Vendor Independent Cyst Segmentation in Retinal SD-OCT Volumes using a Combination of Multiple Scale Convolutional Neural Networks

Freerk G. Venhuizen[a], Mark J.J.P. van Grinsven[a], Carel B. Hoyng[b], Thomas Theelen[b], Bram van Ginneken[a], and Clara I. Sánchez[a]

[a]Diagnostic Image Analysis Group, RadboudUMC Nijmegen, the Netherlands
[b]Department of Ophthalmology, RadboudUMC Nijmegen, the Netherlands

1 Purpose

Major causes of blindness in developed countries are retinal vascular diseases, essentially neovascular age-related macular degeneration (AMD) [1, 2], retinal vein occlusion (RVO) and diabetic maculopathy (DMP). Automated computer-aided detection and diagnosis (CAD) systems capable of detecting, classifying and quantifying the characteristics of the pathology associated with these retinal diseases is highly beneficial in the diagnosis, treatment prediction and the assessment of treatment progression [3]. Among others, the presence of retinal cysts are an important biomarker in AMD and RVO, thus their detection and segmentation is beneficial to clinical disease analysis. Optical Coherence Tomography (OCT) imaging has the ability to visualize and analyze the morphology of the retina, as well as its abnormalities [4,5]. Due to the technological advances in OCT imaging with regard to imaging speed, image quality, and functional analysis, OCT is rapidly becoming one of the main imaging modalities used in clinical practice, and for the quantification and analysis of disease-specific biomarkers such as cyst volume. In this work we propose a fully automatic CAD system for retinal cyst segmentation in OCT volume scans. The system is capable of detecting cysts in OCT volumes acquired with OCT scanners from different vendors, for which the amount of noise, image quality and contrast varies strongly.

2 Materials and Methods

2.1 OCT Data

For this study a total of 30 OCT volumes, containing a wide variety of retinal cysts together with manual cyst delineations, were provided by the OPTIMA cyst segmentation challenge. The resolution and density of the volumes vary from 512x496 to 512x1024 and 49 to 128 B-scans, respectively. The challenge data consists of a training data, stage 1 testing set and stage 2 testing set with 15, 8 and 7 scans, respectively. At the current time point only the training set, containing 15 scans, has been made available. The scans were acquired using four different OCT scanners. An overview of the dataset is shown in Table 1. To

Table 1. Challenge dataset

Set	Spectralis	Cirrus	Topcon	Nidek	Total
Training	4	4	4	3	15
Testing 1	2	2	2	2	8
Testing 2	2	2	2	1	7

analyze the performance of the system in the absence of the dedicated test set, we split the provided training data into a training and a validation set, which will be used for later performance evaluation. The validation set is formed of one randomly selected scan from each OCT vendor (4 scans in total).

2.2 Overview of CAD Pipeline

The pipeline of the CAD system is visualized schematically in Figure 1 and follows a two-stage approach. In the first stage, multiple convolutional neural networks (CNNs) [6] are used to obtain a segmentation at different image scales. In the second stage, the individual scale segmentations are merged, redefining the borders of the segmented cysts by combining local information obtained with the lower scale network with contextual information obtained from the higher scale networks [7].

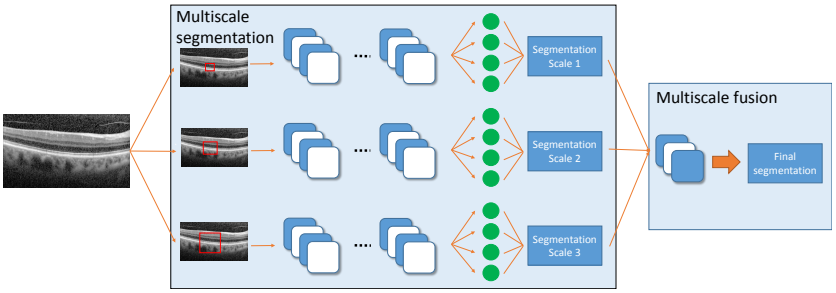


Fig. 1. Two-stage proposed cyst segmentation approach based on multiple scales CNNs: In the first stage every B-scan is processed at three different scales with an increasing contextual window. In the next stage, the segmentations obtained from each scale are fused to obtain the final segmentation.

2.3 First stage: Multiple scale segmentation

Three different CNNs are independently trained at different scales to predict the label of a single pixel as belonging to a cyst or background by considering a certain region around the pixel of interest. Using a small neighbourhood results in a sharp delineation of the boundaries of the cyst but also in many false positives, even in regions outside the retina. Using increasingly larger regions around the

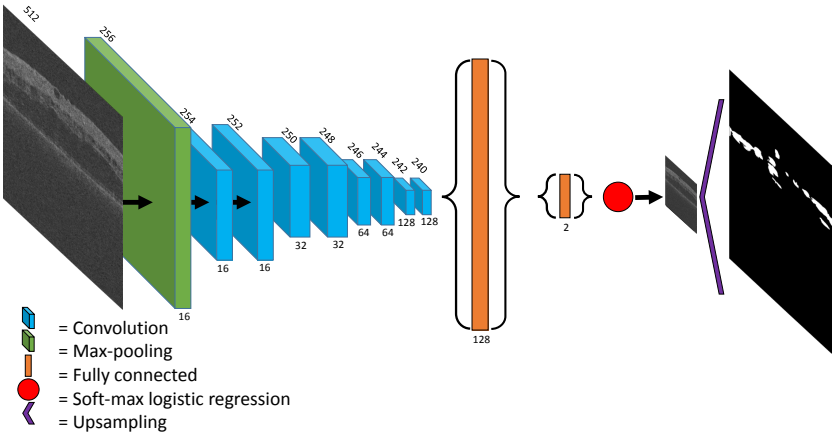


Fig. 2. Schematic overview of individual convolutional neural network architecture.

pixel, i.e. larger scale, spatially distant information is included and a rough but accurate delineation of the cysts is obtained. By combining the higher scales, which are responsible for the localisation of cysts, with the lower scales, which are focused on refining the borders of the cyst, an accurate cyst delineation is obtained [7].

Network Architecture Each network is an individually trained deep convolutional neural network, comprised of eight convolutional layers based on the Oxfordnet [8] architecture which only uses 3×3 filters. An overview of the network architecture is shown in Figure 2. The input of the first layer is a 2-dimensional image patch extracted from the OCT volume centered on the pixel to be evaluated. The size of the patches is 21×21, 41×41 and 81×81 for the three different networks, respectively. After every convolutional layer, the resulting feature maps are passed through a non-linear mapping (RELU). Spatial 2 by 2 pixels max-pooling with a stride of 2 pixels was applied prior to the first convolutional layer for the network using the 41×41 patches. Similarly, for the network processing the 81×81 patches, spatial 4 by 4 pixels max-pooling with a stride of 4 was applied. Max pooling functions are applied as a downsampling operator on the input patch, keeping the size of the feature maps and number of system parameters manageable.

Training the System A set of 4400000 million patches per scale were randomly extracted from the training data, i.e. 200000 positive and 200000 negative patches per OCT volume. Patches labeled as cyst were sampled with replacement to obtain a balanced distribution of classes. Network parameters were optimized during the training phase of the network using stochastic gradient descent. The networks for each scale were trained independently, without parameter sharing or pretrained initializations, in order to discern distinct features for each scale.

After training the networks on patch level, the three networks are convolutionalized [7] by transforming the fully connected layers into convolutional layers, in order to obtain a classification over a whole OCT volume in a single pass through the network and speed up the classification process. As the networks operating at higher scales include a downsampling step, in-network bilinear upsampling is performed before combination to obtain a full resolution output.

2.4 Second stage: Multiscale fusion

After the first stage a pair of probabilities per scale and per pixel are obtained, indicating the confidence of that pixel belonging to a cyst and to the background, respectively. A combination of these confidence scores is used to obtain a final classification score. We first create a binary segmentation for every scale by subtracting the two probability scores per pixel, and subsequent thresholding. We optimized the threshold based on the average dice score over the training data. Finally, we combine the binary segmentation maps of the three scales by applying an AND operation.

2.5 Postprocessing

In order to discard detected regions outside the retina we remove detections outside the delimiting retina layers. These layers are detected by applying the Iowa Reference Algorithms (Retinal Image Analysis Lab, Iowa Institute for Biomedical Imaging, Iowa City, IA) to find the boundary of the inner limiting membrane (ILM) and the Boundary of myoid and ellipsoid of inner segments (BMEIS) [9–11]. Only detections within these two boundaries are being considered as cysts. Additionally, we remove regions smaller than the minimum cyst size in the training set in order to discard spurious detection due to noise. A hole filling algorithm based on connected component analysis is finally applied to fill holes in binary segmentations without affecting the outer border of the segmented cyst.

3 Results

As only the training dataset has been released at the current timepoint, the validation set was used for performance evaluation. The provided manual cyst delineations by two human graders were used as reference standard for validation. The Sorensen dice coefficient [12] is used to asses the performance.

The first two columns of Table 2 give an overview of the dice coefficients obtained by the proposed system and calculated over the entire OCT volume when compared to graders 1 and 2, respectively. The third column shows the performance of the system compared to the delineations performed by the intersection of the two graders. To give an indication of human performance, the third column indicates the dice coefficient between the two graders, as a measure

Table 2. Dice coefficients of the obtained cyst segmentation for the entire volume.

OCT Traindata	**Dice Obs1**	**Dice Obs2**	**Dice Obs1 ∩ Ob2**	**Inter-rater variability**
Cirrus 1	0.696	0.717	0.613	0.691
Cirrus 2	0.853	0.845	0.821	0.887
Cirrus 3	0.644	0.612	0.568	0.762
Nidek 1	0.784	0.743	0.698	0.813
Nidek 2	0.784	0.760	0.701	0.818
Spectralis 1	0.729	0.763	0.679	0.797
Spectralis 2	0.841	0.845	0.814	0.905
Spectralis 3	0.726	0.752	0.652	0.682
Topcon 1	0.636	0.629	0.561	0.681
Topcon 2	0.541	0.542	0.471	0.678
Topcon 3	0.780	0.744	0.742	0.851
Average	0.728 ±0.09	0.726±0.09	0.665±0.10	0.778±0.09
OCT Testdata				
Cirrus 4	0.424	0.380	0.422	0.650
Nidek 3	0.284	0.266	0.220	0.662
Spectralis 4	0.644	0.674	0.633	0.831
Topcon 4	0.516	0.524	0.520	0.813
Average	0.469±0.15	0.461±0.18	0.448±0.18	0.739±0.09

of inter-rater variability. The same analysis has been performed for the central 3 mm region of the retina. These results are shown in Table 3. On average, the result on the training set (0.728 and 0.742 for the results in the entire volume and inside the 3mm region, respectively) is approaching the performance of the graders (0.778). The average dice coefficient on the validation set is 0.469. When comparing to the intersection of the two independent graders, a slight drop in performance can be observed for all images, obtaining an average score of 0.665 and 0.448 for the training and validation images, respectively. Additional sensitivity analysis was performed on the system, with resulting sensitivities of 0.900 (±0.093) and 0.57(±0.175) for the training and validation set, respectively.

4 Discussion

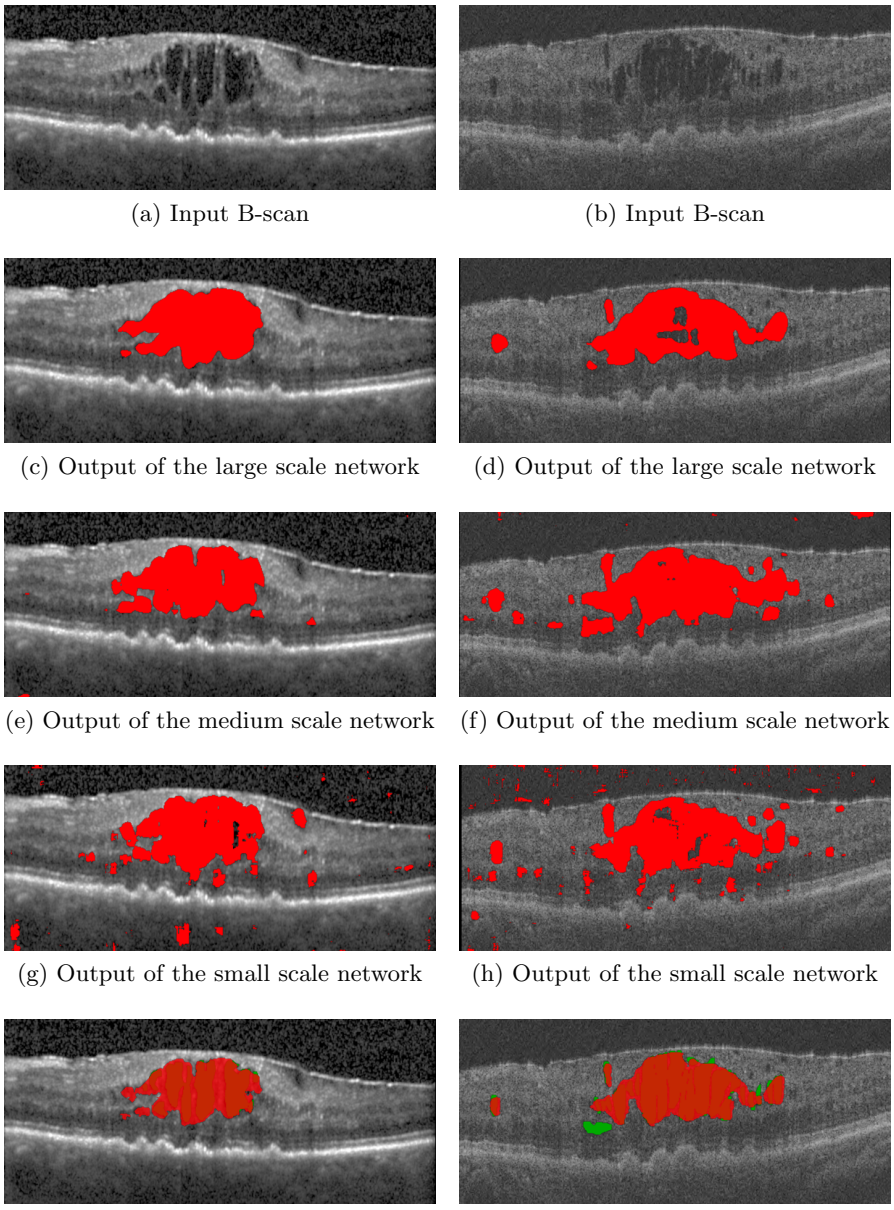
A CAD system for a fully automatic vendor independent segmentation of cysts based on OCT images was presented in this paper. The method used a combination of CNN predictions at different scales in order to provide an effective method to include contextual information for pixel classification. In order to allow a dense prediction for a full OCT volume, an efficient implementation using

Table 3. Dice coefficients of the obtained cyst segmentation within the central 3mm diameter retinal region.

OCT Traindata	**Dice Obs1**	**Dice Obs2**	**Dice Obs1 ∩ Ob2**	**Inter-rater variability**
Cirrus 1	0.705	0.726	0.622	0.771
Cirrus 2	0.854	0.845	0.821	0.941
Cirrus 3	0.714	0.687	0.594	0.811
Nidek 1	0.786	0.746	0.70	0.918
Nidek 2	0.801	0.772	0.772	0.922
Spectralis 1	0.730	0.764	0.680	0.850
Spectralis 2	0.841	0.845	0.814	0.940
Spectralis 3	0.727	0.752	0.572	0.707
Topcon 1	0.650	0.643	0.238	0.394
Topcon 2	0.569	0.568	0.509	0.786
Topcon 3	0.785	0.779	0.746	0.930
Average	0.742 ±0.08	0.738±0.08	0.638±0.16	0.815±0.16
OCT Testdata				
Cirrus 4	0.454	0.403	0.257	0.539
Nidek 3	0.378	0.347	0.294	0.857
Spectralis 4	0.654	0.684	0.638	0.883
Topcon 4	0.517	0.525	0.521	0.873
Average	0.500±0.12	0.489±0.15	0.428±0.18	0.788±0.16

convolutionalization was applied, reducing the classification time dramatically. Figure 3 shown an example of the segmentation process for two different OCT scans. It can be observed that the network trained at larger scale accurately localizes the cysts, but due to the large contextual window, the accuracy near the cyst borders is low. The two lower scales achieve a more accurate border delineation but the amount of false positives increases substantially. The combination of the different scales provides a good trade-off between accurate cyst segmentation and number of false detections.

The system was trained and evaluated on a very small set of images, which makes it difficult to generalize to new unseen images. It can be observed that the performance of the system is highly dependent on the OCT scanner used to acquire the scan. The performance of the system when applied to higher quality OCT images, like those obtained by the Spectralis and the Topcon scanners, is substantially higher than those obtained using scans from the Cirrus and Nidek scanner. Specially, low performance was obtained for the Cirrus scan included in the validation set. This can be explained due to the small size of the manually segmented cysts on this volume. Dice coefficient measurements are highly sensi-



(a) Input B-scan (b) Input B-scan

(c) Output of the large scale network (d) Output of the large scale network

(e) Output of the medium scale network (f) Output of the medium scale network

(g) Output of the small scale network (h) Output of the small scale network

(i) Combination of the three scales (j) Combination of the three scales

Fig. 3. Intermediate and final output of the proposed method (red regions) a Spectralis (left) and a Topcon (right) B-scans. In the last row, green indicates false negative regions annotated by both observers.

tive to a over/under-segmentation of small regions, having a detrimental effect on the final result. When limiting the analysis to the central 3 mm region of the retina, the performance increases for all scans (see Table 3). A more accurate result is needed for this central region as the presence of any abnormality within this area is correlated with a higher risk of vision loss.

The combination of the different scales was performed using a logical combination of the outputs. Although this provides a simple, fast approach, a more sophisticated fusion step using a additional CNN might help in reducing the over/under-segmentation errors. This approach, together with parameter sharing training, will be investigated in future steps.

References

1. C. Klaver, R. Wolfs, J. Assink, C. van Duijn, A. Hofman, and P. de Jong, "Genetic risk of age-related maculopathy. population-based familial aggregation study," *Archives of Ophthalmology* **116**, pp. 1646–51, 1998.
2. N. M. Bressler, "Age-related macular degeneration is the leading cause of blindness," *Journal of the American Medical Association* **291**, pp. 1900–1901, 2004.
3. M. J. J. P. van Grinsven, Y. T. E. Lechanteur, J. P. H. van de Ven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Automatic drusen quantification and risk assessment of age-related macular degeneration on color fundus images," *Investigative Ophthalmology and Visual Science* **54**, pp. 3019–3027, 2013.
4. L. de Sisternes, N. Simon, R. Tibshirani, T. Leng, and D. L. Rubin, "Quantitative sd-oct imaging biomarkers as indicators of age-related macular degeneration progression.," *Invest Ophthalmol Vis Sci* **55**, pp. 7093–7103, 2014.
5. N. Jain, S. Farsiu, A. A. Khanifar, S. Bearelly, R. T. Smith, J. A. Izatt, and C. A. Toth, "Quantitative comparison of drusen segmented on sd-oct versus drusen delineated on color fundus photographs," *Investigative Ophthalmology and Visual Science* **51**, pp. 4875–4883, 2010.
6. C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, pp. 1915–1929, 2013.
7. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *arXiv preprint arXiv:14114038* , 2015.
8. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:14091556* , 2014.
9. M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering* **3**, pp. 169–208, 2010.
10. M. K. Garvin, M. D. Abràmoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images.," *IEEE Trans Med Imaging* **28**, pp. 1436–1447, 2009.
11. X. Chen, M. Niemeijer, L. Zhang, K. Lee, M. D. Abramoff, and M. Sonka, "Three-dimensional segmentation of fluid-associated abnormalities in retinal oct: probability constrained graph-search-graph-cut," *IEEE Transactions on Medical Imaging* **31**, pp. 1521–1531, 2012.
12. L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**, pp. 297–302, 1945.