Highlights

- A real-time algorithm for assessing retinal image quality is proposed.
- The neural network used is thousands of times smaller than state-of-the-art.
- The algorithm is able to accurately locate the fovea at no extra cost.
- Performance on a clinical database is similar to human one.
- The method can be used to save both patient and physician time.

Fast Macula Detection And Application to Retinal Image Quality Assessment

Robin Alais[a,*], Petr Dokládal[a], Ali Erginay[b], Bruno Figliuzzi[a], Etienne Decencière[a]

[a]*MINES ParisTech - PSL Research University, CMM, Center for Mathematical Morphology, 35 rue Saint Honoré, Fontainebleau, France*
[b]*AP-HP, Hôpital Lariboisière, Service d'Ophtalmologie, 2 rue Ambroise Paré, 75475 Paris*

Abstract

In this article, we present a segmentation algorithm for assessing retinal image quality with respect to the visibility of the macular region. An image is considered of acceptable quality if the macular region is clearly visible and entirely in the field of view. Additionally, for acceptable images, the method is able to locate the fovea with a maximal error of 0.34 mm. The algorithm is based on a lightweight fully-convolutional network, several thousand times smaller than state-of-the-art networks investigated so far in preliminary studies. We obtain near-human performance for assessing macula visibility and fovea localization. The presented method can easily be embedded in tabletop or handheld retinographs, decreasing the number of ungradable images, saving both patient and physician time. It is an important step towards automatic screening of retinal pathologies, including diabetic retinopathy, which is a major global healthcare issue.

Keywords: Image quality assessment, Convolutional neural networks, Retinal imaging, Macula Detection

1. Introduction

1.1. Problem Presentation

It is estimated that there were 415 million people with diabetes in 2015, and this number is expected to reach 642 million by 2040 [1]. A common complication of diabetes is diabetic retinopathy (DR), which is one of the main causes of blindness and visual loss [2, 3]. Due to the heterogeneity of protocols between different epidemiology studies, it is hard to give a precise prevalence of DR; however, the percentage of diabetic patients found with DR in recent studies is relatively stable and ranges from 21.9% to 36.8% [4].

DR is detectable and treatable, but regular clinical examination of all patients diagnosed with diabetes is infeasible in practice; in many developing countries, there is a significant lack of ophthalmologists, and in developed countries, the number of people aged 60+ is growing at twice the rate of the profession [5].

In most clinically significant cases, DR is detectable on eye fundus photographs. Telemedicine networks [6, 7, 8, 9] have been created in various countries in order to perform mass screening; international and local guidelines recommend one funduscopic examination per year for diabetic patients [10, 8]. Photographs can be taken by technicians in hospitals, specific screening centers, pharmacies or even prisons equipped with non-mydriatic cameras. In recent years, portable retinographs have been developed, which allow for even more massive screening. Photographs are then sent to ophthalmologists, who grade them and indicate the course of action to be followed. Both patient and practician time can be saved this way, provided that images are of good enough quality.

Due to the increasing amount of data, in conjunction with the limited number of ophthalmologists, computer retinal image understanding is of utmost interest. The literature concerning eye fundus image processing is abundant (a non-exhaustive review can be found in [11]), including many segmentation methods to extract anatomical structures such as the optic disk, the macula or the vascular network, or pathological structures [12, 13, 14, 15, 16, 17], as well as automatic predictions of DR severity [18, 19, 20].

Image quality estimation is a necessary preliminary step to these tasks, since it would make little sense applying an automatic diagnosis algorithm to images too noisy, blurred or not contrasted enough. Most publicly available datasets, like the Kaggle Diabetic Retinopathy dataset or the Messidor database [21], contain only gradable images, and in [13], it is clearly mentioned that for building a local database, "acceptable image quality, as judged by the screening program ophthalmologists, was a selection criterion".

In the context of a telemedicine network, the diagnosis is performed by a human expert, but the photographs are taken at a different time and location, by operators whose skill and level of experience can vary. A significant portion of images - around 10% for the OPHDIAT network [8] - are deemed uninterpretable by ophthalmologists. This could be prevented by automatically estimating the quality at

*Corresponding author
Email addresses: robin.alais@mines-paristech.fr (Robin Alais), petr.dokladal@mines-paristech.fr (Petr Dokládal), ali.erginay@aphp.fr (Ali Erginay), bruno.figliuzzi@mines-paristech.fr (Bruno Figliuzzi), etienne.decenciere@mines-paristech.fr (Etienne Decencière)

acquisition time, sending a warning to the operator if the photograph should be re-taken.

For a diagnosis to be made, an essential requirement is that the region of the macula must be clearly visible. In the present work, we focus on this task and present a lightweight CNN architecture capable of evaluating if the macula is 1) clearly visible 2) entirely within the field of view. In addition, our algorithm is able to accurately locate the macula if both conditions are met. This position can be used as well to check if a central image is correctly centered on the macula, or as part of an automated diagnosis algorithm, where the distance between lesions and the macula is an important information.

1.2. Related Work

So far, proposed methods for locating the macula either require images to be of sufficient quality [13], or attempt at providing a location based on contextual information like the optic disk and vascular network, even if the macula itself is not visible in the image [14, 15]. The originality of our approach lies in the fact that we use real clinical data, including very low-quality images; we automatically assess image quality, and we deliberately aim at detecting the macular region *only* if its visibility is sufficient.

The earliest attempts at defining a score for eye fundus image quality relied on properties of intensity histograms [22, 23]. Image structure clustering, introduced in [24], applies a bank of filters in order to perform unsupervised segmentation into several clusters roughly corresponding to anatomical structures of the retina, such as optic disk, vessels or retinal background. In this article, authors summarize a retinal photograph as a 20-dimensional vector composed of the histogram of the image structure clusters (5 bins), along with the three histograms of each color plane (5 bins each); good quality images are then separated from bad quality images by means of a Support Vector Machine. A similar approach was used in [25], although for images centered on the optic nerve of only 22.5° field of view. The authors also used a Support Vector Machine as their final classifiers, but they use Haralick [26] and sharpness features.

Other features have been proposed, such as the density of visible blood vessels, either in the whole image [27, 28] or near the macula [29]. Measures of clarity [30] and blurring [31] have also been defined. Finally, some methods combine both general image features and retina-specific ones, making use of vessel density, histogram, textural, and local sharpness [32, 33, 34].

In recent years, convolutional neural networks have been proven very efficient on difficult computer vision tasks, notably winning the ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2012 competition [35]. CNNs have then been applied to various tasks, including segmentation of retinal images [36, 37], glaucoma and DR grading [38, 39, 40].

In this context, estimating retinal image quality with convolutional neural networks is an interesting research direction. In [41], a convolutional network is trained in order to discriminate gradable images from artificial ungradable images obtained by adding noise to the original images of the DRISHTI dataset [42], which contains 101 acquisitions centered on the optic disk, with a 30° field of view, all images being taken with the pupils dilated. In [43], the same authors have experimented with both a 'shallow' network (the total number of weights cannot be calculated, since the number of neurons in the two fully-connected layers is not given, but the weights in the convolutional layers alone exceed 1 million) trained from scratch, and AlexNet [35] fine-tuned on a dataset consisting in 908 ungradable and 944 gradable non-mydriatic images. On a larger dataset (9653 ungradable retinal images and 11347 gradable images), they also evaluated the possibility of using a hybrid method combining saliency maps and CNNs [44]. Finally, [45] compare the performance of fine-tuning four CNN architectures - AlexNet [35], GoogLeNet [46], VGG-16 [47] and ResNet-50 - on a 3000-image subset of the Kaggle database. These preliminary studies report that large networks are hard to train, and must deal with overfitting issues, due to the huge amount of parameters. In an attempt to overcome this problem, data augmentation is extensively used, to the point where so much distortion is introduced that most of the training data is not constituted of real or even realistic examples. An example of this is generating artificial new data by applying large rotations (up to 210° !) to images; the resulting images are unrealistic, and it makes the network more or less rotational invariant, which is not a desirable feature for analyzing eye fundus images.

Another drawback of very large networks is their integration in embedded systems. Several million weights can constitute a significant fraction of the available memory: 233MB for AlexNet, 528MB for VGGNet for weights and biases alone, and the prediction times on embedded CPUs can be on the order of a second [48].

In this work, we propose a lightweight solution, with only 8329 parameters, and a reduced number of convolutions to be performed, meaning low power consumption as well. A comparison of memory requirements and computation times on embedded systems between our algorithm and other classic convolutional networks is given in Table 1. Our algorithm, including disk access and post-processing, was benchmarked on a Raspberry Pi; timings for the other networks come from [48] and were obtained on a more powerful 1.9GHz quad-core ARM Cortex-A57 64bit CPU (NVIDIA TX1). For a given task, we should expect it to be performed faster on the TX1 than on the Raspberry Pi. Despite this, our algorithm is the fastest, being more than three times as fast as ResNet, and more than 17 times as fast as VGGNet. It is also the lightest by far, requiring only 98kB for storing the network's weights. We have also benchmarked a modified version of the segmentation network U-Net [49], whose performance for our task is evaluated in section 4.

Network	Time (ms) TX1 (CPU)	Time (ms) Rasp.Pi	Weights (MB)
AlexNet	893		233
VGGNet	2809		528
GoogLeNet	638		26
ResNet	567		97
Us		161	0.1
U-Net		92	0.6

Table 1: Computation times and memory use of various convolutional networks on embedded systems. The first four networks have been benchmarked in [48] on a NVIDIA Jetson TX1. Our network and our implementation of U-Net were benchmarked on a Raspberry Pi; it should be expected that they would have run faster on the TX1.

2. Database

We extracted 6098 eye fundus images from the e-ophtha database [50]. This database has itself been extracted from the OPHDIAT telemedicine network for DR screening. These images are either *central* (centered on the macula), or *nasal* (centered on the optic disk). Different retinographs were used for the acquisitions, with resolutions ranging from 1440x960 to 3504x2336 pixels. Two different readers, independently from each other, indicated whether or not the macula was both entirely within the field of view, and clearly visible, meaning that the fovea and the small vessels around the avascular region can be seen. When that was the case, the fovea's position (x- and y- coordinates) was labeled. When there was a disagreement on the macula's visibility, a decision was made on which annotation to use. Out of the 6098 images that were considered, the macula was deemed visible by both readers on 3142. The remaining images correspond either to bad quality central images or to nasal images where the macula is at least partly outside the field of view (FOV). Sample images are shown in Fig. 1.

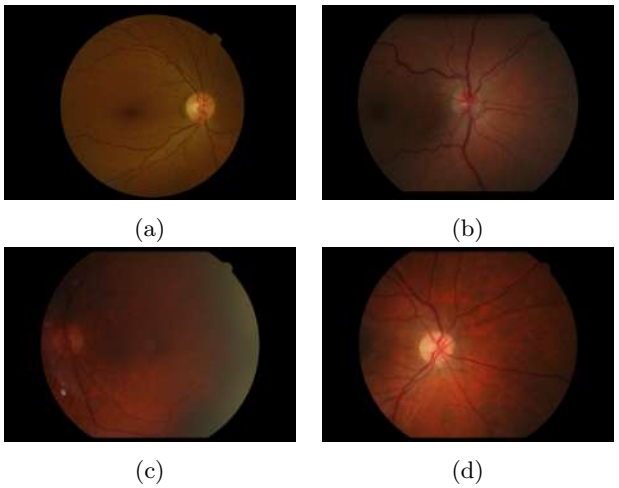


Figure 1: Sample images from the e-ophtha dataset: the macula is considered clearly visible and entirely within the FOV in both (a) and (b). It is in the center of image (c) but the quality is insufficient for grading, and it is partly outside the FOV in (d), which is a good quality nasal image.

3. Methodology

3.1. Image Preprocessing

Briefly put, the idea is to train a network to segment the macular region: with an ideal algorithm, if the segmentation is unsuccessful, it means that no macula is visible; if the segmentation succeeds, it means that image quality in the macular region was sufficient, and we immediately get fovea localization as a byproduct.

To train the network, we used the green channel, which was cropped, zero-padded in order to retain a square frame when necessary, and resized to a 128x128 image (see Fig. 2). This can seem aggressive, and some details might be lost, but approaches for assessing the severity of DR use networks with inputs as small as 512x512 [18], even though the task is much more complex and can rely on the presence of small structures like microaneurysms. A previous work for localizing both macula and optic disk in good-quality images used as input the green channel resized to 256x256 [37]. The aim of this work is a bit different, since we are less interested in predicting a precise location - although our algorithm never predicts a location outside the macula (see Sec. 4.2) - but rather in estimating the quality of the macular region. A 128x128 resolution is sufficient to visually assess the visibility of the macula, and means that smaller networks can be used, easier to train and less costly to run in an embedded framework, in real time.

Where the macula was visible, we used the mean of the two annotations as reference for the fovea's position; in 128x128 resolution, the average distance between readers was 1.25 pixels. We considered the macula to be about 20 pixels wide, and we used as ground truth a disk of ones of radius 10 pixels, the rest of the image being set to zero. When the macula was not visible because of low image quality or because it was at least in part outside the FOV (Fig. 1c and 1d), the ground truth was an image of zeros.

The only pre-processing consisted in dividing the (8-bit) images by 255, in order to get images valued between 0 and 1. No contrast enhancement or filtering was applied, since we wanted to evaluate the quality and macula visibility of raw images. Data augmentation was used by applying vertical symmetries (transforming a right eye into a left eye or the other way around), but no horizontal symmetries or rotations were used, to avoid creating unrealistic training data.

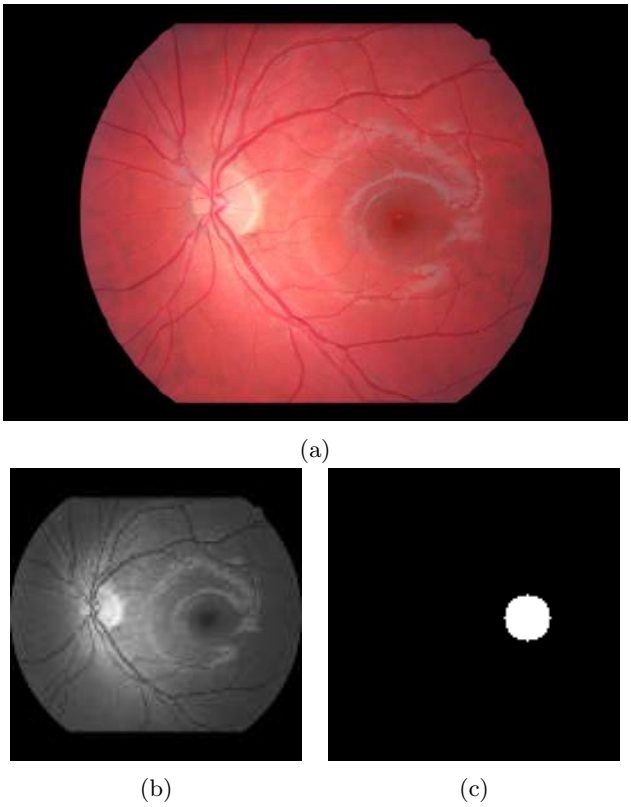
The dataset was split in three parts: a training set, used for learning network parameters, a validation set, used to estimate network performance during learning and setting some hyperparameters, and a test set, exclusively used to assess the method's performance. Images corresponding to the same patient are necessarily in the same set in order to avoid any evaluation bias. The number of images in each set is given in Table 2.

Additionally, in order to further estimate the general-

	Training	Validation	Test	All
Visible	2193	639	310	3142
Not visible	2056	596	304	2956
Total	4249	1235	624	6098

Table 2: Number of images in the training, validation and test sets.

izability of localization performance, we will use the ARIA Database C 8, for which fovea localization ground truth is available, as an additional test set.



(a)

(b) (c)

Figure 2: The green channel of the original color image (a) has been cropped around the region of interest, zero-padded on top and bottom so as not to introduce distortion, then resized to 128x128 (b). Image (c) is the ground truth used to train our convolutional networks.

3.2. Network architecture

We trained a fully-convolutional network consisting only of 3x3 convolutional layers (each followed by a rectified linear unit) piled up, with all convolutional layers having 8 channels. We used zero-padding at every step, so that the output image is the same size as the input one. A similar architecture has been shown to provide good results for cell nucleus segmentation [51].

The number of convolutional layers we pile up determines the receptive field of each pixel in the output image: that is, if there are L layers in the network (including the output layer), the value of one pixel in the output image depends on the values in a $2L+1$ by $2L+1$ square centered at this position in the input image. We tried different values of L, ranging from 10 to 20; the best validation loss was achieved for networks with $L = 16$. The receptive field is then 33x33 pixels. When centered on a pixel at the edge of the macula, it contains the fovea and most - but not all - of the ground truth macular region.

This network has very few parameters: each channel of the first layer is defined by a 3x3 convolutional kernel and a bias term. In the following layers, each channel is defined by a 3x3x8 convolutional kernel and an extra bias term (73 parameters). Finally, the output layer is defined by a 3x3x8 convolutional kernel and a bias term. The network has a total 8329 parameters, which is extremely few (in comparison, AlexNet has 60 million parameters).

The network was initialized with truncated normal distributions with standard deviation $\sigma = 0.1$ for the convolution weights and zeros for the biases. The objective function we minimized was the L^2 distance between images. The gradient was estimated at each step on a mini-batch of 8 images, using the RMSProp optimizer [52]. The network was trained for 4000 epochs.

3.3. Network Output Post-processing

Given an input 128x128 gray-scale image, the network outputs a 128x128 nonnegative array. A rectified linear unit is used in the output layer. We did also experiment using a sigmoid output activation, along with logistic loss, but it turned out that network convergence was harder to reach, and the obtained performance was lower.

Based on an output image, we have to answer the question "Is the macula visible in the original image ?". Since the ground truth image for a positive instance is a binary disk of radius 10, this is equivalent to answering the (ill-defined) question "Does the output image look like a binary disk of radius 10 ?".

The strategy we implemented goes as follows: the output image is thresholded above a value t in order to obtain a binary image. Then, only connected components of area greater than a value A are kept. If there is exactly one component remaining, we assume that the macula is visible, and the fovea can be located as the centroid of this component. If there is zero component of large enough area, we assume that the macula is not visible. If there are two or more components with area greater than A, this means that the algorithm is behaving oddly, and since our application is sensitivity-driven (it is important to correctly identify ungradable images), as a measure of precaution, we also consider that the macula is not clearly visible. Examples of network outputs and illustration of our post-processing are given in figure 3.

4. Results

4.1. Macula Visibility Estimation

4.1.1. Parameter Influence

There are two parameters to be set: the threshold t and minimum area A. In order to pick the best values

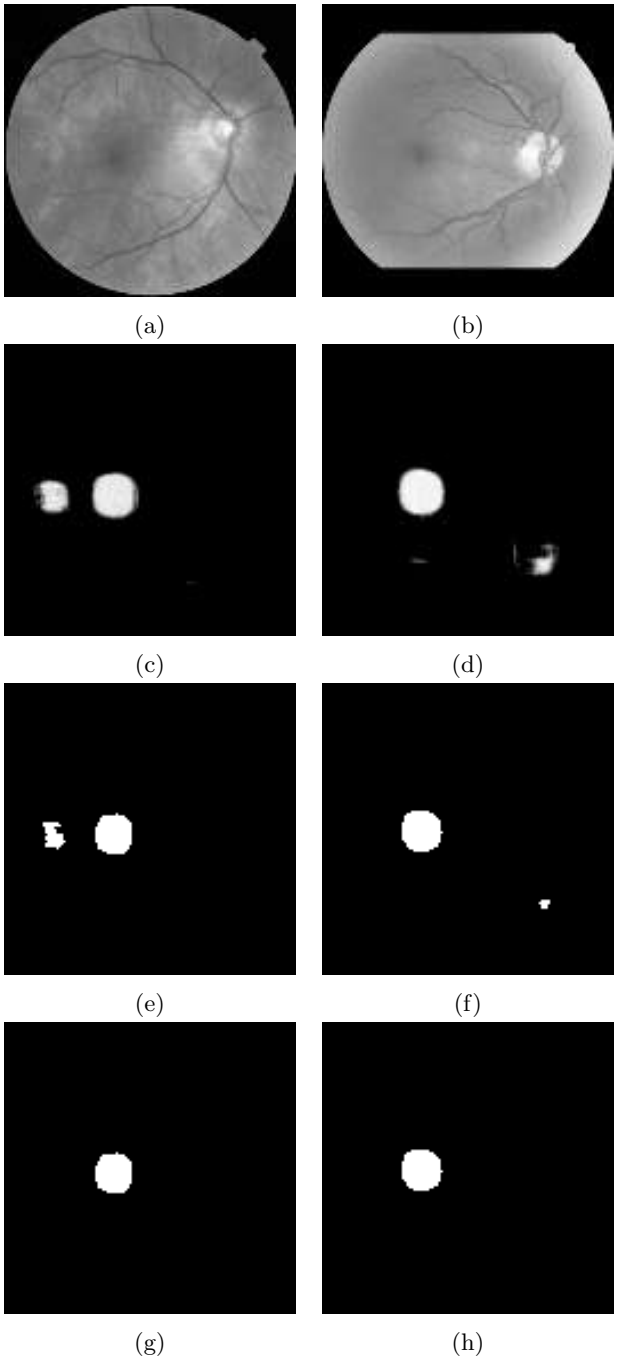


Figure 3: (a-b): input images, (c-d): corresponding network outputs, (e-f): thresholded network outputs ($t = 0.9$), (g-h): connected components of area greater than $A = 200$ pixels. These images belong to the test set and were chosen specifically because more than one component remained after thresholding, but this actually happens in only about 5% cases.

for these parameters, we looked at specificity (the fraction of images where the macula was annotated as visible correctly classified), sensitivity (the fraction of images where nothing was annotated classified as such) and overall accuracy on the validation set. Doing so on the training set could lead to overfitting, while doing so on the test set would give a biased estimation of the algorithm's ability to generalize.

The specificity, sensitivity and accuracy curves are shown in Fig. 4. Unsurprisingly, the higher the threshold, the higher the sensitivity, but even taking $t = 0.5$ and $A = 1$ (i.e. demanding there is only one connected component after thresholding), almost 97% images where macula visibility was considered insufficient by human readers are correctly classified. The main observation is that for $A \leq 200$, the choice of parameters actually has very little influence on the results. For the three considered thresholds, there is a steep fall when setting A above a certain value: in order to maximize sensitivity while keeping a certain margin to this critical value, we chose to set $t = 0.9$ and $A = 200$, which leads to a reasonable tradeoff, with 99% sensitivity, 95.3% specificity and 97.1% overall accuracy. We use those parameters in the following.

4.1.2. Test Set Results

Accuracy on the test set reaches 96.4%. In comparison, the agreement ratio between the two annotators before a consensus was made was only of 89.9%. Sensitivity reaches 98.7%; in other words, out of 304 images where the macula was not annotated in the ground truth, the algorithm makes 4 mistakes. The images on which these errors are made can be seen in Fig. 5. As can be seen on the figure, these correspond, if not to annotating errors, at least to borderline cases. In two out of the four images, the small vessels around the macula can even be distinguished.

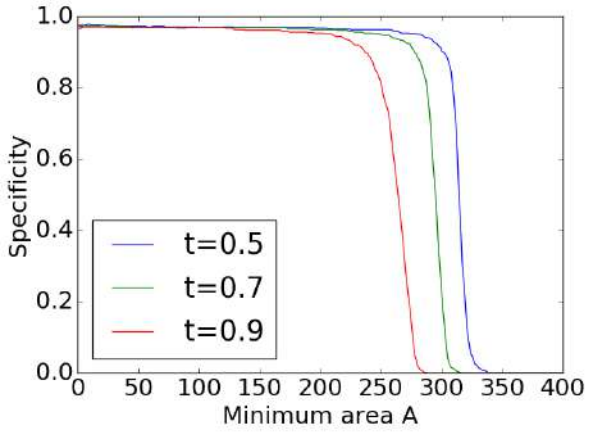
Specificity reaches 94.2%: out of 310 images where the macula was annotated as clearly visible, the algorithm correctly classifies 292.

4.1.3. Performance on Pathological Images

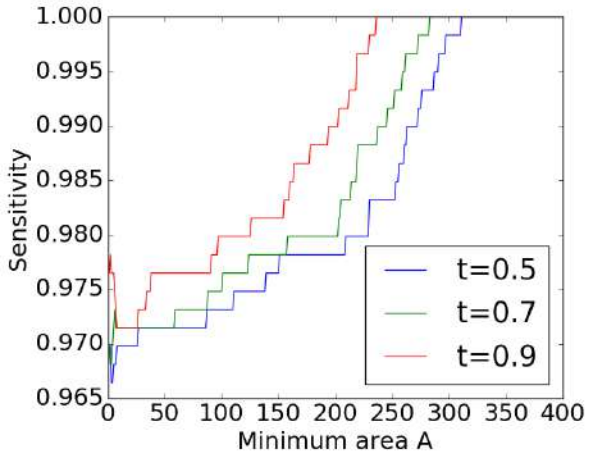
Although we obtain a very good performance on our test set in terms of accuracy, sensitivity and specificity, it is interesting to look more specifically at images where automatic macula detection would be expected to be hard. Figure 6 shows examples of network outputs for images where lesions are present. Although there is one case of unsuccessful detection, we can see that the network is able to detect the macula even when there are hemorrhages or exudates in the macular zone.

4.1.4. U-Net Comparison

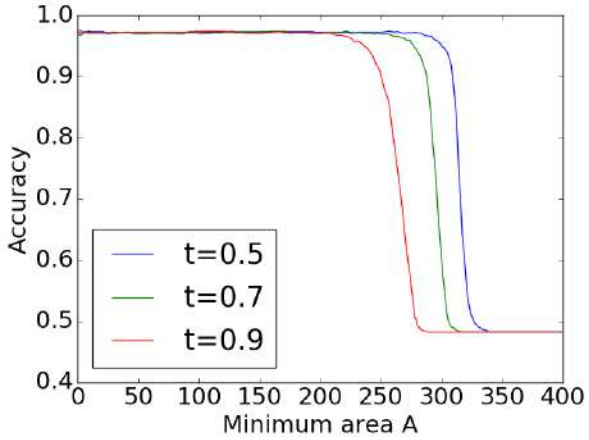
U-Net [49] is a popular network for segmentation and has been successfully used in a variety of applications. It makes perfect sense trying to apply it to our problem, however we have found it to slightly underperform compared to our network, in terms of both specificity and sensitivity. Several configurations of U-Net were tested. The best



(a) Algorithm specificity on the validation dataset plotted against A for different threshold values.



(b) Algorithm sensitivity on the validation dataset plotted against A for different threshold values.



(c) Algorithm accuracy on the validation dataset plotted against A for different threshold values.

Figure 4: Influence of the threshold t and the minimum area A on the classification performance.

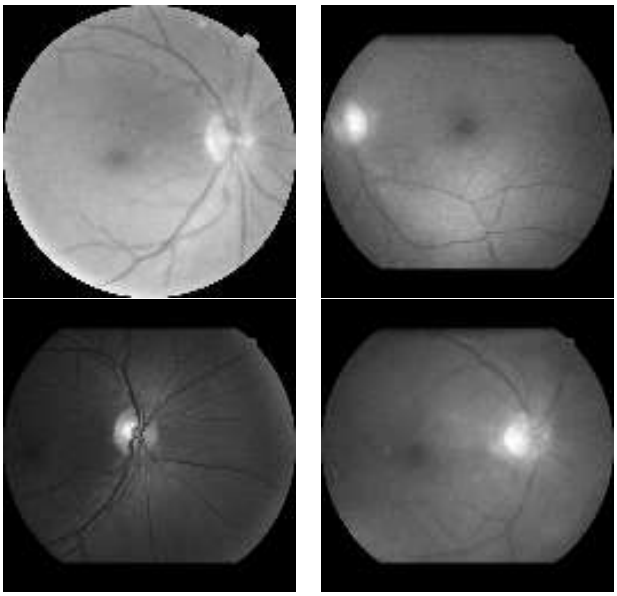


Figure 5: The four test set images for which the algorithm incorrectly predicts that the zone of the macula is of good quality.

one had 4 filters in the initial convolutional layer, and a gaussian noise layer at the end. On the test set, this network incorrectly predicts that the macula is not visible on 19 images, which is similar to our network (18 false negatives). The main drawback is that it has more than twice as many false positives (9 versus 4). It also has many more parameters (122,953) than our network. As for fovea localization, which is detailed in the next section, the average error for U-Net is 1.22 pixels, which is again more than our network's error (0.95 pixel).

4.2. Fovea Localization Results

Although the main task our algorithm addresses is assessing the quality of the macula region, it can also be used to segment the macula, or localize the fovea, as mentioned in section 3.3. In this section, we evaluate the performance of our algorithm for localizing the fovea, on the database we extracted from the e-ophtha database and on the ARIA database.

4.2.1. e-ophtha Database

As mentioned in the previous section, if we use $t = 0.9$ and $A = 200$, our algorithm predicts 292 images where the macula is visible, out of the 310 images where it was annotated. By lowering either parameter or defining another strategy, we could predict a location for the macula for more images, but it would make little sense in this context trying to localize it if we are not even confident it is in the FOV. In order to remain consistent, we leave the parameters unchanged and present localization results only on the images identified by our algorithm.

As previously mentioned in section 3.3, we use the centroid of the (only) connected component of the thresholded output image as our estimation for fovea location. The

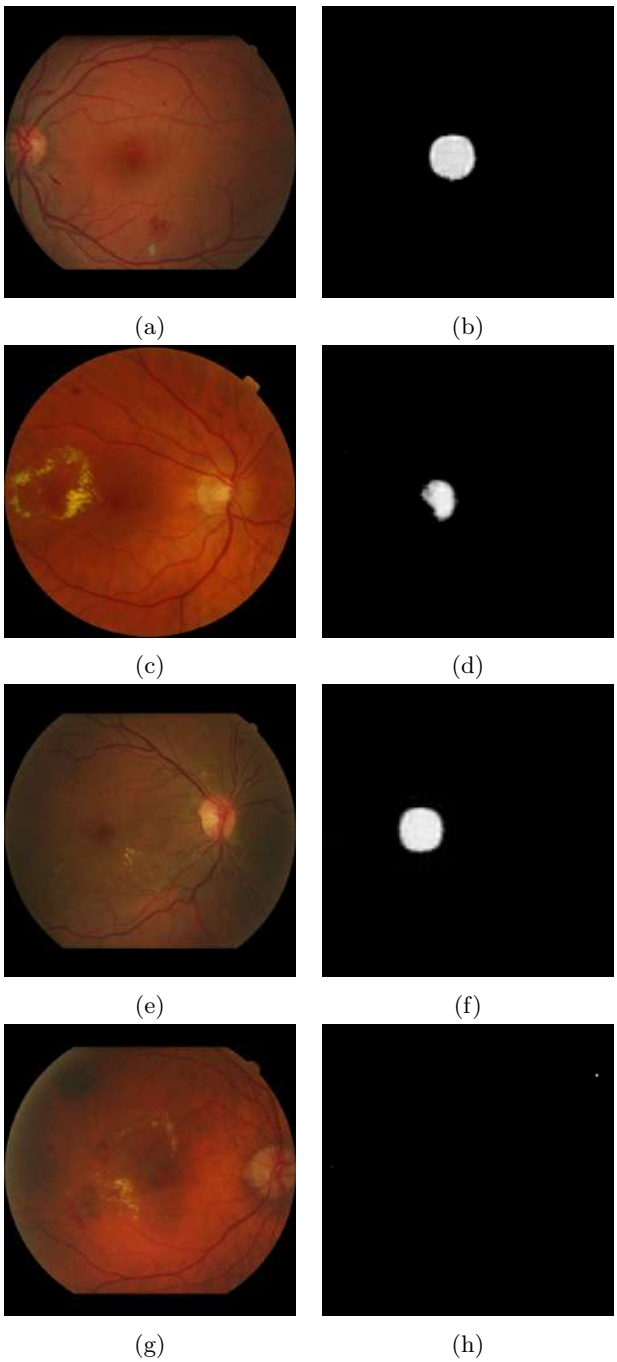


Figure 6: Network outputs for pathological images. In the four cases, there are hemorrhages and/or exudates in the macular region. Despite this, the network provides a reasonable output in the first three cases (the areas after thresholding at $t = 0.9$ are respectively 274, 91 and 250). The last case is an example of unsuccessful detection.

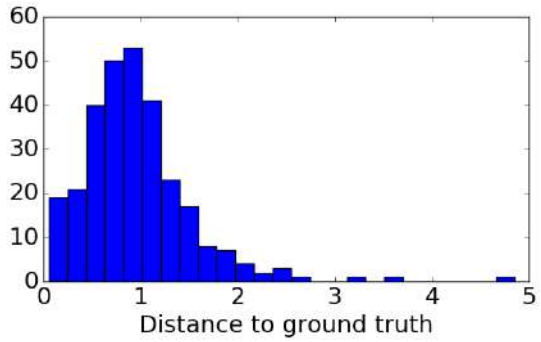


Figure 7: Histogram of distances to ground truth on the test base (in pixels after resizing to 128x128).

mean of the two annotators' positions is used as ground truth. The histogram of distances to the ground truth is shown in Fig. 7.

The average test error is 0.95 pixel, which is less than the average distance between the annotators. One pixel in 128x128 resolution represents 0.075 mm. The largest test error is 4.85 pixels, or 0.33 mm; it has to be noted that, since the macula was considered to be 10 pixels in radius, all of the predicted values lie within the macula.

4.2.2. ARIA Database C

The ARIA database C contains 61 central eye fundus images with corresponding ground truth annotations for optic disk and fovea. Our network, with the same post-processing parameters as before, predicts a position for 46 of them, with a mean error of 1.4 pixel (0.1 mm) and a maximum error of 6 pixels (0.44 mm). The average error is again comparable to the average distance between two human readers (1.25 pixels, as mentioned in Sec. 2) and the maximum error is about half the macula's radius.

The 15 remaining images correspond to poor quality acquisitions; some examples can be seen in Fig. 8. Although it is quite easy for an experimented human observer to approximately locate the fovea region, the macular regions of these images are clearly of limited to no interest for an ophthalmologist.

5. Discussion

Guidelines for teleophthalmology recommend taking two photographs per eye, one centered on the macula, the other centered on the optic disk. A mandatory condition for a couple of images to be gradable by an ophthalmologist is that the macula is clearly visible in at least one of the two images. In practice, a significant proportion of examinations - about 10% for the OPHDIAT network - does not meet this required quality criterion. The algorithm detailed in the present work could significantly reduce the fraction of ungradable central images. Since the algorithm also provides the location of the macula, it can also be used in order to assert that a central image is indeed well

Figure 8: Examples of images from the ARIA database for which the network does not return a fovea position.

centered around it. It can also be used as part of an automated diagnosis algorithm, for which macula segmentation is often a crucial step [53].

Eye fundus photographs so far are often made by healthcare professionals, using tabletop non-mydriatic cameras; however, recent years have seen the emergence of portable, handheld retinographs, cheaper and allowing for screening in remote locations. These generally produce images of lower quality than tabletop retinographs. The algorithm described in the present work is very lightweight: it requires very little memory for storage, is very fast, and consumes little energy, compared to the much larger state-of-the-art networks. It can easily be integrated in an embedded device, telling in real time the operator whether another acquisition should be made.

6. Conclusion

In the present work, we use a fully-convolutional network to segment the macular region, in order to assess the quality of eye fundus images. Our algorithm is also able to provide fovea localization, within 0.1 mm of human performance in average, in the case image quality is deemed sufficient.

Although macula visibility is not the only requirement for a retinal image to be gradable for DR, the method we propose is able to significantly reduce the number of ungradable images sent to medical experts in teleophthalmology networks, saving both patient and physician time, when the number of ophthalmologists is insufficient for current needs and the diabetic population is expected to grow much faster than that of the profession. It can also be combined with other quality criteria, such as contrast and sharpness, to build a complete quality assessment system.

Acknowledgements

This work was funded by the French "Fonds Unique Interministériel" through the RetinOpTIC project and supported by the Medicen and Systematic competitive clusters.

References

[1] K. Ogurtsova, J. da Rocha Fernandes, Y. Huang, U. Linnenkamp, L. Guariguata, N. Cho, D. Cavan, J. Shaw, L. Makaroff, IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040, Diabetes Research and Clinical Practice 128 (2017) 40–50. doi:10.1016/j.diabres.2017.03.024.

[2] A. K. Sjølie, J. Stephenson, S. Aldington, E. Kohner, H. Janka, L. Stevens, J. Fuller, B. Karamanos, C. Tountas, A. Kofinis, Retinopathy and vision loss in insulin-dependent diabetes in Europe: the EURODIAB IDDM Complications Study, Ophthalmology 104 (2) (1997) 252–260.

[3] D. R. Matthews, I. M. Stratton, S. J. Aldington, R. R. Holman, E. M. Kohner, UK Prospective Diabetes Study Group, Risks of progression of retinopathy and vision loss related to tight blood pressure control in type 2 diabetes mellitus: Ukpds 69, Archives of ophthalmology (Chicago, Ill. : 1960) 122 (11) (2004) 1631—1640. doi:10.1001/archopht.122.11.1631.

[4] C. Delcourt, P. Massin, M. Rosilio, Epidemiology of diabetic retinopathy: expected vs reported prevalence of cases in the french population, Diabetes & metabolism 35 (6) (2009) 431–438.

[5] S. Resnikoff, W. Felch, T.-M. Gauthier, B. Spivey, The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners, British Journal of Ophthalmology 96 (6) (2012) 783–787. doi:10.1136/bjophthalmol-2011-301378.

[6] J. Choremis, D. R. Chow, Use of telemedicine in screening for diabetic retinopathy, Canadian Journal of Ophthalmology / Journal Canadien d'Ophtalmologie 38 (7) (2003) 575 – 579. doi:https://doi.org/10.1016/S0008-4182(03)80111-4.

[7] M. C. Boucher, G. Desroches, R. Garcia-Salinas, A. Kherani, D. Maberley, S. Olivier, M. Oh, F. Stockl, Teleophthalmology screening for diabetic retinopathy through mobile imaging units within Canada, Canadian Journal of Ophthalmology / Journal Canadien d'Ophtalmologie 43 (6) (2008) 658–668. doi:10.3129/i08-120.

[8] P. Massin, A. Chabouis, A. Erginay, C. Viens-Bitker, A. Lecleire-Collet, T. Meas, P.-J. Guillausseau, G. Choupot, B. André, P. Denormandie, OPHDIAT©: A telemedical network screening system for diabetic retinopathy in the Île-de-France, Diabetes & Metabolism 34 (3) (2008) 227 – 234. doi:https://doi.org/10.1016/j.diabet.2007.12.006.

[9] K. Tozer, M. A. Woodward, P. A. Newman-Casey, Telemedicine and Diabetic Retinopathy: Review of Published Screening Programs., Journal of endocrinology and diabetes 2 (4).

[10] American Diabetes Association, Executive Summary: Standards of Medical Care in Diabetes–2012, Diabetes Care 35 (Supplement_1) (2012) S4–S10. doi:10.2337/dc12-s004.

[11] K. S. Sreejini, V. K. Govindan, A Review of Computer Aided Detection of Anatomical Structures and Lesions of DR from Color Retina Images, International Journal of Image, Graphics and Signal Processing 7 (11) (2015) 55–69. doi:10.5815/ijigsp.2015.11.08.

[12] C. Agurto, V. Murray, H. Yu, J. Wigdahl, M. Pattichis, S. Nemeth, E. S. Barriga, P. Soliz, A Multiscale Optimization Approach to Detect Exudates in the Macula, IEEE Journal of Biomedical and Health Informatics 18 (4) (2014) 1328–1336. doi:10.1109/JBHI.2013.2296399.

[13] M. Niemeijer, M. D. Abràmoff, B. van Ginneken, Fast detection of the optic disc and fovea in color fundus photographs, Medical Image Analysis 13 (6) (2009) 859–870. doi:10.1016/j.media.2009.08.003.

[14] R. Veras, F. Medeiros, R. Silva, D. Ushizima, Assessing the accuracy of macula detection methods in retinal images, in: Digital Signal Processing (DSP), 2013 18th International Conference on, IEEE, 2013, pp. 1–6.

[15] D. Welfer, J. Scharcanski, D. R. Marinho, Fovea center detection based on the retina anatomy and mathematical morphology, Computer Methods and Programs in Biomedicine 104 (3) (2011) 397–409. doi:10.1016/j.cmpb.2010.07.006.

[16] G. Gupta, K. Ram, S. Kulasekaran, N. Joshi, M. Sivaprakasam, R. Gandhi, Detection of retinal hemorrhages in the presence of blood vessels, in: Proceedings of the Ophthalmic Medical Image Analysis First International Workshop, Boston, Massachusetts, 2014, pp. 105–112.

[17] X. Zhang, G. Thibault, E. Decencière, B. Marcotegui, B. Laÿ, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin, A. Chabouis, Z. Victor, A. Erginay, Exudate detection in color retinal images for mass screening of diabetic retinopathy, Medical Image Analysis 18 (7) (2014) 1026–1043. doi:10.1016/j.media.2014.05.004.

[18] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, M. Lamard, Deep image mining for diabetic retinopathy screening, Medical Image Analysis 39 (2017) 178–193. doi:10.1016/j.media.2017.04.012.
URL https://linkinghub.elsevier.com/retrieve/pii/S136184151730066X

[19] Z. Xiao, X. Zhang, L. Geng, F. Zhang, J. Wu, J. Tong, P. O. Ogunbona, C. Shan, Automatic non-proliferative diabetic retinopathy screening system based on color fundus image, BioMedical Engineering OnLine 16 (1). doi:10.1186/s12938-017-0414-z.
URL http://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-017-0414-z

[20] G. G. Gardner, D. Keating, T. H. Williamson, A. T. Elliott, Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool., British Journal of Ophthalmology 80 (11) (1996) 940–944. doi:10.1136/bjo.80.11.940.
URL http://bjo.bmj.com/cgi/doi/10.1136/bjo.80.11.940

[21] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, J.-C. Klein, FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE, Image Analysis & Stereology 33 (3) (2014) 231. doi:10.5566/ias.1155.
URL https://www.ias-iss.org/ojs/IAS/article/view/1155

[22] S. Lee, Y. Wang, Automatic retinal image quality assessment and enhancement, in: Proceedings of SPIE Image Processing, Vol. 3661, 1999, pp. 1581–1590.

[23] M. Lalonde, L. Gagnon, M.-C. Boucher, Automatic visual quality assessment in optical fundus images, in: Proceedings of vision interface, Vol. 32, Ottawa, 2001, pp. 259–264.

[24] M. Niemeijer, M. Abramoff, B. Vanginneken, Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening, Medical Image Analysis 10 (6) (2006) 888–898. doi:10.1016/j.media.2006.09.006.

[25] J. Paulus, J. Meier, R. Bock, J. Hornegger, G. Michelson, Automated quality assessment of retinal fundus photos, International Journal of Computer Assisted Radiology and Surgery 5 (6) (2010) 557–564. doi:10.1007/s11548-010-0479-7.

[26] R. M. Haralick, K. Shanmugam, et al., Textural features for image classification, IEEE Transactions on systems, man, and cybernetics (6) (1973) 610–621.

[27] L. Giancardo, M. D. Abràmoff, E. Chaum, T. P. Karnowski, F. Meriaudeau, K. W. Tobin, Elliptical local vessel density: a fast and robust quality metric for retinal images, in: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE, 2008, pp. 3534–3537.

[28] L. Giancardo, F. Meriaudeau, T. P. Karnowski, E. Chaum, K. Tobin, Quality assessment of retinal fundus images using elliptical local vessel density, in: New developments in biomedical engineering, InTech, 2010.

[29] A. Hunter, J. A. Lowell, M. Habib, B. Ryder, A. Basu, D. Steel, An automated retinal image quality grading algorithm, in: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, IEEE, 2011, pp. 5955–5958.

[30] A. D. Fleming, S. Philip, K. A. Goatman, P. F. Sharp, J. A. Olson, Automated clarity assessment of retinal images using regionally based structural and statistical measures, Medical Engineering & Physics 34 (7) (2012) 849–859. doi:10.1016/j.medengphy.2011.09.027.

[31] R. Pires, H. F. Jelinek, J. Wainer, A. Rocha, Retinal Image Quality Analysis for Automatic Diabetic Retinopathy Detection, IEEE, 2012, pp. 229–236. doi:10.1109/SIBGRAPI.2012.39.

[32] H. Yu, C. Agurto, S. Barriga, S. C. Nemeth, P. Soliz, G. Zamora, Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening, in: Image analysis and interpretation (SSIAI), 2012 IEEE southwest symposium on, IEEE, 2012, pp. 125–128.

[33] J. M. Pires Dias, C. M. Oliveira, L. A. da Silva Cruz, Retinal image quality assessment using generic image quality indicators, Information Fusion 19 (2014) 73–90. doi:10.1016/j.inffus.2012.08.001.

[34] J. M. P. Dias, C. M. Oliveira, L. A. d. S. Cruz, Evaluation of Retinal Image Gradability by Image Features Classification, Procedia Technology 5 (2012) 865–875. doi:10.1016/j.protcy.2012.09.096.

[35] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[36] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, L. Van Gool, Deep retinal image understanding, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 140–148.

[37] B. Al-Bander, W. Al-Nuaimy, B. M. Williams, Y. Zheng, Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc, Biomedical Signal Processing and Control 40 (2018) 91–101. doi:10.1016/j.bspc.2017.09.008.

[38] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, Y. Zheng, Convolutional neural networks for diabetic retinopathy, Procedia Computer Science 90 (Supplement C) (2016) 200 – 205, 20th Conference on Medical Image Understanding and Analysis (MIUA 2016). doi:https://doi.org/10.1016/j.procs.2016.07.014.

[39] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, Jama 316 (22) (2016) 2402–2410.

[40] B. Graham, Competition Report, Tech. rep., Department of Statistics and Centre for Complexity Science, Coventry, UK (2015).

[41] D. Mahapatra, P. K. Roy, S. Sedai, R. Garnavi, A CNN based neurobiology inspired approach for retinal image quality assessment, in: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the, IEEE, 2016, pp. 1304–1307.

[42] J. Sivaswamy, S. Krishnadas, A. Chakravarty, G. Joshi, A. S. Tabish, et al., A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis, JSM Biomedical Imaging Data Papers 2 (1) (2015) 1004.

[43] R. Tennakoon, D. Mahapatra, P. Roy, S. Sedai, R. Garnavi, Image Quality Classification for DR Screening Using Convolutional Neural Networks, University of Iowa, 2016, pp. 113–120. doi:10.17077/omia.1054.

[44] D. Mahapatra, P. K. Roy, S. Sedai, R. Garnavi, Retinal Image Quality Classification Using Saliency Maps and CNNs, in: L. Wang, E. Adeli, Q. Wang, Y. Shi, H.-I. Suk (Eds.), Machine Learning in Medical Imaging: 7th International Workshop, MLMI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Proceedings, Springer International Publishing, Cham, 2016, pp. 172–179, dOI: 10.1007/978-3-319-47157-0_21.

[45] J. Sun, C. Wan, J. Cheng, F. Yu, J. Liu, Retinal Image Quality Classification Using Fine-Tuned CNN, in: M. J. Cardoso, T. Arbel, A. Melbourne, H. Bogunovic, P. Moeskops, X. Chen, E. Schwartz, M. Garvin, E. Robinson, E. Trucco, M. Ebner,

Y. Xu, A. Makropoulos, A. Desjardin, T. Vercauteren (Eds.), Fetal, Infant and Ophthalmic Medical Image Analysis, Vol. 10554, Springer International Publishing, Cham, 2017, pp. 126–133, dOI: 10.1007/978-3-319-67561-9_14.

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[47] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR abs/1409.1556.

[48] Z. Lu, S. Rallapalli, K. Chan, T. La Porta, Modeling the Resource Requirements of Convolutional Neural Networks on Mobile Devices, Proceedings of the 2017 ACM on Multimedia Conference - MM '17 (2017) 1663–1671ArXiv: 1709.09503. doi:10.1145/3123266.3123389.
URL http://arxiv.org/abs/1709.09503

[49] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Vol. 9351, Springer International Publishing, Cham, 2015, pp. 234–241, dOI: 10.1007/978-3-319-24574-4_28.

[50] E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, B. Laÿ, A. Chabouis, TeleOphta: Machine learning and image processing methods for teleophthalmology, IRBM 34 (2) (2013) 196–203. doi:10.1016/j.irbm.2013.01.010.

[51] B. Pang, Y. Zhang, Q. Chen, Z. Gao, Q. Peng, X. You, Cell nucleus segmentation in color histopathological imagery using convolutional networks, in: Chinese conference on pattern recognition, 2010, pp. 1–5.

[52] G. Hinton, N. Srivastava, K. Swersky, Rmsprop: Divide the gradient by a running average of its recent magnitude, Neural networks for machine learning, Coursera lecture 6e.

[53] X. Zhang, Image processing methods for computer-aided screening of diabetic retinopathy, Ph.D. thesis, École nationale supérieure des mines Paris (2014).