaneously segment the optic disc (OD) and fovea using an OD-fovea model and an evolutionary algorithm.

In general, methods based on handcrafted features strongly depend on the properties of the images considered (illumination, contrast, presence of artifacts, ...), allowing to take advantage of the anatomical knowledge a priori, as well as to detect structures such as the optic disc and vessels. These are methods that do not require a large number of examples to adjust their parameters. The proposed method would fall into this category.

The rapid development of Deep Learning techniques in recent years has also allowed them to be applied to automatically locate the fovea. The goal in [2] is to simultaneously detect the optic disc and fovea by using a deep multiscale sequential convolutional neural network. [34] uses two convolutional networks, one to generate the ROI that contains the fovea (coarse network), and the other to obtain the final location (fine network). [20] proposes a network to locate the optic disc (region proposal network), and another to locate the fovea (a three-level cascaded convolutional neural network), taking into account the geometric relationship between the disc and fovea. In [40], a hierarchical coarse-to-fine deep regression neural network is used to locate the fovea. In [18], an end-to-end encoder-decoder network (DRNet) is proposed to segment and locate the centers of the optic disc and fovea. [21] presents a method that detects the disc and fovea at the same time by using a modified U-Net++ architecture with the EfficientNet-B4 model as a backbone. [27] reformulates the segmentation task as a pixel-wise regression task to also segment the disc and cup, simultaneously applying a U-Net deep network. It must be said that of all the Deep Learning methods discussed, some of them use transfer learning to adapt pretrained models to the task they want to solve. Specifically, [34, 40] use VGG-based architectures and [20] is based on the Resnet50 model.

In general, methods based on Deep Learning do not depend on anatomical features or retinal landmarks, which is an advantage over methods based on hand-crafted features. On the other hand, the fact of requiring large sets of images to train the models and the associated computational cost can be significant handicaps in practice.

3 Material and methods

3.1 Binarized spatial color histograms

The combined use of spatial and color information in images has been a common practice for many years in the field of computer vision. [29] presents the Color Coherence Vectors technique, which allows differentiating pixels not only by color, but also by texture, location, etc. Another technique worth noting is the Color Correlogram, presented in [19]. [31] proposes a technique to add information about the spatial distribution of pixels in the image, of which three variants are discussed: annular color histogram, angular color histogram, and hybrid histogram, which combines the two techniques. In [6], the authors propose a way of combining spatial and color information called Spatial-Chromatic Histogram (SCH). Another measure is the one presented in [37], called Color Distribution Entropy (CDE). In 2011, [1] presented a variant of this technique, which they call DCDEN.

All the techniques reviewed in the paragraph above, in one way or another, use histograms to combine spatial and color information. Accordingly, the proposal we present in this paper starts from the idea of a conventional color histogram to incorporate, in a natural and direct way, spatial information as one more dimension. Specifically, we show how a binarized two-dimensional histogram obtained as a function of a spatial coordinate and a color component is particularly intuitive and easy to handle, while at the same time

providing much richer information about the content of the image than the original color histogram.

We start from the definition of a conventional color histogram, in which it is convenient to view an image as the realization of a random field modeled by a spatially arranged, three-dimensional random variable c . In this way, the color histogram H of an image I , with $M \times N$ pixels can be defined as an estimate of the probability function of this variable:

$$H(c) = \frac{1}{M \times N} \sum_{x=1}^N \sum_{y=1}^M \delta(I(x, y) - c) \tag{1}$$

where δ is defined as:

$$\delta(v) = \begin{cases} 1, & v = (0, 0, 0) \\ 0, & \text{elsewhere} \end{cases} \tag{2}$$

It should be noted that this definition is entirely general and, therefore, applicable to the case of a monochromatic standard histogram, in which c would become a one-dimensional random variable.

We can use the definition of a color histogram to define the spatial color histogram, since the only difference lies in considering the spatial components as random variables also, such that the expression in (1) can be written in its more general form:

$$H(c, x) = \frac{1}{M \times N} \sum_{y=1}^M \delta(I(x, y) - c) \tag{3}$$

$$H(c, y) = \frac{1}{M \times N} \sum_{x=1}^N \delta(I(x, y) - c) \tag{4}$$

where the function δ is defined as in (2). For convenience, in what follows, when referring to spatial color histograms, the definitions in (3) and (4) will be assumed, but without dividing by $M \times N$.

To make this type of histogram more manageable and intuitive, it may be convenient, in practice, to take the spatial variable as x or y , and make c equal to a given color component. The resulting histogram can be seen as an image to which we can apply operations, such as binarization, in accordance with:

$$H^B(c, x) = H(c, x) > T_x \tag{5}$$

$$H^B(c, y) = H(c, y) > T_y \tag{6}$$

where T_x and T_y represent the chosen thresholds, whose specific value may depend on the concrete problem in question, although oftentimes, a very small value turns out to be the most appropriate to retain almost all the information of interest while removing some of the noise. In the example shown in Fig. 1, these thresholds have been assigned a value of 1.

The example image has nothing to do with fovea localization, but it has been chosen to show the generality of the technique presented. It is particularly interesting to observe how the binarized histograms, despite their simplicity, retain interesting and useful information about the original image, information that is much richer than that provided by the conventional color histogram. Indeed, the spatial color histograms make it possible to distinguish the four most relevant objects in the image, even though some of them are very small: the camel, the sand, the sky, and the sun. Furthermore,

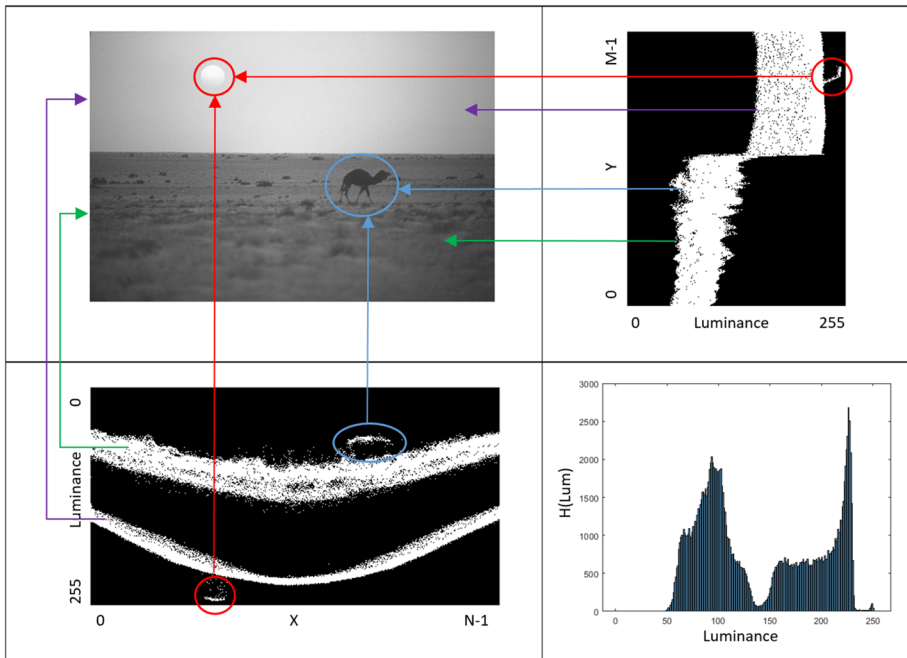


Fig. 1 Binarized spatial color histograms (camel image taken from <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>)

the histogram directly provides the approximate location of the objects in the x and y spatial coordinates, while preserving the vertical and horizontal distances. This property has been used to estimate the fovea location, as will be explained in the next section. There is even information about the spatial distribution of the lighting in the image, which, as we can see, is not uniform; rather, there is a slight gradient, with higher values towards the center of the image on the x -axis. Figure 2 shows the binarized spatial color histograms for an example retinal image.

3.2 Proposed fovea localization method

As commented in Section 2, most of the handcrafted feature-based methods make use of a priori anatomical information to limit the search space in which to locate the fovea. One of the anatomical features that is usually considered is the fact that the distance between the center of the optic disc (OD) and the center of the fovea is approximately 2.5 times the OD diameter following the horizontal raphe of the retina, that is, the line of symmetry separating the superior and inferior retinal regions [14]. We have also taken advantage of this feature to obtain an initial estimate of the fovea center as described in Section 3.2.1. In Section 3.2.2, the method itself for accurate fovea localization is explained based on the binarized spatial color histograms presented in Section 3.1.

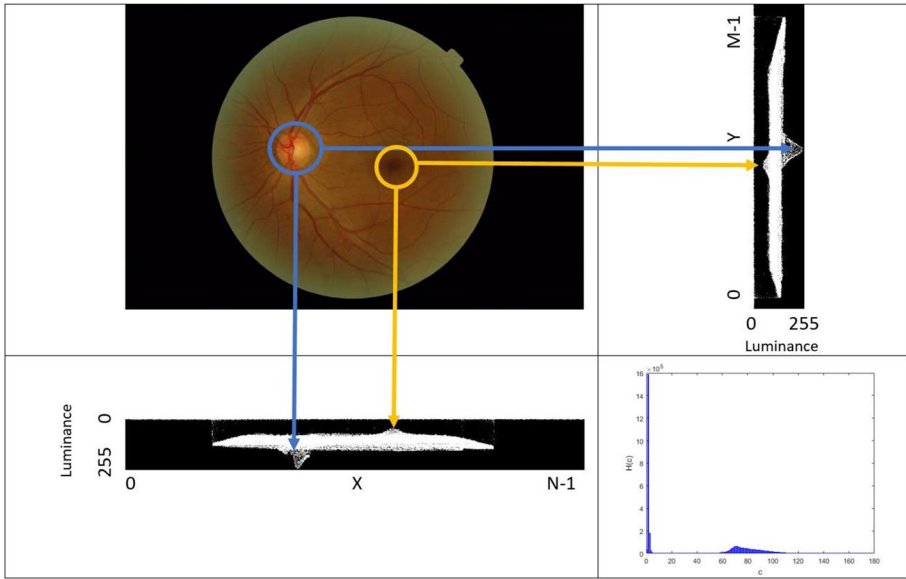


Fig. 2 Binarized spatial color histograms for a retinal image

3.2.1 Approximate fovea localization

Our procedure for OD localization was already published and described in detail in [35]. For the sake of completeness, a summary of this procedure is included in this subsection. The method consists of two main steps: creating a mask based on vascular information to shrink the search space and filtering the image with a detector which combines vascular and brightness information.

The first step exploits the fact that the OD is the entry point for the major blood vessels that supply the retina. Consequently, the OD region usually exhibits high vessel density, and can also be seen as a convergence point of this structure. The high vessel density is captured by filtering and thresholding the vessel image. The convergence of the branches of the vascular tree is estimated by finding the intersections of lines that are used to approximate the branches of this tree. The lines are the result of applying the Hough Transform to the output of a Canny edge detector computed using a vessel-enhanced image. The final constraint mask is obtained as a logical AND of the vessel density image and a thresholded version of the vessel convergence image. Figure 3 shows the constraint mask for a sample input image.

In the second step, an OD detector is obtained as the difference between the output of two averaging filters computed using the intensity image: a circular averaging filter and a rectangular averaging filter. This detector is only applied to those parts of the image preserved by the constraint mask. The coordinates of the OD center are those corresponding to the pixel with the maximum value provided by the detector, as shown in Fig. 4.

As in [12], the OD diameter, D , is obtained as:

$$D = 0.15D_{FOV} \tag{7}$$

where D_{FOV} is the retinal diameter.

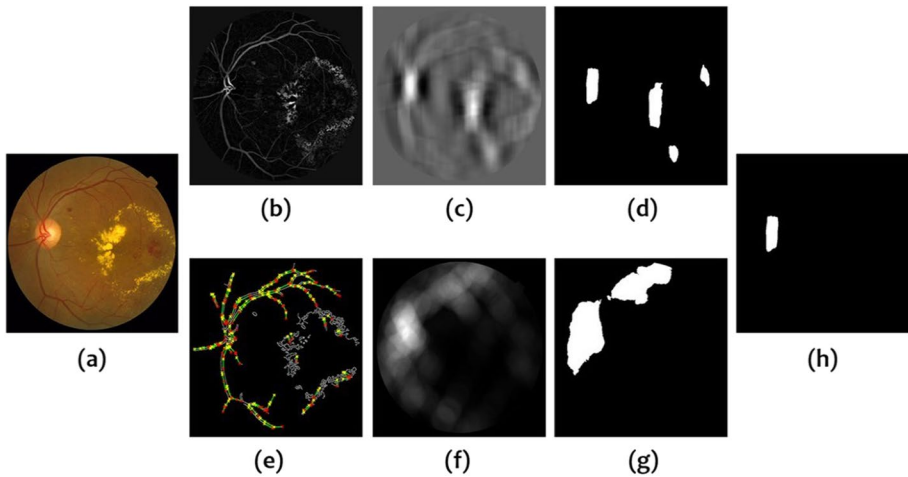


Fig. 3 **a** Original image, **b** Vessel enhanced image, **c** Vessel density image, **d** Vessel density mask, **e** Hough Transform of the Canny edge detector output, **f** Vessel convergence image, **g** Vessel convergence mask, **h** Logical AND of (d) and (g)

Regarding the orientation of the fovea with respect to the OD, instead of estimating the raphe, a simpler approach is carried out based on the vessel convergence image so that the initial location of the fovea is on a horizontal line at 2.5 times the OD diameter, measured in the opposite direction of maximum vessel density.

3.2.2 Accurate fovea localization

Once the fovea is approximately located, the following steps are used to determine the coordinates of its center more accurately, as shown in the example in Fig. 5:

1. The original image is cropped using a rectangular window centered in the approximate location of the fovea. The size and position of the rectangle, $[y_{min}, x_{min}, width, height]$, are given by:

$$y_{min} = y_{OD} - \frac{D}{4} \quad (8)$$

$$x_{min} = x_{OD} \pm 2.5D - \frac{D}{3} \quad (9)$$

$$width = \frac{2D}{3} \quad (10)$$

$$height = \frac{3D}{4} \quad (11)$$

where x_{OD} , y_{OD} are the estimated coordinates of the OD center. We are left with the G channel, which offers, in general, a better contrast. It should be noted that, despite

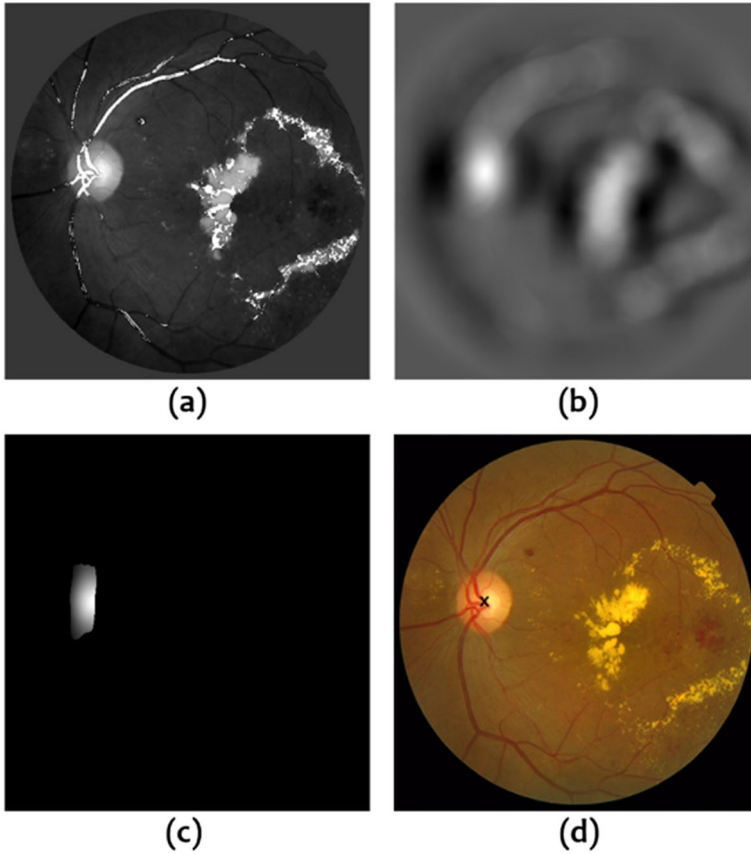


Fig. 4 **a** Intensity image with superimposed vessel mask, **b** Output of OD detector, **c** OD detector combined with constraint mask, **d** Estimated OD center

the crop, the resulting image is still of considerable size. The one in the example is 259x155 pixels.

2. The spatial color histograms for the x and y coordinates are computed, in combination with the G component. Previous to this, a Gaussian smoothing filter with variance 1 is applied.
3. The histograms are binarized as per (5) and (6) with $T_x, T_y = 2$, and only the largest connected components, $H_{CCmax}^B(c, x)$ and $H_{CCmax}^B(c, y)$, are retained. Both operations pursue the objective of eliminating any noise that may interfere with the calculation of interest.
4. This step takes advantage of another anatomical feature of the fovea, namely that it usually appears as a dark area relative to its surroundings. For this reason, the coordinates of its center, x_c, y_c , can be identified as those which correspond to the extreme points on $H_{CCmax}^B(c, x)$ and $H_{CCmax}^B(c, y)$ with the lowest G values, i.e., those that satisfy:

$$x_c = x / (G_{min}, x) \in H_{CCmax}^B(c, x) \tag{12}$$

$$y_c = y / (G_{min}, y) \in H_{CCmax}^B(c, y) \tag{13}$$

Fig. 5 **a** Approximate fovea location and rectangular window (step 1), **b** spatial color histograms computed for the cropped image (step 2), **c** binarized spatial color histograms (step 3), **d** $H_{CCmax}^B(c, x)$ and $H_{CCmax}^B(c, y)$ are used to obtain the coordinates of the center of the fovea (step 4), **e** accurate fovea location in the original image

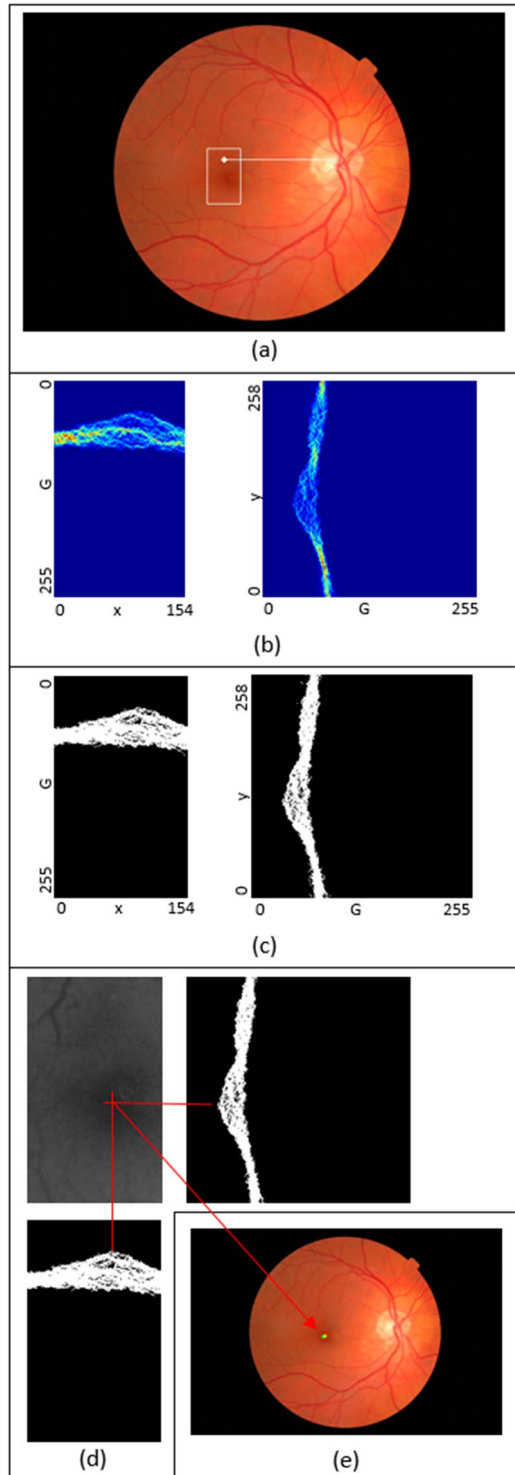


Figure 5 shows how x_C and y_C are shifted with respect to the initial estimate. If there is more than one value of x , y that meets the conditions in (12) and (13), we are left with an average value. These values must be scaled based on the actual size of the image to obtain the final fovea location.

3.3 Datasets and methodology for evaluation

Three datasets have been used to evaluate the proposed fovea localization method: Messidor, REFUGE1 and DIARETDB1.

The Messidor dataset [8] was created in the framework of Diabetic Retinopathy screening and diagnosis. This dataset can be downloaded from [26]. It contains 1200 images acquired in three different ophthalmology departments using a 3CCD color video camera on a Topcon TRC NW6 non-mydratic retinograph with a 45° FOV and three different resolutions: 1440×960, 2240×1488 and 2304×1536 pixels. To the best of our knowledge, it is the only dataset for which a ground truth of fovea locations is publicly available¹⁰. Because of this, it has become the most widely used benchmark for this problem, though the x , y coordinates of the center of the fovea are not provided for the whole set, but only for 1136 images.

As part of REFUGE1 (Retinal Fundus Glaucoma Challenge) [28], a dataset of 1200 images was published for the participants in the event. This dataset can be downloaded from [32]. We deemed its inclusion in this work appropriate because Diabetic Retinopathy and Glaucoma are two of the most important pathologies that affect the retina. The ground truth for these images was built using the fovea center positions manually marked by an ophthalmologist with 14 years of experience. Since most of the methods proposed in the challenge were based on Machine Learning, the dataset was split into three subsets—for training (2124×2056 pixels), validation (1634×1634 pixels) and testing (1634×1634 pixels)—each composed of 400 samples. The proposed method has been evaluated using the test set.

DIARETDB1 is a popular dataset for benchmarking Diabetic Retinopathy detection from digital images [23]. This dataset can be downloaded from [9]. It consists of 89 images with a resolution of 1500×1152 pixels and has been included in our study because some researchers have also used it for fovea localization. However, as far as we know, there is no public ground truth available for this data, so the published results are based on annotations carried out by the same expert as in the case of the REFUGE 1 dataset.

Regarding the evaluation methodology, we will follow [12], where the error in fovea localization is calculated as the Euclidean distance, $D(c_{exp}, c_{real})$, between the real fovea coordinates and the experimental ones. Since the size of the retinal images under consideration may be different, a normalized distance, $D^*(c_{exp}, c_{real})$, becomes a more convenient measure:

$$D^*(c_{exp}, c_{real}) = \frac{D(c_{exp}, c_{real})}{D_{FOV}} 100 \quad (14)$$

where D_{FOV} is the retinal diameter as in (7).

In order to compare our technique with other methods, a usual way to evaluate the accuracy obtained consists of counting the number of cases where is less than (1/8)R, (1/4)R, (1/2)R and R, where the OD radius, R, is calculated as D/2, as per (7). This methodology has become the standard for evaluating accuracy in fovea localization.

Table 1 General performance of the proposed method

	(1/8)R (%)	(1/4)R (%)	(1/2)R (%)	R (%)	D* Mean (St. Dev.)
Messidor	58.71	92.69	98.94	99.56	1.0231 (1.1918)
REFUGE 1	66.25	95.50	98.25	99.00	0.9917 (1.3376)
DIARETDB1	26.97	55.06	84.27	95.51	2.4933 (2.9083)

4 Results

Several experiments have been developed to: 1) assess the performance of the proposed method, 2) analyze the influence of the method's parameters on its performance, 3) compare it with other existing methods using the same datasets, and 4) compare their computational times.

Table 1 shows the result of applying our method to the three sets of images, using the evaluation methodology described in the Section 3.3.

The main parameters on which the proposed method depends are the location of the center of the disc, the initial estimate of the center of the fovea, the dimensions of the cropping box, and the binarization threshold for the histograms. It is important to note that these parameters were adjusted beforehand using a set of internal images, and their values were kept fixed in all the experiments conducted with the three sets of images considered. Table 2 shows how some of these parameters influenced the performance of the method with the Messidor images.

The comparison with other methods was made based on the data available in each case. As discussed in Section 3.3, the existence of a public ground truth for the coordinates of the center of the fovea makes Messidor the most widely used image set for this problem. Table 3 shows the results obtained by our method and some cutting-edge methods based on both classical computer vision techniques and Deep Learning, in the case of Al-Bander et al. [2], Huang et al. [20], Xie et al. [40], and Meyer et al. [27]. Since many studies do not provide the results for (1/8)R and D^* , it was decided not to include them in the table. Although the published ground truth is for 1136 images, some authors report performance values for the total of 1200 images, and others only for a subset of 800 images. The latter were not considered for comparison purposes.

In the case of REFUGE1, since no results published in this format are available, for comparison purposes, some methods that rely on Convolutional Neural Networks (CNN) to

Table 2 Influence of the method's parameters on its general performance

	(1/8)R (%)	(1/4)R (%)	(1/2)R (%)	R (%)	D* Mean
Original method	58.71	92.69	98.94	99.56	1.0231
OD manual localization	58.71	92.69	99.12	99.65	1.0065
Fovea approximate localization	1.23	5.99	27.64	85.21	5.2303
Tx, Ty = 0	57.92	91.11	98.50	99.38	1.0755
Tx, Ty = 1	59.24	92.17	98.68	99.38	1.0394
Tx, Ty = 3	51.06	89.26	98.86	99.65	1.1221
Tx, Ty = 4	41.55	81.60	98.15	99.65	1.3112

Table 3 Performance comparison on Messidor database

	(1/4)R (%)	(1/2)R (%)	R (%)
Our method	92.69	98.94	99.56
Gegundez-Arias et al. [12]	76.32	93.84	98.24
Aquino [3]	83.01	91.28	98.24
Dashtbozorg et al. [7]	66.50	93.75	98.87
GeethaRamani et al. [11]	85.00	94.08	99.33
Zheng et al. [42]	91.36	98.32	99.03
Guo et al. [14]	91.91	98.50	99.58
Al-Bander et al. [2]	66.80	91.40	96.60
Huang et al. [20]	70.10	89.20	99.25
Xie et al. [40]	98.15	99.74	99.82
Meyer et al. [27]	94.01	97.71	99.74
Carmona & Molina [4]	94.10	97.30	99.00

segment the fovea were implemented. Accordingly, the coordinates of the center of the segmented region were taken as the estimated fovea coordinates. To train these networks, the training and validation sets mentioned in Section 3.3 were used; specifically, the networks used were the well-known U-Net [33] and the Pyramid Scene Parsing Network (PSPNet) [41]. The encoder in the U-net network can be implemented with a pre-trained CNN. In our case, we opted for ResNet50-Unet and VGG-Unet. Similarly, we combined the PSP module with the same type of CNN to obtain a ResNet50-ppspNet and a VGG-ppspNet.

More specifically, the training of these networks was performed using the Python library *keras-segmentation* v0.3.0 [15, 16], running on Jupyter Notebook with a TensorFlow backend v1.15.4, inside a Docker container. The docker image tagged as tensorflow/tensorflow:1.15.4-gpu-py3-jupyter was used. A Geforce GTX 1080 Ti GPU with 11 GB of RAM was used to accelerate the computations. Additionally, to be able to run the *keras-segmentation* library properly, we had to install *Keras* v2.3.1 and *opencv-python* v4.1.2.30. The input size was set to 576×576 , so the input images were re-sized accordingly. The fovea masks were generated by drawing a circle of 50 pixels radius centered on the provided ground truth. Two different labels were assigned to the fovea region and the background. The models were trained for 40 epochs using the data augmentation function called *aug_all* [17], which consists of random geometric and non-geometric transformations. The rest of the parameters were set to the defaults provided by the library (no preprocessing of the input images, batch size of 2, 512 steps per epoch, categorical_crossentropy as the loss function, and an Adam optimizer). The pre-trained weights from ImageNet were used, both for VGG and ResNet50 CNNs. By default, the library tunes all the layers of the model. The validation subset of the REFUGE1 dataset was used to choose the model weights from the epoch that yielded the best results for the segmentation of the fovea based on the Intersection over Union (IoU) metric. Table 4 shows the results obtained.

Regarding the DIARETDB1 image set, Table 5 shows the results for distance R with different methods. No results are reported for the other distances considered.

Finally, Table 6 shows a comparison of times published with other methods.

The codes for calculating the spatial color histogram and the location of the fovea were implemented in Matlab and are available at [13]. The CNNs for REFUGE 1 were trained using Python-KERAS as explained above.

Table 4 Performance comparison on REFUGE1 database

	(1/8)R (%)	(1/4)R (%)	(1/2)R (%)	R (%)	D* Mean
Our method	66.25	95.50	98.25	99.00	0.9917
ResNet50-ppspNet	67.50	98.50	1	1	0.7805
ResNet50-Unet	87.00	99.00	99.25	99.50	0.9669
VGG-ppspNet	55.75	93.75	98.25	98.25	1.2268
VGG-Unet	86.25	94.50	94.75	95.00	1.4717

5 Discussion

Analyzing the general performance results of our method (Table 1), the accuracy values obtained for Messidor and REFUGE1 are similar, and exhibit a certain drop in performance for (1/8)R. This behavior could be explained by the fact that the end points of H_{CCmax}^B are directly taken as the solution to the problem, as per (12) and (13). Any small disturbance in the image that slightly alters this end value may affect the accuracy for (1/8)R, and to a lesser extent that of the other reference values. If more accuracy is needed at that level, some type of processing could be done to smooth out the binary image before looking for the extreme points.

In the case of DIARETDB1, the success of the method is a little lower as these images are generally of poorer quality, and the assumption that the fovea is an area that is differentiated by being dark with respect to its surroundings is not satisfied at times. It is also a much smaller sample of images.

In Fig. 6, we have selected some of the common situations that can occur in practice for both correct and incorrect localizations. In (a), we see how taking the largest connected region in $H^B(c, x)$ prevents the minimum value of G from being reached for an incorrect value of x_c due to the presence of noise. (b) shows the robustness of the method against the presence of exudates, since, given that these pixels have high G values, they do not affect the part of the histogram that is used to obtain the final result. In general, the method is robust against bright degenerations but the presence of dark degenerations in the area of the fovea could affect, in some cases, the accuracy of the location, since there could be a certain overlap with the part of the histogram that

Table 5 Performance comparison on DIARETDB1 database

	R (%)
Our method	95.51
Chalakkal et al. [5]	95.51
Welfer et al. [39]	92.13
Aquino [3]	94.38
Qureshi et al. [30]	98.74
Sinthanayothin et al. [36]	65.17
Medhi et al. [25]	95.51
Kao et al. [22]	92.10
Syed et al. [38]	94.38
Guo et al. [14]	97.63
Carmona & Molina [4]	100.00

Table 6 Computation time comparison

	Time (sec.)
Our method	1.4083 (0.1860)
Chalakkal et al. [5]	25
Al-Bander et al. [2]	0.007
Aquino [3]	10.88
Guo et al. [14]	10.4
Carmona & Molina [4]	29.35 (0.91)

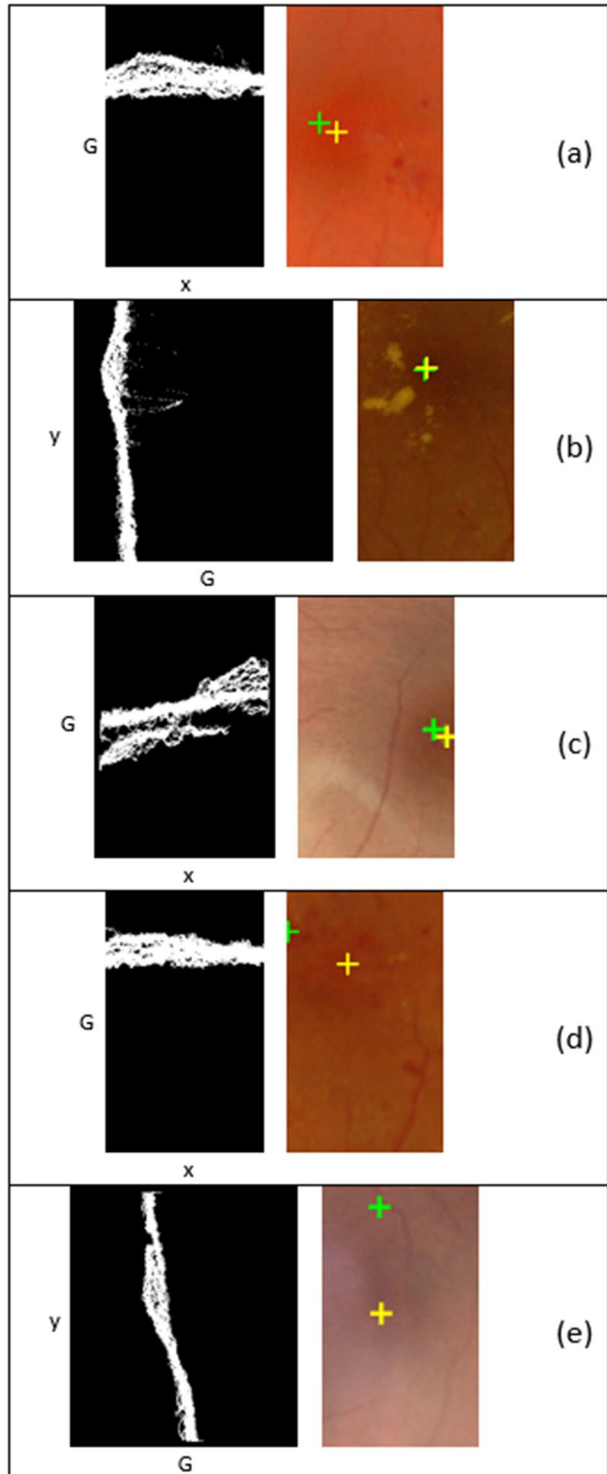
occupies the fovea. In (c), the cropping box leaves out part of the fovea, but despite this, it is possible to correctly determine the coordinates of its center. (d) shows an example of incorrect localization, especially in x . The error stems from the presence of some pixels that are connected to the main region in $H_{CCmax}^B(c, y)$, which alter the value of x_c . Finally, in (e), another situation is shown that leads to an error in the calculation of the coordinates. In this case, the problem comes from the existence of a significant gradient in the values of G , which causes the minimum of G to be reached for an incorrect value of y_c .

Regarding to the study of how the method's parameters influenced its performance (Table 2), we see that the manual localization of the center of the optic disc barely changes the results with respect to the automatic localization. This seems logical since it only influences the initial estimate of the fovea's location. However, if we take this initial estimate as the final one, we can see how the accuracy drops sharply, highlighting the importance of using histograms to locate the fovea much more precisely. Regarding the threshold values for the binarization of the spatial color histograms, for $T_x, T_y < 2$, there is not much difference with respect to $T_x, T_y = 2$. For $T_x, T_y > 2$, we start to see a slight drop in the accuracy, especially for (1/8)R and (1/4)R. This can be explained by the elimination of not only noise, but also of significant parts of the histogram that end up affecting the calculations. With regard to the dimensions of the cropping box, there are many possible variants, which is why we will only comment qualitatively on its influence. In essence, what we have observed is that a window that is too small sometimes leaves out the fovea, which leaves no option to locate it correctly. By contrast, an overly large window entails the presence of more noise and elements that can significantly complicate the localization.

In the performance comparison using Messidor (Table 3), note that the accuracy of our method exceeds that of the methods based on classical techniques, except for R in Guo et al. [14], which surpasses ours very slightly, and Carmona and Molina [4] for (1/4)R. In general, the best results are obtained with the method of Xie et al. [40], although our method manages to overcome two methods based on Deep Learning, and is slightly below the one proposed by Meyer et al. [27], although it exceeds it for (1/2)R.

With the REFUGE1 data, as shown in Table 4, the networks that used ResNet50 as the encoder outperform the proposed method. However, our method performs better than the networks using VGG with the exception of VGG-Unet for (1/8)R. It is not surprising to find that techniques based on Deep Learning provide the best results, as we already saw this in the case of Messidor. This is consistent with the success achieved when applying these techniques to many other problems in the field of computer vision. However, as commented in Section 2, the availability of training samples and the robustness against test images captured in different conditions can be an issue.

Fig. 6 A yellow cross indicates ground truth, green cross indicates fovea location as estimated by our method. **a** Correct localization. The noise can be discarded by taking the largest connected component, **b** correct localization even in the presence of exudates, **c** correct localization with partially occluded fovea, **d** incorrect localization due to the presence of noise, **e** incorrect localization due to gradient in G values



In the comparison with DIARETDB1, the result obtained with our method is quite competitive, being surpassed only by the methods of Qureshi et al. [30], Guo et al. [14] and Carmona and Molina [4].

Regarding computation times, it should be noted that the calculation of the histograms in itself is very fast, on the order of milliseconds with Matlab. What slows down the localization of the fovea considerably is the initial determination of the center of the optic disc. As shown in Table 6, the total time of the proposed method is the lowest, by far, only surpassed by Al-Bander et al. [2], which uses Deep Learning. In that case, most of the time is consumed training the neural networks. Once trained, the calculation is very fast.

6 Conclusions

This paper presents a simple method for determining the center of the fovea based on spatial color histograms. Despite its simplicity, the experiments carried out and the comparison with other state-of-the-art techniques show that it is an effective and fast procedure that is capable of surpassing many of these techniques, even some based on Deep Learning. The authors believe that spatial color histograms can be a valuable tool for other types of problems in the field of Medical Image Processing, such as image enhancement, segmentation, and any other in which conventional color histograms can be replaced by these more powerful versions.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by José Sigut, Omar Nuñez, Francisco Fumero, Silvia Alayon and Tinguaro Diaz-Aleman. The first draft of the manuscript was written by José Sigut and Silvia Alayon and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data Availability All data sets used in this article are publicly available data.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alamdar F, Keyvanpour M (2011) A new color feature extraction method based on dynamic color distribution entropy of neighborhoods. *Int J Comput Sci Issues*. 8(5)1:42–48. arXiv:1201.3337
2. Al-Bander B, Al-Nuaimy W, Williams BM, Zheng Y (2018) Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. *Biomed Signal Process Control* 40:91–101. <https://doi.org/10.1016/j.bspc.2017.09.008>

3. Aquino A (2014) Establishing the macular grading grid by means of fovea centre detection using anatomical-based and visual-based features. *Comput Biol Med* 55:61–73. <https://doi.org/10.1016/j.compbiomed.2014.10.007>
4. Carmona EJ, Molina-Casado JM (2020) Simultaneous segmentation of the optic disc and fovea in retinal images using evolutionary algorithms. *Neural Comput Appl* 1–19:2020. <https://doi.org/10.1007/s00521-020-05060-w>
5. Chalakkal RJ, Abdulla WH, Thulaseedharan SS (2018) Automatic detection and segmentation of optic disc and fovea in retinal images. *IET Image Process* 12(11):2100–2110. <https://doi.org/10.1049/iet-ipl.2018.5666>
6. Cinque L, Ciocca G, Levaldi S, Pellicanò A, Schettini R (2001) Color-Based Image Retrieval Using Spatial-Chromatic Histograms. *Image Vis Comput* 2:969–973. <https://doi.org/10.1109/MMCS.1999.778621>
7. Dashtbozorg B, Zhang J, Huang F, and ter Haar Romeny BM (2016) Automatic optic disc and fovea detection in retinal images using super-elliptical convergence index filters. In: Campilho A., Karray F. (eds) *Image Analysis and Recognition. ICIAR 2016. Lecture Notes in Computer Science*, vol 9730. Springer, Cham
8. Decencière E et al (2014) Feedback on a publicly distributed database: the Messidor database. *Image Anal Stereol* 33(3):231–234. <https://doi.org/10.5566/ias.1155>
9. DIARETDB1 dataset: <https://academictorrents.com/details/817b91fd639263f6f644de4ccc9575c20b005c6c>. Accessed 08/07/2023
10. Escorcía-Gutiérrez J, Torrents-Barrena J, Gamarra M, Romero-Aroca P, Valls A, Puig D (2020) Convexity shape constraints for retinal blood vessel segmentation and foveal avascular zone detection. *Comput Biol Med*. 127. <https://doi.org/10.1016/j.compbiomed.2020.104049>
11. GeethaRamani R, Balasubramanian L (2018) Macula segmentation and fovea localization employing image processing and heuristic based clustering for automated retinal screening. *Comput Methods Programs Biomed* 160:153–163. <https://doi.org/10.1016/j.cmpb.2018.03.020>
12. Gegundez-Arias ME, Marin D, Bravo JM, Suero A (2013) Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques. *Comput Med Imaging Graph* 37(5–6):386–393. <https://doi.org/10.1016/j.compmedimag.2013.06.002>
13. Github of Medical Image Analysis Group (MIAG): <https://github.com/miag-ull>. Accessed 08/07/2023
14. Guo X, Wang H, Lu X, Hu X, Che S, Lu Y (2020) Robust Fovea Localization Based on Symmetry Measure. *IEEE J Biomed Health Inform* 24(8):2315–2326. <https://doi.org/10.1109/JBHI.2020.2971593>
15. Gupta D (2020) Implementation of various Deep Image Segmentation models in keras. *GitHub*. <https://github.com/divamgupta/image-segmentation-keras>. Accessed 08/07/2023
16. Gupta D (2020) Image Segmentation toolkit for Keras. *PyPI*. <https://pypi.org/project/keras-segmentation/>. Accessed 08/07/2023
17. Gupta D (2020) Image augmentation model. *GitHub*. <https://bit.ly/keras-segmentation-image-augmentation-model>. Accessed 08/07/2023
18. Hasan K, Alam A, Elahi TE, Roy S, Martí R (2021) DRNet: Segmentation and localization of optic disc and Fovea from diabetic retinopathy image. *Artif Intell Med*. 111. <https://doi.org/10.1016/j.artmed.2020.102001>
19. Huang J, Kumar SR, Mitra M, Zhu WJ, Zabih R (1997) Image indexing using color correlograms. *Proc. of Computer Vision and Pattern Recognition*. 762–768. <https://doi.org/10.1109/CVPR.1997.609412>
20. Huang Y, Zhong Z, Yuan J, Tang X (2020) Efficient and robust optic disc detection and fovea localization using region proposal network and cascaded network. *Biomedical Signal Processing and Control*. 60. <https://doi.org/10.1016/j.bspc.2020.101939>
21. Kamble R, Samanta P, Singhal N (2020) Optic Disc, Cup and Fovea Detection from Retinal Images Using U-Net++ with EfficientNet Encoder. In: H. Fu, M.K. Garvin, T. MacGillivray, Y. Xu, Y. Zheng (eds) *Ophthalmic Medical Image Analysis. OMIA 2020. Lecture Notes in Computer Science*, vol. 12069. Springer, Cham
22. Kao EF, Lin PC, Chou MC, Jaw TS, Liu GC (2014) Automated detection of fovea in fundus images based on vessel-free zone and adaptive Gaussian template. *Comput Methods Programs Biomed* 117(2):92–103. <https://doi.org/10.1016/j.cmpb.2014.08.003>
23. Kauppi T, Kalesnykiene V, Kamarainen JK, Lensu L, Sorri I, Raninen A, Voutilainen R, Usitalo H, Kälviäinen H and Pietilä J (2007) DIARETDB1 diabetic retinopathy database and evaluation protocol. *Proc. of the 11th Conf. on Medical Image Understanding and Analysis*. <https://doi.org/10.5244/C.21.15>
24. Mahony NO, Campbell S, Carvalho A, Harapanahalli S, Velasco-Hernández GA, Krpalkova L, Riordan D, Walsh J (2019) Deep learning vs. traditional computer vision. *Advances in Computer*

- Vision Proceedings of the 2019 Computer Vision Conference (CVC). Springer Nature Switzerland AG, pp. 128–144. arXiv:1910.13796. <https://doi.org/10.1007/978-3-030-17795-9>
25. Medhi JP, Dandapat S (2016) An effective fovea detection and automatic assessment of diabetic maculopathy in color fundus images. *Comput Biol Med* 74:30–44. <https://doi.org/10.1016/j.combiomed.2016.04.007>
 26. Messidor dataset: <https://www.adcis.net/en/third-party/messidor/>. Accessed 08/07/2023
 27. Meyer MI, Galdran A, Mendonça AM, Campilho A (2018) A pixel-wise distance regression approach for joint retinal optical disc and fovea detection. *Proc. Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference*. 39–47. https://doi.org/10.1007/978-3-030-00934-2_5
 28. Orlando JI et al (2020) REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med Image Anal* 59:101570. <https://doi.org/10.1016/j.media.2019.101570>
 29. Pass G, Zabih R, Miller J (1996) Comparing images using color coherence vectors. *Proc. of the fourth ACM International Conference on Multimedia*. 65–73. <https://doi.org/10.1145/244130.244148>
 30. Qureshi RJ, Kovacs L, Harangi B, Nagy B, Peto T, Hajdu A (2012) Combining algorithms for automatic detection of optic disc and macula in fundus images. *Comput Vision Image Understanding* 116(1):138–145. <https://doi.org/10.1016/j.cviu.2011.09.001>
 31. Rao A, Srihari RK and Zhang Z (1999) Spatial color histograms for content-based image retrieval. *Proc. of 11th IEEE International Conference on Tools with Artificial Intelligence*. 183–186. <https://doi.org/10.1109/TAI.1999.809784>
 32. REFUGE1 dataset: <https://refuge.grand-challenge.org/>. Accessed 08/07/2023
 33. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. 234–241, 2015. arXiv:1505.04597
 34. Sedai S, Tennakoon R, Roy P, Cao K, Garnavi R (2017) Multistage segmentation of the fovea in retinal fundus images using fully convolutional neural networks. *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*. 1083–1086. <https://doi.org/10.1109/ISBI.2017.7950704>.
 35. Sigut J, Nunez O, Fumero F, Gonzalez M, Arnay R (2017) Contrast based circular approximation for accurate and robust optic disc segmentation in retinal images. *Peer J* 5:e3763. <https://doi.org/10.7717/peerj.3763>
 36. Sinthanayothin C, Boyce JF, Cook HL, Williamson T (1999) Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol* 83:902–910. <https://doi.org/10.1136/bjo.83.8.902>
 37. Sun J, Zhang X, Cui J, Zhou L (2006) Image Retrieval Based on Color Distribution Entropy. *Pattern Recogn Lett* 27(10):1122–1126. <https://doi.org/10.1016/j.patrec.2005.12.014>
 38. Syed AM, Akram MU, Akram T, Muzammal M, Khalid S, Khan MA (2018) Fundus images-based detection and grading of macular edema using robust macula localization. *IEEE Access* 6:58784–58793. <https://doi.org/10.1109/ACCESS.2018.2873415>
 39. Welfer D, Scharcanski J, Marinho DR (2011) Fovea center detection based on the retina anatomy and mathematical morphology. *Comput Methods Programs Biomed* 104(3):397–409. <https://doi.org/10.1016/j.cmpb.2010.07.006>
 40. Xie R, Liu J, Cao R, Qiu CS, Duan J, Garibaldi J, Qiu G (2021) End-to-End Fovea Localisation in Colour Fundus Images With a Hierarchical Deep Regression Network. *IEEE Trans Med Imaging* 40(1):116–128. <https://doi.org/10.1109/TMI.2020.3023254>
 41. Zhao H, Shi J, Qi X, Wang X, Jia J (2016) Pyramid scene parsing network. *Computer Vision and Pattern Recognition*. arXiv:1612.01105
 42. Zheng S, Zhu Y, Pan L, Zhou T (2019) New Simplified Fovea and Optic Disc Localization Method for Retinal Images. *J Med Imaging Health Inform* 9(4):847–855. <https://doi.org/10.1166/jmihi.2019.2665>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.