

Deep Learning based Retinal OCT Segmentation

M. Pekala^{*}, N. Joshi^{*}, D.E. Freund^{*}, N. M. Bressler[†],
 D. Cabrera DeBuc[‡], and P. Burlina^{*}

^{*} Johns Hopkins University Applied Physics Laboratory, Laurel MD

[†] Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore MD

[‡] Bascom Palmer Eye Institute,
 University of Miami Miller School of Medicine, Miami FL

Abstract. Objective

To evaluate the efficacy of methods that use deep learning (DL) for the automatic fine-grained segmentation of optical coherence tomography (OCT) images of the retina.

Methods

OCT images from 10 patients with mild non-proliferative diabetic retinopathy were used from a public (U. of Miami) dataset. For each patient, five images were available: one image of the fovea center, two images of the perifovea, and two images of the parafovea. For each image, two expert graders each manually annotated five retinal surfaces (i.e. boundaries between pairs of retinal layers). The first grader's annotations were used as ground truth and the second grader's annotations to compute inter-operator agreement. The proposed automated approach segments images using fully convolutional networks (FCNs) together with Gaussian process (GP)-based regression as a post-processing step to improve the quality of the estimates. Using 10-fold cross validation, the performance of the algorithms is determined by computing the per-pixel unsigned error (distance) between the automated estimates and the ground truth annotations generated by the first manual grader. We compare the proposed method against five state of the art automatic segmentation techniques.

Results

The results show that the proposed methods compare favorably with state of the art techniques, resulting in the smallest mean unsigned error values and associated standard deviations, and performance is comparable with human annotation of retinal layers from OCT when there is only mild retinopathy.

Conclusions

The results suggest that semantic segmentation using FCNs, coupled with regression-based post-processing, can effectively solve the OCT segmentation problem on par with human capabilities with mild retinopathy.

Keywords: Fully convolutional networks, Gaussian process regression, OCT segmentation

1 Introduction

Optical coherence tomography (OCT) is an important retinal imaging modality as it is a non-invasive, high-resolution imaging technique capable of capturing micron-scale structure within the human retina. The retina is organized into layers (e.g. see figure 1 in [1]) and abnormalities in this structure have been associated with ophthalmic, neurodegenerative and vascular disorders. One such example is age-related macular degeneration (AMD), a retinal condition that is among the leading causes of blindness and visual impairment. For individuals over 50 years of age in the United States, if left untreated, it is the leading cause of irreversible central vision loss [2–4]. Studies have shown that advanced AMD lesions correlate with thinning of the outer retina in geographic atrophy as well as overlying choroidal neovascularization [5].

As a part of the central nervous system (CNS), the retina is also subject to a number of specialized immune responses similar to those in the brain and spinal cord; changes in the retinal structure have been associated with CNS disorders such as stroke, multiple sclerosis, Parkinson’s disease, and Alzheimer’s disease. In particular, thinning of the retinal nerve fiber layer (RNFL) is often associated with the aforementioned disorders and, in some cases, its thickness correlates directly with the progression of neurological impairment [6]. Furthermore, ocular manifestations of CNS disorders can sometimes precede symptoms within the brain itself, while thickening of the retina with cystoid abnormalities or subretinal fluid represents one of the most common causes of vision impairment, i.e., retinal pathology from macular edema as a result of diabetes or retinal vein occlusions. Since the retinal structure can be imaged relatively easily via OCT, automated retinal analysis using OCT provides a compelling complement to traditional CNS detection methodologies. Currently, commercial OCT devices provide a map to describe the retinal thickness, typically between the surface of the retina and the retinal pigment epithelial layer of the retina. However, these measurements may not fully incorporate the data available on OCT regarding retinal pathology.

Work in automated retinal image analysis (ARIA) has steadily progressed in the past two decades, as datasets have become more plentiful and machine vision and machine learning techniques have become more proficient (e.g. [7–12]). This has also favorably impacted work in automatic OCT segmentation, where most standard algorithms employ classical (e.g. graph based [13]) segmentation techniques (see e.g. [1, 14–20] and specifically [20] for a recent review of the practice).

Work in deep learning has had substantial impact recently on medical imaging (see examples such as [21, 22]) and also ARIA, for instance to automatically detect patients with referable age related macular degeneration from fundus images [23, 24] or OCT [25]. For OCT segmentation, some recent studies have featured the use of convolutional neural networks (ConvNets): [26] uses ConvNets to delineate macular edema, [27] uses a cascaded U-Net like architecture [28] and shows performance close to that of a classical approach based on random forests, and [29] uses a hybrid ConvNets and graph based method to identify

OCT boundary layers. Recent efforts at U. of Miami [1] have also taken steps to develop publicly available OCT datasets with clinical gold standards for comparing performance among methods, including a number of OCT segmentation algorithms of record.

The salient/novel features of the present work include: a new OCT segmentation method using a combination of fully convolutional networks (FCNs) based on DenseNet and Gaussian process regression, and a performance comparison with methods of record showing that the proposed approach performs on par with a human annotator and compares favorably against other methods of record when used on a publicly available dataset [1]. In particular, our method exhibits the smallest unsigned boundary estimation errors, a result which has potential clinical implications given that ophthalmic, neurological, and vascular disorders have manifestations in retinal layers visible in OCT.

2 Methods

2.1 Data

For our study we utilize the publicly available U. of Miami OCT dataset [1]. This includes 50 OCT images spanning 10 different patients with mild, non-proliferative diabetic retinopathy. Each image consists of 768×496 pixels with transversal and axial resolutions of $11.11\mu\text{m}/\text{pixel}$ and $3.867\mu\text{m}/\text{pixel}$. There are five images available for each patient, which includes one image of the fovea center, two of the perifovea, and two of the parafovea. Two expert graders each annotated five retinal surfaces per image, where a “surface” is defined as the boundary between a pair of adjacent retinal layers. The result is a total of 250 annotated surfaces per grader. The annotated surfaces are numbered 1,2,4,6 and 11 (following the convention introduced in [1]). These surfaces and the associated layers are described in Table 1. Also following the approach in [1], we use the first grader’s annotations as ground truth and the second grader’s annotations as a measure of inter-operator agreement.

Surface ID	Upper Layer	Lower Layer
1	Pre-retinal space	Nerve fiber layer
2	Nerve fiber layer	Ganglion cell layer
4	Inner plexiform layer	Inner nuclear layer
6	Outer plexiform layer	Henle’s Fiber layer and Outer nuclear layer
11	Bruch’s complex	Choriocapillaris

Table 1. Annotated surfaces provided by dataset in [1].

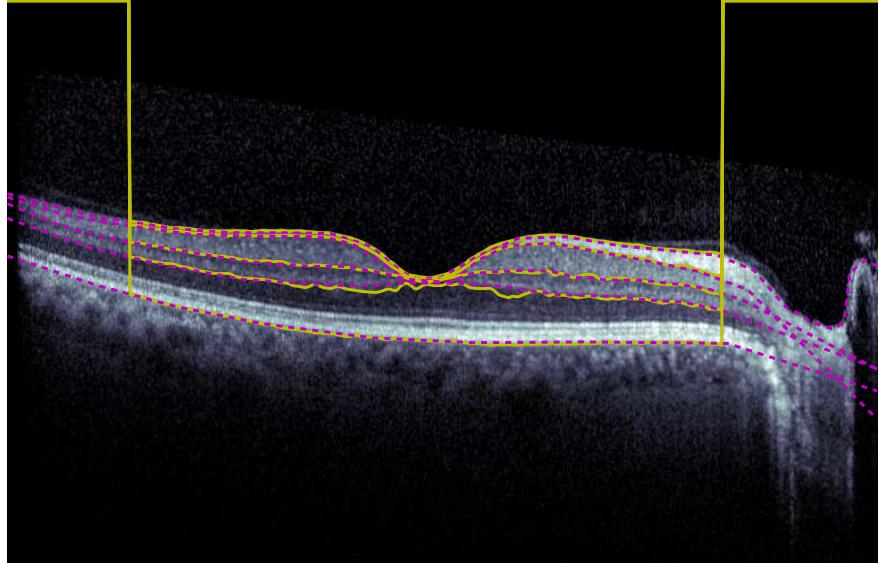


Fig. 1. Example annotations from the dataset of [1]. The yellow lines depict the AURA surface estimate; note the estimate does not span the entire image. Magenta lines denote the estimates generated by human observer #1 which are used as ground truth in this study.

2.2 Segmentation approach

Our approach for estimating retinal surfaces consists of two primary steps. The first step employs a classification algorithm to identify, for each pixel, the most likely corresponding retinal layer. These per-pixel classification estimates are then used as the inputs to the second step, a regression procedure which leverages our prior knowledge that retinal surfaces can be modeled as smooth functions that partition layers along the axial dimension.

For the retinal layer classification step we employ a fully convolutional neural network (FCN) based on the DenseNet [30,31] architecture. FCNs are a subcategory of ConvNets that take tensor-like data as input and produce class estimates having the same spatial dimensions; when the inputs are images, FCNs provide per-pixel class estimates. This is in contrast with more traditional “whole image” classification schemes whereby a single class estimate is produced for the entire input. Note that, while this work only attempts to estimate the layer associated with each pixel, the ability to generate per-pixel class estimates might also be used to identify additional clinically relevant features or lesions in OCT images.

Many convolutional networks process data serially, in that each layer operates solely upon the output of the previous layer. The DenseNet architecture, in contrast, permits each layer of the network to directly process the outputs from all previous layers. This construction allows information (i.e. features) ex-

tracted in the early layers to propagate throughout the network without being perturbed by the action of intermediate layers. Directly passing feature maps from early to later layers also has benefits with regard to efficient training via backpropagation. Other FCNs, such as U-Nets [28], also directly propagate a subset of features maps; however these intra-layer connections are less abundant relative to the DenseNet architecture. In fact, our initial experiments were based on U-Nets; however, we empirically found the DenseNet architecture provided superior performance in this setting.

Variations in thickness of retinal layers introduces a non-trivial amount of class imbalance in the aforementioned classification procedure (there are fewer pixels corresponding to the thin, inner retinal layers). To mitigate the impact of this class imbalance we increase the weight in the loss penalty for the pixels associated with minority classes during training by a factor of 10 (roughly corresponding to the level of class imbalance).

At this point, one might attempt to directly extract surfaces from the layer estimates by identifying locations where class estimates change along the axial dimension. However, surfaces are defined by a unique location for each pixel in the transversal dimension, a constraint not explicitly enforced by the per-pixel classification procedure. For example, Fig. 2 shows an example classification output that, while fairly accurate, includes a few undesirable artifacts that may introduce duplicate or missing surface estimates at some locations. One option is to employ local heuristics to address these issues. In this heuristic, if the classification procedure generates more than one candidate for a layer at a given location, the point which is nearest in Euclidean distance to the prior surface is used (for surface 1, distance to surface 2 is used as the adjudication method). Alternately, if a layer estimate is missing for any given location, an estimate is imputed from the nearest available value for that layer. The combination of methods employed above for segmentation and post-processing constitutes a baseline algorithm which we term “SEG”.

An alternative to making local repairs is to explicitly use our prior knowledge that retinal surfaces (in two-dimensional images) can be modeled as scalar-valued functions with an appropriate level of smoothness and apply a post-processing module that solves a regression problem for each surface. For this study we employ Gaussian processes (GP) with a Radial Basis Function kernel for this purpose [32]. We used a value of 50 pixels for both the variance and length scale hyper-parameters of this kernel; this choice was based on qualitative evaluation of the smoothness of the resulting estimates. In the future, improved performance might be obtained by formal hyper-parameter selection. With enough data, hyper-parameters could also be tuned on a per region and/or per-surface basis. Other regression techniques are of course possible; in addition to providing a clean mechanism for specifying smoothness priors, GPs also have the advantage of providing a mechanism for solving regression problems in higher dimensions (an important consideration in settings where volumetric OCT data is available). We term this combined FCN and GP approach “SEG+REG”.

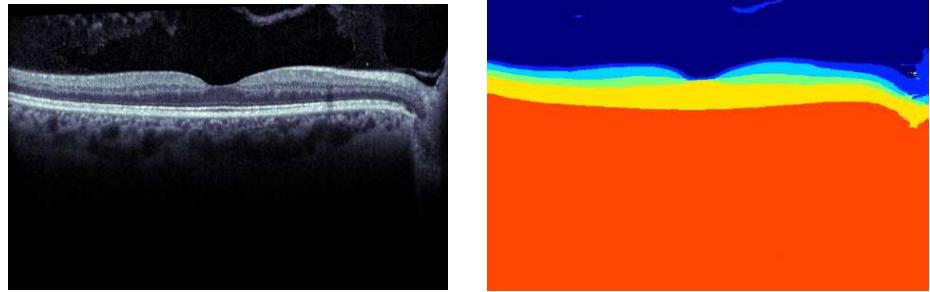


Fig. 2. Example segmentation; original image (left); neural network segmentation output, before post-processing (right).

2.3 Comparison with other state of the art algorithms

In addition to OCT images and ground truth, the publicly available U. of Miami OCT dataset [1] also includes annotations generated by five commonly used OCT segmentation software packages and/or algorithms of record. These reference algorithms/implementations are: Spectralis 6.0 [15], IOWA Reference Algorithm [16], AUTomated retinal analysis tools (AURA) [17], Dufour's (Bern) algorithm [18], and OCTRIMA3D [19]. We refer the reader to [1] for a complete description of these algorithms. Note that these automated annotations do not always span the entire OCT image (e.g., see Fig. 1). Therefore, our performance evaluation is based solely upon the subset of each image for which all algorithms produced a valid surface estimate.

2.4 Evaluation methods and metrics

We use a K-fold cross validation ($K=10$) process where we use nine sets of five images (resulting in a total of 45 images) from nine patients for training the FCN, and testing is done on the remaining test patient's five images. Then the patient used for testing is rotated as is done in conventional K-fold testing approaches, resulting in testing performed on all images. This stratification allowed us to train the network on representative data while ensuring that the segmented images for a given patient were not a by-product of training on that patient's images. A few of the images contain regions that consist of all zero pixels; these regions were not used during training (although they are evaluated at test time).

Following the approach in [1], we measure the accuracy of surface estimates by computing the per-pixel differences between the estimate and the ground truth annotations generated by the first manual grader. Metrics calculations are limited to the regions for which all automated algorithms in the dataset had valid estimates (therefore excluding remote/lateral regions where artifacts are more prevalent). We used mean unsigned errors and mean signed errors as performance metrics for both the proposed algorithms and algorithms of record.

Table 2. Mean unsigned error aggregated across all eye regions. Values in bold indicate when an algorithm meets or exceeds human performance.

	SEG	SEG+REG	Spectralis	OCTRIMA	AURA	IOWA	Bern	Inter-Observer
surface 1	1.13	1.10	1.09	0.95	1.35	2.03	1.71	0.87
surface 2	1.14	1.06	1.45	1.18	1.19	1.74	2.77	1.14
surface 4	0.95	0.92	1.92	0.99	1.12	1.79	1.60	1.10
surface 6	1.23	1.19	1.19	1.52	1.54	1.51	1.72	1.29
surface 11	1.06	1.02	0.99	1.20	0.96	1.22	1.24	1.12
mean	1.10	1.06	1.33	1.17	1.23	1.66	1.81	1.10
max	1.23	1.19	1.92	1.52	1.54	2.03	2.77	1.29
std	0.10	0.10	0.37	0.23	0.22	0.30	0.57	0.15

For a given surface, the estimate v_{est} and the corresponding ground truth v_{ref} are both vectors (with dimension equal to the width of the evaluation region, in pixels) and the signed error is defined to be

$$e_s = v_{ref} - v_{est};$$

the unsigned error is just the absolute value of e_s taken component-wise.

3 Results

We report the performance of both the SEG and SEG+REG compared with other algorithms. Table 2 reports the mean unsigned errors for each algorithm and surface, and the average and max values across all testing data. Values in bold font indicate when an algorithm meets or exceeds human performance (e.g. inter-operator error). The table suggests that in aggregate the proposed methods match human performance, and perform favorably when compared to other algorithms of record. These results also indicate particularly good performance of the proposed methods on the inner retinal surfaces. Table 3 shows the signed errors for the corresponding regions, from which it appears that our method may be slightly overestimating the support of the retinal layers as evidenced by a relatively large positive error on surface 1 and a relatively large negative error on surface 11. Following [1] we also provide the mean unsigned error broken down by ocular regions in Table 4 ¹.

4 Discussion

We present results demonstrating that semantic segmentation using a DenseNet fully convolutional Network coupled with a regression-based post-processing using GPs can effectively address the problem of fine-grained automated OCT

¹ Note there is some minor difference between these results and table 5 of [1] for the algorithms of records which may be attributed to variations in the extent of the macular region that was evaluated; many of the automated methods tend to exhibit greater variation towards the edges of the scans.

Table 3. Mean signed error across all eye regions.

	SEG	SEG+REG	Spectralis	OCTRIMA	AURA	IOWA	Bern	Inter-Observer
surface 1	0.90	0.89	-0.82	0.66	1.22	1.99	1.65	0.26
surface 2	-0.12	-0.14	0.76	0.16	0.34	1.47	2.53	0.29
surface 4	0.18	0.18	1.43	0.12	0.41	1.59	1.30	0.29
surface 6	-0.30	-0.30	-0.51	-0.92	-0.51	0.78	1.13	0.09
surface 11	-0.66	-0.66	-0.44	-0.94	-0.58	1.04	0.90	-0.69

Table 4. Mean unsigned error for all surfaces and regions.

	SEG	SEG+REG	Spectralis	OCTRIMA	AURA	IOWA	Bern	Inter-Observer
surface1 fovea	1.18	1.13	0.90	0.90	0.90	2.14	1.67	0.85
surface1 parafovea	1.12	1.10	1.14	1.00	1.31	1.98	1.81	0.89
surface1 perifovea	1.12	1.09	1.13	0.92	1.62	2.01	1.62	0.86
surface2 fovea	1.34	1.24	1.39	1.15	1.29	2.42	2.02	1.31
surface2 parafovea	1.03	0.97	0.92	1.03	0.92	1.59	2.45	0.97
surface2 perifovea	1.15	1.05	2.02	1.35	1.42	1.54	3.47	1.22
surface4 fovea	1.10	1.08	1.30	1.12	1.25	1.81	1.44	1.13
surface4 parafovea	0.91	0.89	1.32	0.91	1.02	1.67	1.52	1.08
surface4 perifovea	0.92	0.88	2.82	1.00	1.14	1.89	1.76	1.11
surface6 fovea	1.45	1.40	1.79	2.75	2.58	1.58	1.86	1.50
surface6 parafovea	1.26	1.22	1.10	1.36	1.42	1.50	1.74	1.36
surface6 perifovea	1.08	1.04	0.99	1.08	1.14	1.49	1.62	1.11
surface11 fovea	0.92	0.88	0.81	1.02	0.88	1.08	1.23	1.12
surface11 parafovea	1.07	1.04	0.98	1.19	0.95	1.14	1.16	1.12
surface11 perifovea	1.11	1.08	1.07	1.31	1.02	1.38	1.32	1.11
mean	1.12	1.07	1.31	1.21	1.26	1.68	1.78	1.12
max	1.45	1.40	2.82	2.75	2.58	2.42	3.47	1.50
std	0.15	0.14	0.53	0.45	0.43	0.37	0.57	0.18

segmentation, a capability that has many clinical applications. The results show that the proposed methods compare well with state of the art, resulting in the smallest mean unsigned error values and associated standard deviations; overall, performance is comparable with human annotation. We should note however that caution should be exercised when interpreting such strict comparisons since the algorithms of record we compare against were developed and optimized using datasets which may not match exactly the U. of Miami evaluation dataset used here, in aspects such as resolution, noise characteristics, and artifacts.

In addition, the benefit of using the proposed approaches are their relative simplicity. Another advantage of the fully convolutional architectures and regression used here is that these approaches can be naturally expanded in a number of ways, including the direct analysis of 3D volumetric data (e.g. see [33]) and to the problem of identifying additional structures within the OCT scans, such as drusen or other lesions.

As mentioned previously, in our study we also originally used FCNs based on U-Nets and ensembles of U-Nets [28]; however, we found DenseNet provided superior performance for this application.

We anticipate these results could be further improved with additional training data and/or a more exhaustive selection of training hyper-parameters (e.g. weighting of minority class pixels or per-layer tuning of the downstream regression). It is also important to note that the dataset used here only represents the mild spectrum of diabetic retinopathy. A future study with an analysis which includes more advanced conditions would also be of value. Further studies would be indicated with more severe retinal pathology.

Overall, the results show that deep learning and FCNs can provide a competitive approach for OCT automatic segmentation that is fully automated and holds promise for clinical applications.

5 Conclusion

We propose novel OCT automated segmentation methods. Results suggest that semantic segmentation using FCNs, coupled with regression-based post-processing, can effectively produce results that are on par with human capabilities and meet or exceed the prior methods of record considered here.

Acknowledgments This work is supported by the JHU/APL Independent Research and Development Program. We thank Dr. Jun Kong for interesting discussions on OCT.

References

1. Jing Tian, Boglarka Varga, Erika Tatrai, Palya Fanni, Gabor Mark Somfai, William E Smiddy, and Delia Cabrera DeBuc. Performance evaluation of automated segmentation software on optical coherence tomography volume data. *Journal of biophotonics*, 9(5):478–489, 2016.
2. Ronald Klein and Barbara EK Klein. The prevalence of age-related eye diseases and visual impairment in aging: Current estimates. *Investigative ophthalmology & visual science*, 54(14), 2013.
3. AC Bird, NM Bressler, SB Bressler, IH Chisholm, G Coscas, MD Davis, PTVM De Jong, CCW Klaver, BEKlein Klein, R Klein, et al. An international classification and grading system for age-related maculopathy and age-related macular degeneration. *Survey of ophthalmology*, 39(5):367–374, 1995.
4. Neil M Bressler. Age-related macular degeneration is the leading cause of blindness... *JAMA*, 291(15):1900–1901, 2004.
5. Glenn J Jaffe, Daniel F Martin, Cynthia A Toth, Ebenezer Daniel, Maureen G Maguire, Gui-Shuang Ying, Juan E Grunwald, Jiayan Huang, Comparison of Age-related Macular Degeneration Treatments Trials Research Group, et al. Macular morphology and visual acuity in the comparison of age-related macular degeneration treatments trials. *Ophthalmology*, 120(9):1860–1870, 2013.

6. Anat London, Inbal Benhar, and Michal Schwartz. The retina as a window to the brain—from eye research to cns disorders. *Nature Reviews Neurology*, 9(1):44–53, 2013.
7. P Burlina, DE Freund, B Dupas, and N Bressler. Automatic screening of age-related macular degeneration and retinal abnormalities. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 3962–3966. IEEE, 2011.
8. Frank G Holz, Erich C Strauss, Steffen Schmitz-Valckenberg, and Menno van Lookeren Campagne. Geographic atrophy: clinical features and potential therapeutic approaches. *Ophthalmology*, 121(5):1079–1091, 2014.
9. Freerk G Venhuizen, Bram van Ginneken, Freekje van Asten, Mark JJP van Grinsven, Sascha Fauser, Carel B Hoyng, Thomas Theelen, and Clara I Sánchez. Automated staging of age-related macular degeneration using optical coherence tomography/automated staging of AMD in OCT. *Investigative Ophthalmology & Visual Science*, 58(4):2318–2328, 2017.
10. Philippe Burlina, David E Freund, Neil Joshi, Y Wolfson, and Neil M Bressler. Detection of age-related macular degeneration via deep learning. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 184–188. IEEE, 2016.
11. David E Freund, Neil Bressler, and Philippe Burlina. Automated detection of drusen in the macula. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 61–64. IEEE, 2009.
12. Albert K Feeny, Mongkol Tadarati, David E Freund, Neil M Bressler, and Philippe Burlina. Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images. *Computers in biology and medicine*, 65:124–136, 2015.
13. Radford Juang, Elliot R McVeigh, Beatrice Hoffmann, David Yuh, and Philippe Burlina. Automatic segmentation of the left-ventricular cavity and atrium in 3d ultrasound using graph cuts and the radial symmetry transform. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 606–609. IEEE, 2011.
14. Delia Cabrera DeBuc. A review of algorithms for segmentation of retinal image data using optical coherence tomography. In *Image Segmentation*. InTech, 2011.
15. Heidelberg Engineering GmbH. Spectralis HRA+OCT user manual software, 2014.
16. K Lee, MD Abramoff, M Garvin, and M Sonka. The Iowa reference algorithms (retinal image analysis lab, iowa institute for biomedical imaging, IA), 2014.
17. Andrew Lang, Aaron Carass, Matthew Hauser, Elias S Sotirchos, Peter A Calabresi, Howard S Ying, and Jerry L Prince. Retinal layer segmentation of macular OCT images using boundary classification. *Biomedical optics express*, 4(7):1133–1152, 2013.
18. Pascal A Dufour, Lala Ceklic, Hannan Abdillahi, Simon Schroder, Sandro De Dzanet, Ute Wolf-Schnurrbusch, and Jens Kowal. Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints. *IEEE transactions on medical imaging*, 32(3):531–543, 2013.
19. Jing Tian, Boglárka Varga, Gábor Márk Somfai, Wen-Hsiang Lee, William E Smiddy, and Delia Cabrera DeBuc. Real-time automatic segmentation of optical coherence tomography volume data of the macular region. *PloS one*, 10(8):e0133908, 2015.
20. A Breger, M Ehler, H Bogunovic, SM Waldstein, AM Philip, U Schmidt-Erfurth, and BS Gerendas. Supervised learning and dimension reduction techniques for quantification of retinal fluid in optical coherence tomography images. *Eye*, 2017.

21. Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.
22. Philippe Burlina, Seth Billings, Neil Joshi, and Jemima Albayda. Automated diagnosis of myositis from muscle ultrasound: Exploring the use of machine learning and deep learning methods. *PLoS one*, 12(8):e0184059, 2017.
23. Philippe Burlina, Neil Joshi, Michael Pekala, Katia Pacheco, David E Freund, and Neil M Bressler. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmology*, 2017.
24. Philippe Burlina, Katia D Pacheco, Neil Joshi, David E Freund, and Neil M Bressler. Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis. *Computers in Biology and Medicine*, 82:80–86, 2017.
25. Cecilia S Lee, Doug M Baughman, and Aaron Y Lee. Deep learning is effective for classifying normal versus age-related macular degeneration oct images. *Ophthalmology Retina*, 1(4):322–327, 2017.
26. Cecilia S Lee, Ariel J Tyring, Nicolaas P Deruyter, Yue Wu, Ariel Rokem, and Aaron Y Lee. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *bioRxiv*, page 135640, 2017.
27. Yufan He, Aaron Carass, Yeyi Yun, Can Zhao, Bruno M Jedynak, Sharon D Solomon, Shiv Saidha, Peter A Calabresi, and Jerry L Prince. Towards topological correct segmentation of macular oct from cascaded fcns. In *Fetal, Infant and Ophthalmic Medical Image Analysis*, pages 202–209. Springer, 2017.
28. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
29. Leyuan Fang, David Cunefare, Chong Wang, Robyn H Guymer, Shutao Li, and Sina Farsiu. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative amd patients using deep learning and graph search. *Biomedical Optics Express*, 8(5):2732–2744, 2017.
30. Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1175–1183. IEEE, 2017.
31. Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016.
32. Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, 2006.
33. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d U-Net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.