Fovea Detection in Optical Coherence Tomography using Convolutional Neural Networks

Bart Liefers, Freerk G. Venhuizen, Thomas Theelen, Carel Hoyng, Bram van Ginneken, and Clara I. Sánchez

Radboud University Medical Center, Nijmegen, the Netherlands.

ABSTRACT

The fovea is an important clinical landmark that is used as a reference for assessing various quantitative measures, such as central retinal thickness or drusen count. In this paper we propose a novel method for automatic detection of the foveal center in Optical Coherence Tomography (OCT) scans. Although the clinician will generally aim to center the OCT scan on the fovea, post-acquisition image processing will give a more accurate estimate of the true location of the foveal center. A Convolutional Neural Network (CNN) was trained on a set of 781 OCT scans that classifies each pixel in the OCT B-scan with a probability of belonging to the fovea. Dilated convolutions were used to obtain a large receptive field, while maintaining pixel-level accuracy. In order to train the network more effectively, negative patches were sampled selectively after each epoch. After CNN classification of the entire OCT volume, the predicted foveal center was chosen as the voxel with maximum output probability, after applying an optimized three-dimensional Gaussian blurring. We evaluate the performance of our method on a data set of 99 OCT scans presenting different stages of Age-related Macular Degeneration (AMD). The fovea was correctly detected in 96.9% of the cases, with a mean distance error of 73 µm(±112 µm). This result was comparable to the performance of a second human observer who obtained a mean distance error of 69 µm(±94 µm). Experiments showed that the proposed method is accurate and robust even in retinas heavily affected by pathology.

Keywords: Fovea, Optical Coherence Tomography, Convolutional Neural Networks

1. DESCRIPTION OF PURPOSE

The fovea is a small region (about 1.5 mm in diameter[1]) located roughly in the center of the retina, opposite the lens. This region has the highest concentration of cones, photoreceptor cells responsible for color vision and high spatial acuity. Consequently, abnormalities involving the fovea have a severe effect on visual acuity and central vision. Therefore, the foveal center is an important clinical landmark for the assessment and monitoring of visual impairment. Moreover, the foveal center is routinely used in clinical practice and clinical trials to measure structural biomarkers, such as central retinal thickness, for the evaluation of treatment outcome and decision making regarding retreatment.

The development of Optical Coherence Tomography (OCT) made the in vivo study of foveal morphology possible. An OCT scan is composed of a stack of slices (B-scans) covering the central region of the retina and provides a cross-sectional view of the retina. Compared to en-face modalities, OCT provides more accurate and complete information on the exact location of the fovea, especially in pathological retinas.

During the OCT acquisition, the operator or clinician will generally aim to acquire a scan centered at the fovea. However, it has been shown that the scan center does not always coincide neatly with the foveal center, introducing errors in the extracted biomarkers.[2] Manual detection of the foveal center after acquisition would correct this misalignment but it is a time consuming task. In contrast, automatic post acquisition image analysis may provide a fast and accurate estimate of the foveal center.[3]

In this paper, we propose a method based on a Convolutional Neural Network (CNN) for automated detection of the foveal center in OCT scans that is robust and performs well even in pathological retinas. The proposed method uses dilated convolutional filters.[4] It takes the full OCT scan as input and produces a single coordinate as output. This coordinate represents the center of the fovea.

Medical Imaging 2017: Image Processing, edited by Martin A. Styner, Elsa D. Angelini, Proc. of SPIE
Vol. 10133, 1013302 · © 2017 SPIE · CCC code: 1605-7422/17/$18 · doi: 10.1117/12.2254301

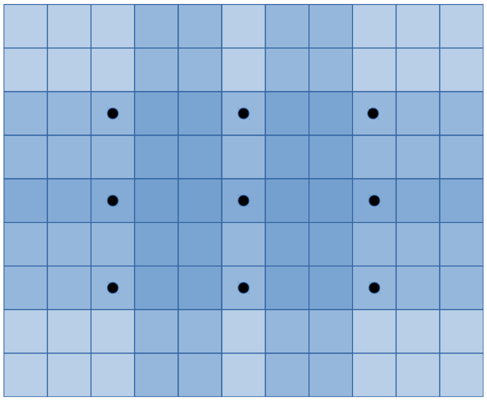

Figure 1: Visualization of the dilated convolution filter at layer 3. Only the pixels with black dots are included in this filter. These pixels have a receptive field of 5 × 5 pixels each, as a result of the previous two convolutional layers.

Layer	Convolution	Dilation	Channels	Receptive Field
1	3 × 3	(1, 1)	32	3 × 3
2	3 × 3	(1, 1)	32	5 × 5
3	3 × 3	(3, 2)	64	11 × 9
4	3 × 3	(6, 4)	64	23 × 17
5	3 × 3	(12, 8)	64	47 × 33
6	3 × 3	(24, 16)	128	95 × 65
7	3 × 3	(48, 32)	128	191 × 129
8	3 × 3	(1, 1)	128	193 × 131
9	1 × 1	(1, 1)	64	193 × 131
10	1 × 1	(1, 1)	2	193 × 131

Table 1: Summary of the network architecture. The dilation refers to the spacing in the dilated filters in the horizontal and vertical direction respectively. Channels refers to the number of channels in the particular layer. The receptive fields indicate the size of the field of all pixels that can influence the network output at that layer.

2. METHODS

2.1 Data

For this study a total of 880 OCT scans were selected from the European Genetic Database (EUGENDA), a large multi-center database for clinical and molecular analysis of Age-related Macular Degeneration (AMD).[5] OCT volumes were acquired using a Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany) at a wavelength of 870 nm, a horizontal resolution ranging from 6 µm to 14 µm and an axial resolution of up to 3.5 µm. The number of slices, i.e. the number of B-scans, in the scans used for this study varied from 18 to 48. Before processing, all B-scans from an OCT scan were resampled to a constant pixel size of 11.5 µm × 3.9 µm (lateral × axial). The distance between B-scans varied from 111.8 µm to 288.2 µm.

For each OCT scan, the AMD severity stage was provided, namely control, early AMD, intermediate AMD, advanced AMD with choroidal neovascularization (CNV) and advanced AMD with geographic atrophy (GA). The data was randomly divided into a training set, consisting of 781 OCT scans (141 graded as control, 49 as early AMD, 63 as intermediate AMD, 104 as advanced AMD CNV and 7 as advanced AMD GA); and a validation set, consisting of 99 OCT scans (20 controls, 20 early AMD, 20 intermediate AMD, 19 advanced AMD CNV, 20 advanced AMD GA). For each OCT scan in the data set, the foveal center was manually annotated by a human observer (reference annotations). One scan, graded as advanced AMD CNV, was excluded because the reference grader was unable to manually place the foveal center because of poor image quality. To assess human performance, a second observer independently annotated the foveal center in the OCT scans of the validation set.

2.2 Network architecture

Given an OCT scan as input, the proposed CNN architecture classifies each pixel in each B-scan with a probability of belonging to the fovea. This architecture is based on dilated convolutions.[4] Dilated convolutional filters operate on pixels at a specific distance from the central pixel instead of the direct neighboring pixels (see Figure 1). If applied in a context with layers with increasing dilation size, pixel-level accuracy can be maintained, while the dilated filters allow to include information from a larger context compared to traditional filters. The CNN architecture used in this study is fully convolutional[6] and includes five layers with dilated convolutions (see Table 1). After the final layer a soft-max classification is applied. In contrast to the architecture proposed in by Yu and Koltun,[4] the dilated filters used in this architecture present asymmetrical spacing in the axial and

E-mail: Bart.Liefers@Radboudumc.nl

lateral directions. The benefit of using this spacing for fovea detection is that a larger receptive field (the set of pixels to which the network is path-connected) is obtained in the lateral direction in order to include the confluence of retinal layers that is typically observed near the foveal center. The final receptive field of this network is 193 × 131 pixels (see Figure 3). After each convolutional layer, we apply a leaky rectify non-linearity with leakiness 0.01. After CNN classification, the obtained probability map for a full OCT scan is smoothed using a Gaussian filter with a sigma of 150 µm × 5 µm × 150 µm (lateral × axial × transverse). The maximum of the smoothed probability map represents the final foveal center.

2.3 Training

To construct the CNN training data, positive samples were drawn from a rectangular area of 60 µm × 20 µm (5 × 5 pixels) around the manually annotated foveal centers in the training set; whereas negative patches were centered on training pixels with a lateral and axial distance of at least 300 µm from the foveal center. Data augmentation by rotation between -10 and +10 degrees and horizontal flipping is applied to artificially increase the number of samples. In order to increase the efficiency of the CNN learning process, a selective sampling strategy to focus on challenging training samples is applied. At each CNN training epoch, a weight is assigned to each negative sample, proportional to its classification error at the current CNN state. This weight represents its sampling probability: higher weight means a higher probability to be selected for the next epoch.[7] The probabilities are calculated according to the following equation:

$$p_i = \frac{w_i}{\sum_{j \in X_-} w_j} \tag{1}$$

Here p_i is the probability of background pixel i to be selected, w_i is the assigned probability of belonging to the fovea for pixel i, and X_- is the set of background pixels with a distance of at least 300 µm to the foveal center.

In this study, for each epoch, 781 batches of 32 positive patches and 32 negative patches are supplied to the network. After every epoch, B-scans containing the foveal center in the training set are classified using the current CNN state and a weight is assigned to each negative patch within these B-scans. The negative samples for the next epoch are dynamically sampled based on the calculated weights. Although we apply the CNN to all B-scans from an OCT volume for classification, only the B-scans containing the foveal center are considered during the training phase in order to reduce the computational complexity. The network is trained for 10 epochs using RMSprop with a learning rate of 10^{-5}.

3. RESULTS

For evaluation, the fovea center was correctly identified if the detected position lies within the fovea region (a circular region of 1.5mm diameter[1]). The foveal center was correctly identified in 96 of 99 cases (96.9% accuracy) of the validation set. For those cases, Figure 2 shows the distance errors of the proposed method and the second observer compared to the reference annotations for each AMD severity level. In Table 2 a comparison of the accuracy of the method and the observer can be found. For the 96 correct detections, the results of the proposed approach are comparable to the human observer (no significant differences were observed, p = 0.74), even in heavily affected retinas, with a global mean distance error of 73 µm(±112 µm) for the automatic method and 69 µm(±94 µm) for the observer.

	Umbo (75 µm)	Foveola (175 µm)	FAZ (250 µm)	Fovea (750 µm)	Total
Method	73	84	91	96	99
Observer	71	86	90	99	99

Table 2: Number of detections within a certain distance of the reference foveal center for the method and the observer.

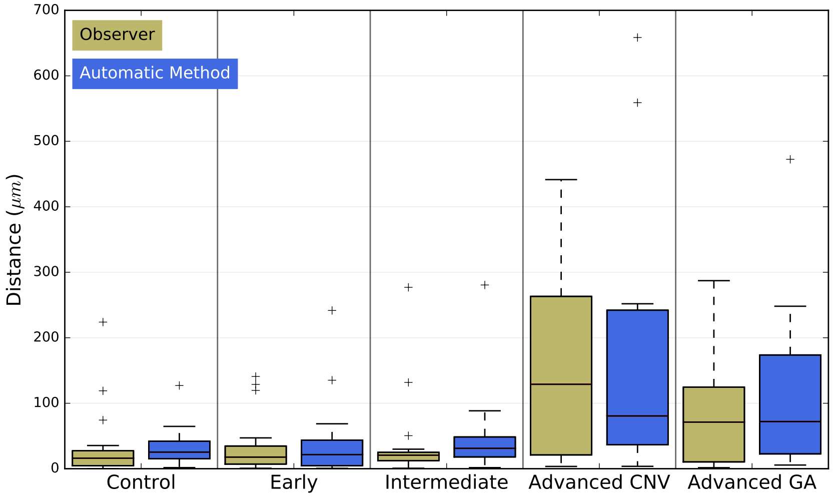

Figure 2: Boxplots showing the distance error of the proposed automatic method and the second observer compared to the reference for each AMD severity level.

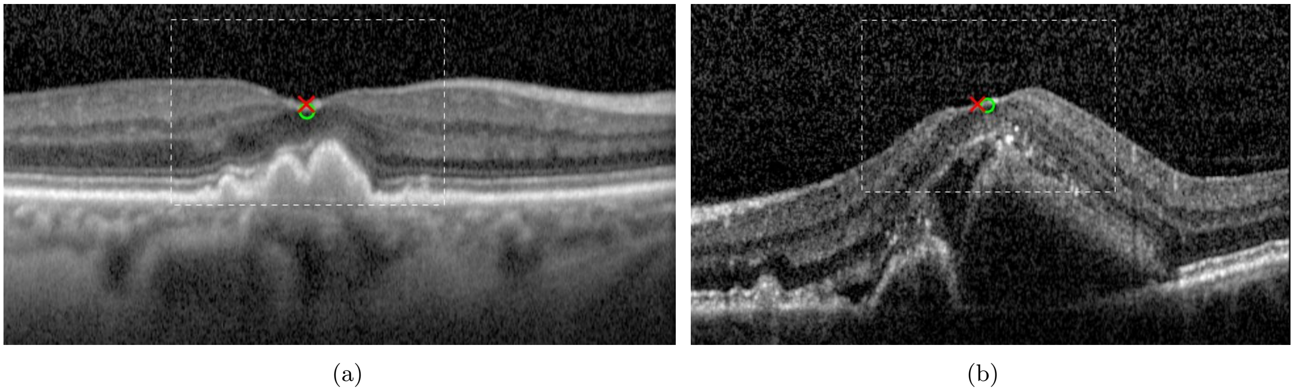

Figure 3: Examples of the automatically identified foveal center in two different OCT scans graded as (a) intermediate AMD (distance error of 5 µm) and (b) advanced AMD CNV (distance error of 81 µm). The cross represents the predicted location. The circle represents the reference annotation. The dashed rectangle indicates the receptive field for the pixel at the predicted fovea location.

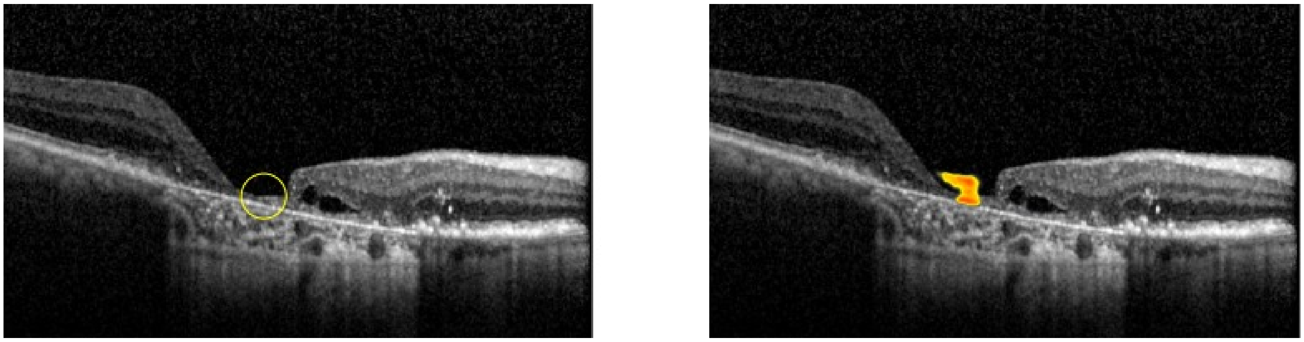

Figure 4: Original image (left), and an overlay heatmap (right) showing the probability of belonging to the fovea for each pixel as classified by the network. Yellow and red colors indicate higher probabilities. Probabilities outside the foveal region are very low and therefore not visible in the image. The reference location for the true fovea is placed at the center of the yellow circle.

4. DISCUSSION

Figure 3 shows examples of the detected foveal center in two affected retinas. The fovea is still accurately detected, even in case of deformations in the retina due to the presence of drusen or fluid. In three cases the foveal center was detected further than 750 µm from the reference location. In these cases the network still gave a positive response near the true foveal center, but a larger response at confounding irregular structures at other locations in the OCT volume. Figure 4 shows the response of the network as a heatmap. In this particular B-scan, the macular hole presents an atypical foveal morphology. The confluence of retinal layers is still visible however, but vertical alignment of the confluence is distorted (the left part of the retina descends deeper towards the choroid compared to the right part). Interestingly, it appears as if the response of the network is also separated into two vertically separated regions. This would indicate the network has learned to respond independently to a confluence in retinal layers from either side.

The OCT scans in the data set on which the method was developed is relatively heterogeneous in transverse resolution (the distance between B-scans). The current method deals with this challenge by applying a 2D method to each B-scan separately. The fovea is then placed in the B-scan with the global maximum response. This solution makes the method invariant to transverse resolution, which is a very desirable feature. On the other hand, the method can never aggregate information across multiple B-scans, which may be a limitation in very hard cases, where a human observer would typically scroll through B-scans to get an idea of the larger context.

Another limitation of the proposed method is that it has been trained only on data from Heidelberg Spectralis on healthy subjects or patients affected by AMD. Data from different OCT scanners may present different levels of noise or contrast and different retinal diseases may present different forms of retinal deformation that the method may confuse with the fovea. Therefore it needs to be validated how well the proposed method would generalize to other data sets.

5. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

To our best knowledge, this is the first attempt to apply CNNs to fovea detection in OCT scans. The resulting method is very robust and performs well even in pathological retinas.

6. CONCLUSION

We have developed a method based on dilated CNNs for the automated detection of the foveal center in OCT scans. Our method has been compared to a human observer and shows comparable performance in both healthy retinas and retinas affected by AMD. In 96 out of 99 cases the fovea was correctly identified with an accuracy that is on a par with a human observer.

ACKNOWLEDGMENTS

This work has not been submitted for publication elsewhere.

REFERENCES

[1] Yanoff, M., Fine, B. S., and Gass, J. D. M., [*Ocular pathology*] (1996).

[2] Wang, F., Gregori, G., Rosenfeld, P. J., Lujan, B. J., Durbin, M. K., and Bagherinia, H., "Automated detection of the foveal center improves SD-OCT measurements of central retinal thickness," *Ophthalmic Surgery, Lasers and Imaging Retina* **43**(6), S32–S37 (2012).

[3] Wu, J., Waldstein, S. M., Gerendas, B. S., Leitner, R., Birta, S., Langs, G., Simader, C., and Schmidt-Erfurth, U., "Disease modelling & prediction: Automated fovea detection as a key registration landmark for construction of a population reference frame," *Investigative Ophthalmology & Visual Science* **56**(7), 5917–5917 (2015).

[4] Yu, F. and Koltun, V., "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122* (2015).

[5] van de Ven, J. P., Smailhodzic, D., Boon, C. J., Fauser, S., Groenewoud, J. M., Chong, N. V., Hoyng, C. B., Klevering, B. J., and den Hollander, A. I., "Association analysis of genetic and environmental risk factors in the cuticular drusen subtype of age-related macular degeneration," (2012).

[6] Long, J., Shelhamer, E., and Darrell, T., "Fully convolutional networks for semantic segmentation," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3431–3440 (2015).

[7] van Grinsven, M. J., van Ginneken, B., Hoyng, C. B., Theelen, T., and Sánchez, C. I., "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE Transactions on Medical Imaging* **35**(5), 1273–1284 (2016).