



Deep learning approaches for segmenting Bruch's membrane opening from OCT volumes

Author

Suñot, D, Alonso-Caneiro, D, Robert Iskander, D, Collins, MJ

Published

2020

Journal Title

OSA Continuum

Version

Version of Record (VoR)

DOI

[10.1364/OSAC.403102](https://doi.org/10.1364/OSAC.403102)

Downloaded from

<http://hdl.handle.net/10072/413105>

Griffith Research Online

<https://research-repository.griffith.edu.au>



Deep learning approaches for segmenting Bruch's membrane opening from OCT volumes

DOMINIKA SUŁOT,^{1,*}  DAVID ALONSO-CANEIRO,² D. ROBERT ISKANDER,¹ AND MICHAEL J. COLLINS²

¹Department of Biomedical Engineering, Wrocław University of Science and Technology, Wrocław, Poland

²Queensland University of Technology (QUT), Contact Lens and Visual Optics Laboratory, Centre for Vision and Eye Research, School of Optometry and Vision Science, Kelvin Grove, Qld 4059, Australia

*dominika.sulot@pwr.edu.pl

Abstract: Automated segmentation of the eye's morphological features in OCT datasets is fundamental to support rapid clinical decision making and to avoid time-consuming manual segmentation of the images. In recent years, deep learning (DL) techniques have become a commonly employed approach to tackle image analysis problems. This study provides a description of the development of automated DL segmentation methods of the Bruch's membrane opening (BMO) from a series of OCT cross-sectional scans. A range of DL techniques are systematically evaluated, with the secondary goal to understand the effect of the network input size on the model performance. The results indicate that a fully semantic approach, in which the whole B-scan is considered with data augmentation, results in the best performance, achieving high levels of similarity metrics with a dice coefficient of 0.995 and BMO boundary localization with a mean absolute error of 1.15 pixels. The work further highlights the importance of fully semantic methods over patch-based techniques in the classification of OCT regions.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Bruch's membrane, which was first discovered by Karl Bruch [1], is a thin 5-layer structure located between the retinal pigment epithelium and the choroid. It partly regulates the reciprocal exchange of biomolecules, nutrients, oxygen, fluids and metabolic waste products between the retina and the general circulation [2]. This membrane stretches for over half of the eye, and can conform its shape to changes in the intraocular pressure [3]. Around the optic nerve this membrane is not present and that gap is defined as the Bruch's membrane opening (BMO). Anatomical changes in Bruch's membrane have been observed with age [4] and degenerative changes are associated with the development of age-related macular degeneration (AMD) [5,6,7].

The most commonly used methods to observe the optic nerve head are fundus photography, scanning laser ophthalmoscopy (SLO) and optical coherence tomography (OCT). While the first two provide en-face images, OCT allows cross-sectional imaging of the three-dimensional tissue morphology with high resolution. The BMO-based outer optic nerve head boundary can provide precise estimation of neuroretinal rim tissue and can be used for defining the optic nerve head (ONH) anatomy [8,9]. Additionally, parameters based on BMO can be used for determining nerve damage at the optic disc, supporting the early diagnosis of glaucoma [10,11]. Since glaucoma can lead to irreversible vision loss, early detection is crucial in order to instigate clinical treatment. BMO-based parameters such as horizontal/perpendicular rim width, minimum rim width and area are useful biomarkers for glaucoma diagnosis in addition to the commonly employed thickness of the retinal nerve fiber layer (RNFL) [12,13,14]. Some of the BMO-based parameters correlate more highly with glaucoma-related consequences such visual field loss, than do changes in RNFL thickness [10,12,15–17] and can be used in assessing the progression of glaucoma [18]. Additionally, combining BMO-based parameters with RNFL thickness can improve the efficacy in diagnosing glaucoma [13]. The location of BMO can also be used for

determining the position and shape of lamina cribrosa [19,20], which also provides information about glaucoma progression.

Accurate, automated, and objective methods to extract the BMO from a series of OCT scans and to estimate the corresponding shape parameters is fundamental for achieving useful diagnostic information that can support clinical decision making. Additionally, given that the retina and choroid layers are not anatomically present in this region, accurate estimation of the BMO location can facilitate segmentation of other nearby retinal features [21,22]. Manual extraction of the BMO, although often treated as the ground truth, is subjective and time consuming. Hence, there are various incentives to develop accurate and precise automated BMO segmentation.

Several approaches to solve the problem of automated segmentation of the BMO using image analysis methods have been proposed. Some of them use information obtained only from the OCT volumetric scan [23–25], and some use that information in combination with images from color fundus photography [26,27]. Three of these methods focus on finding a representative set of features from OCT volumes based on previously segmented intraretinal layers and other image-specific parameters such as pixel intensity. Then, these sets of features are used to classify columns or voxels into two classes (background or BMO) using standard machine learning (ML) classifiers such as a random forest or k-NN (nearest neighborhood) classifier. After classification, post-processing steps are applied to optimize the BMO segmentation [23,24,26]. Another method utilizes an image deconvolution approach to find Bruch's membrane and then based on termination of this membrane and ANN-PCA (artificial neural network-principal component analysis) algorithm, it finds the elliptical shape of BMO path [25]. Further, a method by Chen and colleagues [27] proposed a 2-step deep learning procedure. The first step is a coarse detection of a region of interest (ROI), which is located near the BMO, after which the small ROI image is further processed using a model, with an architecture similar to U-Net fully semantic network, to extract the precise location of the BMO.

Deep learning algorithms have been successfully used in problems of biomedical image segmentation, and have shown state-of-the-art performance in a number of OCT image analysis tasks, including retinal segmentation [28–31], choroid segmentation [32,33] and image classification for pathology detection [34,35]. Deep learning OCT image analysis methods can be applied at different levels (i.e., input sizes). Some of them use ANN on a single vector obtained for every image column (A-scan) [36], whereas others use a convolutional neural network (CNN) applied to a set of image columns (patch or collection of A-scan) [37], and others use the entire OCT image (B-scan) applying a fully convolutional network (FCN) with architecture like U-Net [38].

In this study, a range of deep learning techniques to segment BMO in OCT images, without using any feature extraction techniques, are developed and compared. Different ML segmentation methods (ANN, CNN, and FCN) based on a model that can utilize a different input size (image column, group of columns, and whole image) are developed. Particular attention is paid to understand the effect of the network input size on the segmentation task performance. Additionally, the effect of volumetric input data on the segmentation performance for each of these methods is assessed. Although the presented methods are applied for BMO segmentation problems, the findings of this study may be utilized in other OCT image segmentation tasks with deep learning method [36–38].

2. Material and methods

2.1. Subjects and data acquisition

Data from a retrospective study [39] including a total of 325 participants (219 females, 106 males, age ranging from 42 to 86 years) were used in this study. The study protocol was approved by the Bioethical Committee of the Wrocław Medical University (KB–332/2015) and adhered to the tenets of the Declaration of Helsinki. All participants gave written consent to participate in the

study. Data from one eye, randomly selected from every patient, was included in the dataset. All details about data acquisition and selection of participants were described elsewhere [39].

2.2. Dataset

The study included imaging of the ONH, using spectral domain optical coherence tomography (SD-OCT) (Spectralis, Heidelberg Engineering GmbH, Heidelberg, Germany). For each participant, a volumetric scan was acquired, comprised of 49 horizontal B-scans. Each B-scan is a single channel grey scale image of size 436×384 pixels (height \times width), which corresponds to physical dimensions of about 1.93×4.40 mm. An example of a SLO image with the scanning pattern and a B-scan is shown in Fig. 1. For each volumetric scan, the 16 centrally located scans with visible BMO, were manually annotated from the total of 49 scans that were collected. In total, 5200 annotated B-scans were used in this study. Thus, the dataset consists of those horizontal B-scans that intersect the BMO, the location of BMO was manually annotated in each B-scan by a trained ophthalmologist, setting the coordinates of two points that indicate this opening.

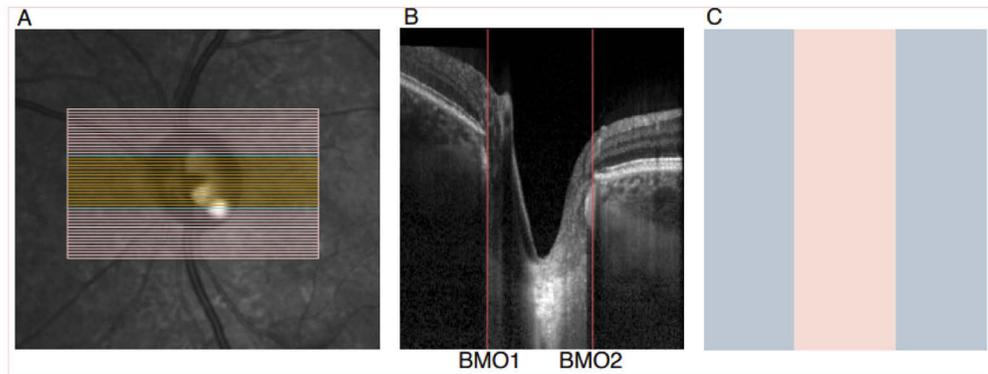


Fig. 1. SLO image with the OCT scanning protocol (A), orange lines represent the position of the annotated scans, whereas blue lines represent the additional scans used in 3-D approaches. Pink lines refer to the position of scans that have not been manually annotated. Illustrative SD-OCT B-scan with overlaid coordinates of Bruch's membrane opening. The termination points of Bruch's membrane opening are BMO1 and BMO2 (B) and the corresponding ground truth mask (C).

The dataset was randomly divided into training and testing parts using a 6 to 4 ratio. Division was made based on patients, which ensured that all images that belong to one patient were assigned to the same group. The training set was further divided into training and validation parts in 8 to 2 ratio and the validation set was used only in the training process to supervise it. The final data split was 2495/625/2080 image for training, validation and testing, respectively. To have a fair comparison of the method, this data split was kept constant in the evaluation of each method.

2.3. Preprocessing of the dataset and data augmentation

Given that the dataset used in this study consists of images and their corresponding pair of coordinates that delineate the limits of the BMO, it was necessary to prepare a ground truth mask. For every annotated scan, every pixel that was located between and including the BMO points in horizontal axis was assigned to BMO class (class 1) and every pixel either side of those points was treated as a background (class 0). An example B-scan with its corresponding mask is shown in Fig. 1(B and C).

For all 3-D approaches, additional preprocessing of the images was done to create an input dataset using three neighboring B-scans, where the BMO coordinates were taken as a reference from the middle scan. Two additional B-scans, to those 16 annotated by the ophthalmologist, were used to create 3-D images (one scan before the first annotated scan and one scan after the last annotated scan) (see Fig. 1(A), where those two additional scans are shown in blue color).

Data were collected to ensure both the BMO and the lamina cribrosa were placed within the instrument's imaging depth. Consequently, the location of the ONH within each volumetric scan was not in the same vertical position. Given this, vertical data augmentation for each of the considered methods (i.e., ANN, CNN, and FCN) was applied. That augmentation includes random vertical shift of images in a range from -90 to 90 pixels (covering about 20% of the image height).

2.4. Comparison of segmentation algorithms

In order to investigate the problem of BMO segmentation using deep learning techniques, it was decided to use different segmentation approaches. Each of them uses the same dataset, but a different input data size. The considered segmentation methods are: one dimensional (A-scan, a column from a B-scan), a two-dimensional patch based (group of consecutive A-scans) and a large two dimensional (entire image). For each of these three methods, their corresponding 3-D approaches, which extend the data to the neighboring B-scans, were also considered. The 3-D approach is an approach which uses a 3-channel image created from 3 neighboring scans, in which each channel represents different spatial information (i.e. different B-scan slice).

The general overview of the utilized methods is presented in Fig. 2. To ensure that the results obtained can be compared across the methods, the same pre-processing and the post-processing steps were used to determine the evaluation metrics. Some minor differences will be noted in the description of each specific method. The entire experiment was run on Python 3.6.4 with following libraries: keras 2.2.4, numpy 1.10.0, scikit-learn 0.22.2, scipy 1.1.0, tensorflow 1.14.0.

2.5. A-scan segmentation

The A-scan method is based on a one-dimensional input, so the input to the ANN is a single A-scan (image column). The ANN architecture, which was adapted from [36], and follows these layer dimensions: 436-200-75-10-2. Each of the 16 annotated scans per subject was divided into a set of single columns and labeled according to the previously created mask. Additionally, a 3-D version of this method was developed by extending the model input. A single A-scan from each neighboring scan (from the 3-D cube) was taken as the input to the ANN model and three of these models were connected by adding an average layer to combine the outputs of the three models, which resulted in a single model consisting of three ANN models.

2.6. Column-based segmentation

Two variations of the column-based method were developed and compared in this study. To prepare the patches, the same strategy was used for the different input images (2-D or 3-D). Overlapping patches of 16 neighboring columns (A-scans) were taken in steps of 4 pixels. Based on the masks created previously, each patch was labeled to the class that the majority of pixels in the mask belong to. The network architecture was adapted from [37] and extended to allow for 3-D input. Note that the only difference between the 2-D and 3-D approaches is the shape of input (number of input channels). For the purpose of testing, in both methods, patches were evaluated in steps of 1 pixel.

2.7. B-scan fully semantic segmentation

The last method is based on the information contained in the entire B-scan. In this approach, a deep learning model was trained with the entire set of images and ground truth masks. A fully

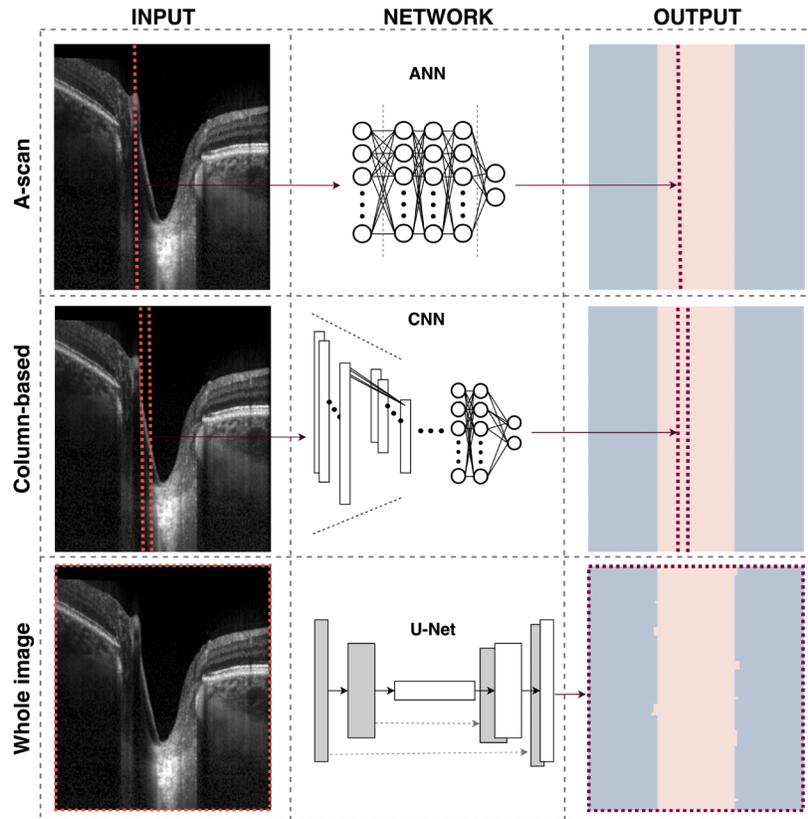


Fig. 2. General overview of the proposed segmentation methods. Columns present (left to right): input type with respect to the whole image, method used (network architecture) and direct output of the model (prediction map). Rows present an overview of the operation of the method based on (top to bottom): a single A-scan, a set of columns (column based), and an entire B-scan (whole image).

convolutional network with U-Net like architecture [38] was used. The architecture is presented in Fig. 3. Note that some padding was added to the input image to be able to down-sample it five times. At the output, the padding was removed to be able to compare the network-predicted maps with the ground truth masks. For the 3-D approach, the network architecture was the same, but with the number of channels in the input image set to 3.

Some additional post-processing was done for this specific method. As this approach provides a per-pixel level classification not a column based one, as does the previous two methods. Thus, the predicted maps can have columns in which two classes are present. To standardize the results, the output prediction maps were post-processed. If the mean value of pixels in one column was greater than 0.5, then all pixel values in that column were set to class 1 (BMO). Otherwise, they were set to class 0 (background).

Additionally, to assess the impact on performance a range of models with modifications to the standard U-net architecture were investigated. Three types of modifications were used including: incorporation of residual learning (Residual) [40], the addition of squeeze-excitation blocks (SE)[41] and with a recurrent neural network bottleneck (RNN) [42]. In case of the last method, the standard RNN layers were replaced with CNN-LSTM layers [43].

determine the coordinates at the beginning and the end of the BMO. This post-processing was conducted using the following procedure and its overview is presented in Fig. 4.

1. BMO1's position was designated as a location of the first transition from class 0 to class 1, not closer than 50 pixels (determined empirically) away from the beginning of the scan.
2. By checking each subsequent column, the location of the transition from class 1 to class 0 was temporarily marked as BMO2. The number of columns being constantly predicted as class 0 was calculated and saved as the gap width.
3. If any subsequent columns were marked as class 1 in the scan, their width (marked as temporary width in the figure) was compared with the gap width. If larger, the last of these columns was saved as BMO2. This step was repeated until the end of the image was found.

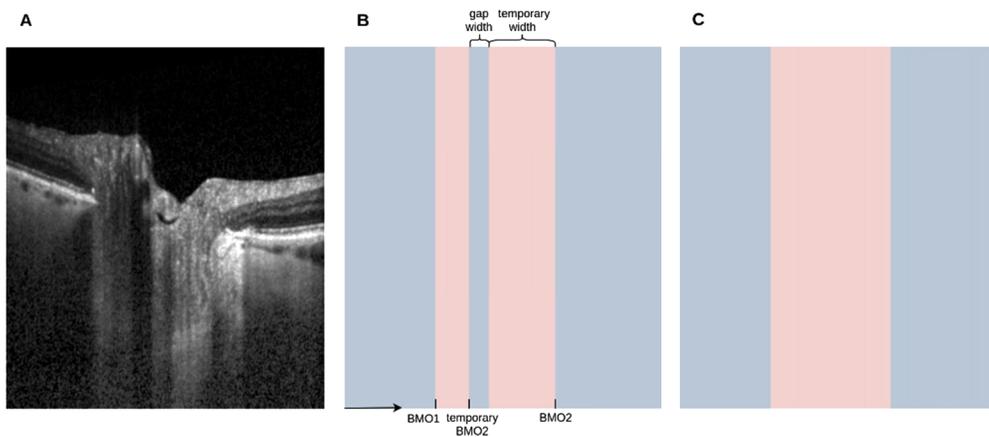


Fig. 4. General overview of the proposed postprocessing. It presents B-scan (A), sample mask with not homogeneous BMO region (B) and a clean mask (C). Width of a gap between two BMO-regions in figure B is marked as gap width. Width of the second BMO region is annotated as temporary width and because it is larger than gap width, temporary BMO2 is moved to the end of the second column and it is marked as BMO2. An arrow indicates processing direction.

Since each scan has two BMO points, the mean value of calculated distance measures of those two points is calculated and presented.

3. Results

The mean value and the standard deviation of the similarity measures are shown in Table 1 and of BMO distance measures in Table 2. For each of the methods, the results are presented in the following order: 2-D, 2-D with data augmentation, 3-D and 3-D with data augmentation. Additionally, results from inter-grader annotation is provided in both tables.

Examining the similarity measurements (Table 1), it is clear that all methods provide a reasonable performance with metrics of DSC > 0.93, JSC > 0.90 and ACC > 0.96. When comparing the methods, the fully semantic one provides superior performance compared to the other two. The best DSC results from each method are: 0.973 for the A-scan, 0.989 for the column-based and 0.994 for the fully semantic approach. Other similarity measures (i.e., JSC and ACC) also show a superior performance for the fully semantic segmentation with U-NET architecture. The result in Table 1 also shows the impact of data augmentation and the 3-D approach on each segmentation method. For the A-scan segmentation, data augmentation does not seem to improve

Table 1. Mean values and the corresponding standard deviations of the similarity measures extracted from 3 independent runs of the different deep learning models and results from inter-grader annotations. The highest values (best performance) are shown in bold font.

| Method (Data) | Input size (pixels) | DA ^a (Y/N) | DSC | JSC | ACC |
|--------------------------------|---------------------------|-----------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| <i>Inter-grader</i> | $436 \times 384 \times 1$ | N | 0.975 | 0.952 | 0.984 |
| <i>ANN (A-scan)</i> | $436 \times 1 \times 1$ | N | 0.949 ± 0.005 | 0.910 ± 0.006 | 0.964 ± 0.003 |
| | $436 \times 1 \times 1$ | Y | 0.945 ± 0.004 | 0.903 ± 0.006 | 0.960 ± 0.003 |
| | $436 \times 1 \times 3$ | N | 0.953 ± 0.027 | 0.925 ± 0.025 | 0.973 ± 0.008 |
| | $436 \times 1 \times 3$ | Y | 0.939 ± 0.025 | 0.902 ± 0.026 | 0.963 ± 0.010 |
| <i>CNN (Column-based)</i> | $436 \times 16 \times 1$ | N | 0.977 ± 0.011 | 0.956 ± 0.021 | 0.984 ± 0.007 |
| | $436 \times 16 \times 1$ | Y | 0.984 ± 0.002 | 0.969 ± 0.004 | 0.989 ± 0.001 |
| | $436 \times 16 \times 3$ | N | 0.981 ± 0.002 | 0.964 ± 0.003 | 0.987 ± 0.001 |
| | $436 \times 16 \times 3$ | Y | 0.984 ± 0.000 | 0.968 ± 0.001 | 0.989 ± 0.000 |
| <i>Fully semantic (B-scan)</i> | $436 \times 384 \times 1$ | N | 0.989 ± 0.001 | 0.979 ± 0.001 | 0.993 ± 0.000 |
| | $436 \times 384 \times 1$ | Y | 0.992 ± 0.000 | 0.983 ± 0.001 | 0.994 ± 0.000 |
| | $436 \times 384 \times 3$ | N | 0.983 ± 0.002 | 0.968 ± 0.004 | 0.989 ± 0.001 |
| | $436 \times 384 \times 3$ | Y | 0.985 ± 0.001 | 0.971 ± 0.001 | 0.990 ± 0.000 |

^aDA = data augmentation. Input size indicates height \times depth \times number of channels.

Table 2. Mean values and the corresponding standard deviations of the distance errors extracted from 3 independent runs of the different deep learning models and results from inter-grader annotations. The lowest values (best performance) are shown in bold font. All measure errors are computed in pixels.

| Method (Data) | Input size (pixels) | DA ^a (Y/N) | MAE | RMSE | MDAE | ME |
|--------------------------------|---------------------------|-----------------------|-----------------------------------|-----------------------------------|-----------------------------------|------------------------------------|
| <i>Inter-grader</i> | $436 \times 384 \times 1$ | N | 3.12 | 3.90 | 2.50 | -0.24 |
| <i>ANN (A-scan)</i> | $436 \times 1 \times 1$ | N | 6.84 ± 0.61 | 14.33 ± 2.35 | 3.33 ± 0.29 | 1.07 ± 0.41 |
| | $436 \times 1 \times 1$ | Y | 7.59 ± 0.63 | 14.78 ± 1.11 | 3.33 ± 0.29 | 1.61 ± 0.27 |
| | $436 \times 1 \times 3$ | N | 5.39 ± 1.68 | 9.33 ± 6.28 | 3.67 ± 0.14 | 2.23 ± 2.12 |
| | $436 \times 1 \times 3$ | Y | 7.19 ± 1.84 | 13.26 ± 5.29 | 3.17 ± 0.14 | 0.58 ± 2.32 |
| <i>CNN (Column-based)</i> | $436 \times 16 \times 1$ | N | 4.75 ± 0.24 | 6.09 ± 0.58 | 4.00 ± 0.00 | -0.35 ± 0.04 |
| | $436 \times 16 \times 1$ | Y | 4.67 ± 0.22 | 5.71 ± 0.57 | 4.00 ± 0.00 | -0.56 ± 0.06 |
| | $436 \times 16 \times 3$ | N | 2.55 ± 0.21 | 15.10 ± 0.12 | 1.83 ± 0.29 | -0.93 ± 0.27 |
| | $436 \times 16 \times 3$ | Y | 2.25 ± 0.06 | 14.82 ± 0.11 | 1.50 ± 0.00 | -0.99 ± 0.12 |
| <i>Fully semantic (B-scan)</i> | $436 \times 384 \times 1$ | N | 1.43 ± 0.04 | 2.25 ± 0.10 | 1.00 ± 0.00 | -0.54 ± 0.12 |
| | $436 \times 384 \times 1$ | Y | 1.15 ± 0.01 | 2.33 ± 0.39 | 1.00 ± 0.00 | -0.44 ± 0.13 |
| | $436 \times 384 \times 3$ | N | 2.02 ± 0.02 | 4.34 ± 1.17 | 1.00 ± 0.00 | -0.42 ± 0.17 |
| | $436 \times 384 \times 3$ | Y | 1.87 ± 0.10 | 3.49 ± 0.32 | 1.00 ± 0.00 | -0.41 ± 0.17 |

^aDA = data augmentation. Input size indicates height \times depth \times number of channels.

the segmentation result, but additional information from neighboring scans (3-D approach) does improve the segmentation result from 0.964 to 0.973-pixel accuracy. For the column-based segmentation with CNN architecture, both data augmentation and the 3-D approach marginally improve the segmentation results, whereas for the fully semantic segmentation method, the best result is achieved with data augmentation, but without any additional scans (2-D method). It is worth noting that the changes in similarity metrics are generally relatively small in magnitude.

The BMO distance measurements in Table 2 show a larger difference between the methods than the similarity measurements (Table 1). This is probably an indication that the overall prediction map is accurate, but some of the methods contain more outliers in the map than the others, negatively impacting the correct detection of the BMO points. However, the main findings remain the same. Looking at the absolute errors, the superior performance of the B-scan fully semantic method is evident. Examining the different metrics, it is also worth noting that the low values of the mean error indicate that models are not biased in any direction (i.e., the models make similar errors on right and left side of the BMO point). RMSE values for the A-scan and column-based segmentation method show that these models are not able to segment all of the scans; indicating that there are some outliers. In some of the scans, the BMO has poor visibility because of lower quality of the image and this could be the cause of outliers. However, the segmentation method based on the fully semantic approach, handles those cases well, as shown by the low RMSE result. MAE indicates the best performance (the lowest value) for fully semantic segmentation with U-Net architecture and data augmentation usage of 1.15 pixels. MDAE, which in this case is equal to 1 pixel, also shows that this segmentation method is capable of deriving the BMO with high accuracy.

Table 3 shows comparison between results obtained from model based on the standard U-net architecture and those with model modifications. Results from standard U-net architecture showed the highest performance for the BMO segmentation task. The models based on modified U-net like architectures achieved similar results, exceeding those from the ANN and CNN methods, although overall the performance across the different metrics were slightly inferior to the standard U-net.

Table 3. Mean values and the corresponding standard deviations of the comparison between models with different architectures. The best values are shown in bold font. All distance errors are computed in pixels.

| Architecture | DSC | JSC | ACC | MAE | RMSE | MDAE | ME |
|------------------|----------------------|----------------------|----------------------|--------------------|--------------------|--------------------|---------------------|
| <i>Standard</i> | 0.992 ± 0.000 | 0.983 ± 0.001 | 0.994 ± 0.000 | 1.15 ± 0.01 | 2.33 ± 0.39 | 1.00 ± 0.00 | -0.44 ± 0.13 |
| <i>Residuals</i> | 0.986 ± 0.001 | 0.971 ± 0.001 | 0.990 ± 0.000 | 2.12 ± 0.07 | 3.56 ± 1.01 | 1.66 ± 0.29 | -1.08 ± 0.17 |
| <i>RNN</i> | 0.984 ± 0.001 | 0.970 ± 0.001 | 0.990 ± 0.001 | 2.29 ± 0.01 | 3.88 ± 0.07 | 1.50 ± 0.00 | -1.00 ± 0.19 |
| <i>SE</i> | 0.989 ± 0.001 | 0.973 ± 0.001 | 0.990 ± 0.001 | 2.17 ± 0.06 | 3.73 ± 2.14 | 2.00 ± 0.00 | -1.07 ± 0.08 |

4. Clinical application

Although the BMO metrics derived from a single OCT scan can be used to capture the cross-sectional shape of the ONH, a 3-D analysis is commonly applied to capture the entire ONH shape. Based on the 2-D results, the best segmentation method was selected (i.e., the fully semantic method with data augmentation approach) for the subsequent 3-D analysis. Because the previous model was trained only on the scans that contained BMO, for further analysis, the selected model was fine-tuned with an additional 10 B-scans for each of the patient datasets, without the BMO present in the image (single class mask). The results obtained after this process are presented in Table 4.

Table 4. Result from one fine-tuned fully semantic model with U-Net like architecture.

| Method (Data) | DSC | JSC | ACC | MAE | RMSE | MDAE | ME |
|---|-------|-------|-------|------|------|------|-------|
| <i>Fine-tuned fully semantic (B-scan)</i> | 0.983 | 0.969 | 0.989 | 2.43 | 4.74 | 1.5 | -1.38 |

Next, the entire volumetric scan (sets of 49 B-scans) was processed sequentially, extracting two BMO points for every processed scan, if present. The collection of points was used for estimating

the BMO outline. In order to compare this result to a ground truth, the testing dataset was manually annotated by a masked research trainee (all 49 B-scans). When comparing the results for the previously annotated 16 scans, the mean absolute difference (± 1 standard deviation) was only 3.1 ± 2.3 pixels. Figure 5 shows an example of the SLO image corresponding to a given OCT volume consisting of 49 B-scans with the segmented BMO outline.

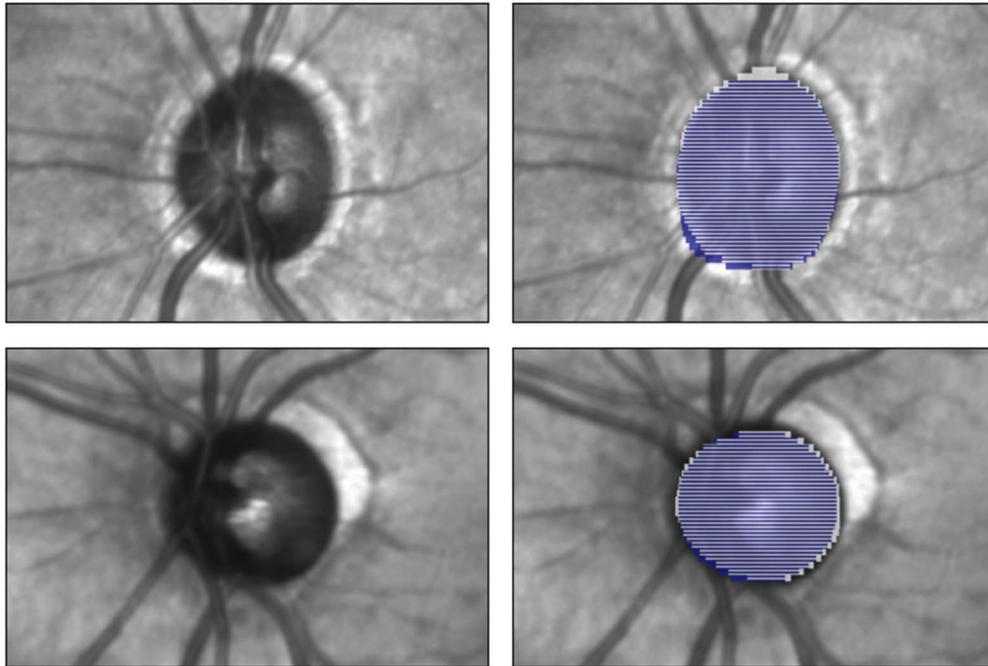


Fig. 5. Illustrative examples of original SLO image (left column) and the same image with overlaid BMO outlines (right column) segmented from the corresponding OCT volume consisting of 49 B-scans. Blue color indicates the BMO outline from manual annotation and the grey one, the BMO outline segmented by the proposed model.

Table 5 shows the mean values and the corresponding standard deviations calculated across test subjects of the two measures assessing the performance (DL vs ground-truth): DSC and the unsigned border positioning error (UBPE) [44]. The UBPE is the minimum value of the Euclidean distance for each point in the BMO outline predicted by a model with all points in ground truth annotations. Raw model predictions were checked, without post-processing, as well as with the modeling of the shape of the BMO, using a convex hull algorithm. Both approaches show high similarity values (DSC) of the predicted BMO outline compared to the ground truth of 0.953 and 0.959, and high accuracy of prediction with $11.4 \mu\text{m}$ and $9.8 \mu\text{m}$ of UBPE, for raw predictions and with convex hull, respectively.

Table 5. Mean value and standard deviation (calculated across test subjects) of dice coefficient and unsigned border positioning error for BMO outline compared to the outline calculated with annotated scans.

| Method | DSC | UBPE (μm) |
|---|-------------------------------------|----------------------------------|
| <i>Raw model predictions</i> | 0.953 ± 0.030 | 11.4 ± 51.9 |
| <i>Convex hull on model predictions</i> | 0.959 ± 0.032 | 9.8 ± 31.9 |

The main difference between these approaches is the value of the UBPE standard deviation, which is higher for the raw model prediction. This indicates the advantage of using the convex hull approach to model the data and reduce variability.

5. Discussion

The results of this study, in which three different methods were considered (i.e., A-scan, column-based, and full image), show that all considered deep learning techniques can be successfully used to segment the Bruch's membrane opening from OCT volumes with high similarity measures (i.e., DSC, JSC, and ACC). Nevertheless, the fully semantic approach with data augmentation appears to provide a superior performance over the other methods. This is particularly evident when the distance error measures for the BMO location are considered. This method achieved an MAE in determining the BMO position of 1.15 pixels, which corresponds to 13.17 μm .

Part of the motivation of this work was to better understand the influence that the size of the input has on the methods performance (final results of segmentation) and thus it was decided to use and compare three different methods. The first of them, ANN, considered a single A-scan. The next, CNN, used a 16 column-based input (patch-based) for segmentation and the last method used the entire B-scan to perform the segmentation task. Although a number of similar methods to segment and classify OCT image have been proposed previously [24–28], to the best of the authors' knowledge, none of them explore in detail the effect of input size on the performance across different deep learning techniques. The results demonstrate that the performance (similarity metrics) increase as the size of the input increases, pointing to the positive effect of information from the whole image on the segmentation of the BMO position. Thus, were possible, fully semantic methods may need to be considered for classification/segmentation tasks that have been previously applied in an image subsection [27]. In addition, in this study we explore the benefit of using the neighboring scans (i.e., the 3-D approach) to improve segmentation results. Examining the similarity metrics, only marginal differences are observed across the methods. The A-scan and column-based segmentation show some improvement, whereas in the case of the fully semantic segmentation, the 3-D approach slightly decreases the performance. It is evident from these findings that there is not a clear benefit from the use of 3-D data for this particular application.

To fully assess the potential use of this method as a clinical or research tool, the fully semantic segmentation with data augmentation was evaluated to extract the entire BMO outline from OCT volumes. All 49-scans were assessed by the trained model and it resulted in a BMO outline segmentation with a Dice coefficient value of 0.959. This shows the potential of using the proposed model for this purpose, making it possible to accurately extract the size of ONH. Although the obtained results cannot be directly compared to those reported elsewhere, because of different datasets, it is worth noting that in work [26] the authors reported a DSC value of 0.855 and in [23] of 0.853, which are significantly lower than the results obtained by the proposed method of 0.959. Additionally, the UBPE in those papers was 67 μm and 61 μm , respectively, which is a larger error than the 9.8 μm achieved here. This further highlights the potential of the proposed method.

Of practical interest is also the computational complexity of the developed methods. When a single segmentation task is considered for the entire image (B-scan), an average (mean \pm 1 standard deviation) processing time of 0.028 ± 0.003 s, 53.61 ± 2.67 s, and 2.36 ± 0.12 s, was obtained for the ANN, CNN and U-net methods, respectively (simulations ran on a personal computer with CPU Intel Core i5-5200U). This indicates that the superior performance of the fully semantic method comes at a time cost when compared to the A-scan method, and that the column-based method cannot match any of the other two methods in terms of computational efficiency.

It is important to acknowledge that in our comparison of different network architectures the number of trainable parameters is different. It was 270 857 (881 367), 24 564 802 (24 567 490), and 12 589 425 (12 589 713) for ANN (ANN 3-D), CNN (CNN 3-D), and U-net (U-net 3-D), respectively. However, in the comparison, it was intended to ascertain whether a particular type of network architecture provides better segmentation result than the other with the size of the input being adequately matched to the particular type of network architecture to maximize its performance.

The findings of the effect of input size need to be confirmed with other larger datasets and this is a study limitation that needs to be further explored. Additionally, it is likely that the performance, particularly error distance metrics, of the ANN and CNN methods could be improved with more sophisticated BMO boundary detection techniques. However, for consistency across the methods, a relatively simple post-processing rule was applied. Although this could be seen as a limitation of the proposed method, it also further demonstrates the robustness of the fully semantic method.

6. Conclusion

There is a continuing interest in the development of machine learning methods for segmenting morphological features of the eye in OCT volumes. In this work, a range of deep-learning-based automated segmentation methods of the Bruch's membrane opening were considered, as this particular retinal feature can provide important clinical information supporting early glaucoma diagnosis [10,11]. A systematic approach for developing and selecting the most appropriate deep learning technique was taken, in which the input size governed the type of method used and its architecture. Using retrospective OCT volumetric data from 325 participants, the best segmentation results, in terms of similarity metrics and position errors, have been achieved with a fully semantic approach with standard U-Net like architecture, followed by modified U-net architectures, the column-based (CNN) method and the A-scan (ANN) method. Additionally, the performance of the selected method on segmentation of the entire Bruch's membrane opening outline was considered for the entire volumetric scan, achieving an unsigned positioning border error of 9.8 μm , which shows the potential of this method.

The contributions of this work can be summarized as follows:

- three DL techniques to segment BMO in OCT images were developed and compared,
- the effect of the network input size on the segmentation task performance was evaluated, and
- the effect of volumetric (3-D) input data on the segmentation performance for each of the DL techniques was assessed.

Although the presented methods are purposely designed for segmenting the Bruch's membrane opening, the findings of this study may be generalized to other OCT image segmentation tasks using deep learning methods. Additionally, given the BMO is an area where the retinal layers are not present, the proposed model also provides useful information to support retina and choroid segmentation methods [21,22]

Funding

Interdisciplinary Doctoral Studies Projects at WUST; Rebecca L. Cooper Medical Research Foundation; National Health and Medical Research Council (APP1186915); Narodowe Centrum Nauki (2018/29/B/ST7/02451); Narodowa Agencja Wymiany Akademickiej (PPI/APM/2019/1/00085/DEC/1).

Disclosures

The authors declare no conflicts of interest.

References

1. K. Bruch, "Untersuchungen zur kenntniss des körnigen pigments der wirbelthiere in physiologischer und pathologischer hinsicht," Meyer und Zeller (1884).
2. J. C. Booiij, D. C. Baas, J. Beisekeeva, T. G. Gorgels, and A. A. Bergen, "The dynamic nature of Bruch's membrane," *Prog. Retinal Eye Res.* **29**(1), 1–18 (2010).
3. C. A. Curcio, M. Johnsonet, and M. Johnson, "Structure, function, and pathology of Bruch's membrane," *Retina* **1**(2), 465–481 (2013).
4. J. Johnstone, M. Fazio, K. Rojananuangnit, B. Smith, M. Clark, C. Downs, C. Owsley, M. J. Girard, J. M. Mari, and C. A. Girkin, "Variation of the axial location of Bruch's membrane opening with age, choroidal thickness, and race," *Invest. Ophthalmol. Visual Sci.* **55**(3), 2004–2009 (2014).
5. M. C. Killingsworth, J. P. Sarks, and S. H. Sarks, "Macrophages related to Bruch's membrane in age-related macular degeneration," *Eye* **4**(4), 613–621 (1990).
6. I. Bhutto and G. Luty, "Understanding age-related macular degeneration (AMD): relationships between the photoreceptor/retinal pigment epithelium/Bruch's membrane/choriocapillaris complex," *Mol. Aspects Med.* **33**(4), 295–317 (2012).
7. M. C. Killingsworth, J. P. Sarks, and S. H. Sarks, "Macrophages related to Bruch's membrane in age-related macular degeneration," *Eye* **4**(4), 613–621 (1990).
8. B. C. Chauhan and C. F. Burgoyne, "From clinical examination of the optic disc to clinical assessment of the optic nerve head: a paradigm change," *Am. J. Ophthalmol.* **156**(2), 218–227.e2 (2013).
9. A. S. Reis, N. O'Leary, H. Yang, G. P. Sharpe, M. T. Nicolela, C. F. Burgoyne, and B. C. Chauhan, "Influence of clinically invisible, but optical coherence tomography detected, optic disc margin anatomy on neuroretinal rim evaluation," *Invest. Ophthalmol. Visual Sci.* **53**(4), 1852–1860 (2012).
10. D. R. Muth and C. W. Hirneiß, "Structure–function relationship between Bruch's membrane opening–based optic nerve head parameters and visual field defects in glaucoma," *Invest. Ophthalmol. Visual Sci.* **56**(5), 3320–3328 (2015).
11. B. C. Chauhan, N. O'Leary, F. A. AlMobarak, A. S. C. Reis, H. Yang, G. P. Sharpe, D. M. Hutchison, M. T. Nicolela, and C. F. Burgoyne, "Enhanced detection of open-angle glaucoma with an anatomically accurate optical coherence tomography–derived neuroretinal tim parameter," *Ophthalmology* **120**(3), 535–543 (2013).
12. L. Reznicek, S. Burzer, A. Laubichler, A. Nasser, C. P. Lohmann, N. Feucht, M. Ulbig, and M. Maier, "Structure–function relationship comparison between retinal nerve fibre layer and Bruch's membrane opening–minimum rim width in glaucoma," *Int. J. Ophthalmol.* **10**(10), 1534 (2017).
13. J. M. Gmeiner, W. A. Schrems, C. Y. Mardin, R. Laemmer, F. E. Kruse, and L. M. Schrems-Hoesl, "Comparison of Bruch's membrane opening minimum rim width and peripapillary retinal nerve fiber layer thickness in early glaucoma assessment," *Invest. Ophthalmol. Visual Sci.* **57**(9), OCT575–OCT584 (2016).
14. P. Enders, W. Adler, F. Schaub, M. M. Hermann, T. Dietlein, C. Cursiefen, and L. M. Heindl, "Novel bruch's membrane opening minimum rim area equalizes disc size dependency and offers high diagnostic power for glaucoma," *Invest. Ophthalmol. Visual Sci.* **57**(15), 6596–6603 (2016).
15. E. J. Lee, K. M. Lee, H. Kim, and T. W. Kim, "Glaucoma diagnostic ability of the new circumpapillary retinal nerve fiber layer thickness analysis based on Bruch's membrane opening," *Invest. Ophthalmol. Visual Sci.* **57**(10), 4194–4204 (2016).
16. G. Rebolleda, A. Casado, N. Oblanca, and F. J. Muñoz-Negrete, "The new Bruch's membrane opening–minimum rim width classification improves optical coherence tomography specificity in tilted discs," *Clin. Ophthalmol.* **10**, 2417–2425 (2016).
17. P. Enders, F. Schaub, W. Adler, R. Nikoluk, M. M. Hermann, and L. M. Heindl, "The use of Bruch's membrane opening–based optical coherence tomography of the optic nerve head for glaucoma detection in microdiscs," *Br. J. Ophthalmol.* **101**(4), 530–535 (2017).
18. A. Belghith, C. Bowd, F. A. Medeiros, N. Hammel, Z. Yang, R. N. Weinreb, and L. M. Zangwill, "Does the location of Bruch's membrane opening change over time? Longitudinal analysis using San Diego automated layer segmentation algorithm (salsa)," *Invest. Ophthalmol. Visual Sci.* **57**(2), 675–682 (2016).
19. G. Rebolleda, J. García-Montesinos, E. De Dompablo, N. Oblanca, F. J. Muñoz-Negrete, and J. J. González-López, "Bruch's membrane opening changes and lamina cribrosa displacement in non-arteritic anterior ischaemic optic neuropathy," *Br. J. Ophthalmol.* **101**(2), 143–149 (2017).
20. P. Krzyżanowska-Berkowska, A. Melińska, I. Helemejko, and D. R. Iskander, "Evaluating displacement of lamina cribrosa following glaucoma surgery," *Graefe's Arch. Clin. Exp. Ophthalmol.* **256**(4), 791–800 (2018).
21. D. Alonso-Caneiro, S. A. Read, and M. J. Collins, "Automatic segmentation of choroidal thickness in optical coherence tomography," *Biomed. Opt. Express* **4**(12), 2795–2812 (2013).
22. M. Heisler, M. Bhalla, J. Lo, Z. Mammo, S. Lee, M. J. Ju, M. F. Beg, and M. V. Sarunic, "Semi-supervised deep learning based 3-D analysis of the peripapillary region," *Biomed. Opt. Express* **11**(7), 3843–3856 (2020).
23. K. Lee, M. Niemeijer, G. M. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abramoff, "Segmentation of the optic disc in 3-D OCT scans of the optic nerve head," *IEEE Trans. Med. Imaging* **29**(1), 159–168 (2010).
24. M. S. Miri, M. D. Abramoff, Y. H. Kwon, M. Sonka, and M. K. Garvin, "A machine-learning graph-based approach for 3-D segmentation of Bruch's membrane opening from glaucomatous SD-OCT volumes," *Med. Image Anal.* **39**, 206–217 (2017).

25. A. Belghith, C. Bowd, R. N. Weinreb, and L. M. Zangwill, "A hierarchical framework for estimating neuroretinal rim area using 3-D spectral domain optical coherence tomography (SD-OCT) optic nerve head (ONH) images of healthy and glaucoma eyes," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 3869–3872 (2014).
26. M. S. Miri, M. D. Abràmoff, K. Lee, M. Niemeijer, J. K. Wang, Y. H. Kwon, and M. K. Garvin, "Multimodal segmentation of optic disc and cup from SD-OCT and color fundus photographs using a machine-learning graph-based approach," *IEEE Trans. Med. Imaging* **34**(9), 1854–1866 (2015).
27. Z. Chen, P. Peng, H. Shen, H. Wei, P. Ouyang, and X. Duan, "Region-segmentation strategy for Bruch's membrane opening detection in spectral domain optical coherence tomography images," *Biomed. Opt. Express* **10**(2), 526–538 (2019).
28. L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express* **8**(5), 2732–2744 (2017).
29. J. Hamwood, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers," *Biomed. Opt. Express* **9**(7), 3049–3066 (2018).
30. S. K. Devalla, P. K. Renukanand, B. K. Speedhar, G. Subramanian, L. Zhang, S. Perera, J. Mari, K. S. Chin, T. A. Tun, N. G. Strouthidis, T. Aung, A. H. Thiery, and M. J. A. Griaard, "DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images," *Biomed. Opt. Express* **9**(7), 3244–3265 (2018).
31. P. Zang, J. Wang, T. T. Hormel, L. Liu, D. Huang, and Y. Jia, "Automated segmentation of peripapillary retinal boundaries in oct combining a convolutional neural network and a multi-weights graph search," *Biomed. Opt. Express* **10**(8), 4340–4352 (2019).
32. J. Kugelman, D. Alonso-Caneiro, S. A. Read, J. Hamwood, S. J. Vincent, F. K. Chen, and M. J. Collins, "Automatic choroidal segmentation in OCT images using supervised deep learning methods," *Sci. Rep.* **9**(1), 13298–13 (2019).
33. X. Sui, Y. Zheng, B. Wei, H. Bi, J. Wu, X. Pan, Y. Yin, and S. Zhang, "Choroid segmentation from optical coherence tomography with graph-edge weights learned from deep convolutional neural networks," *Neurocomputing* **237**, 332–341 (2017).
34. M. Treder, J. L. Laueremann, and N. Eter, "Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning," *Graefes Arch. Clin. Exp. Ophthalmol.* **256**(2), 259–265 (2018).
35. M. Shah, A. R. Ledo, and J. Rittscher, "Automated classification of normal and Stargardt disease optical coherence tomography images using deep learning," *Acta Ophthalmologica* (2020).
36. K. K. Dansingani, K. K. Vupparaboina, S. T. Devarkonda, S. Jana, J. Chhablani, and K. B. Freund, "Amplitude-scan classification using artificial neural networks," *Sci. Rep.* **8**(1), 12451 (2018).
37. J. Loo, L. Fang, D. Cunefare, G. Jaffe, and S. Farsiu, "Deep longitudinal transfer learning-based automatic segmentation of photoreceptor ellipsoid zone defects on optical coherence tomography images of macular telangiectasia type 2," *Biomed. Opt. Express* **9**(6), 2681–2698 (2018).
38. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, Springer, Cham, 234–241 (2015).
39. P. Krzyżanowska-Berkowska, K. Czajor, P. Syga, and D. R. Iskander, "Lamina cribrosa depth and shape in glaucoma suspects. Comparison to glaucoma patients and healthy controls," *Curr. Eye Res.* **44**(9), 1026–1033 (2019).
40. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE (2016).
41. A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," *International conference on medical image computing and computer-assisted intervention*, Springer, Cham, 421–429 (2018).
42. F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville, "Reseg: A recurrent neural network-based model for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2016).
43. J. Wang, L. C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, **2**, 225–230 (2016).
44. M. Sonka, H. Vaclav, and B. Roger, "Image processing, analysis, and machine vision," Cengage Learning (2014).