

U-Net-Based Segmentation of Current Imaging Biomarkers in OCT-Scans of Patients with Age Related Macular Degeneration

Kemal YILDIRIM^{a,1}, Sami AL-NAWAISEH^b, Sophia EHLERS^b, Lukas SCHIEßER^b, Michael STORCK^a, Tobias BRIX^a, Nicole ETER^b and Julian VARGHESE^a

^a *Institute of Medical Informatics, University of Münster, Germany*

^b *Department of Ophthalmology, University Hospital Münster, Münster, Germany*

Abstract. Age-related macular degeneration (AMD) is the leading cause of blindness in the Western world. In this work, the non-invasive imaging technique spectral domain optical coherence tomography (SD-OCT) is used to acquire retinal images, which are then analyzed using deep learning techniques. The authors trained a convolutional neural network (CNN) using 1300 SD-OCT scans annotated by trained experts for the presence of different biomarkers associated with AMD. The CNN was able to accurately segment these biomarkers and the performance was further enhanced through transfer learning with weights from a separate classifier, trained on a large external public OCT dataset to distinguish between different types of AMD. Our model is able to accurately detect and segment AMD biomarkers in OCT scans, which suggests that it could be useful for prioritizing patients and reducing ophthalmologists' workloads.

Keywords. Age-Related Macular Degeneration, Spectral Domain Optical Coherence Tomography, Convolutional neural Networks, Image Segmentation

1. Introduction

In western countries, age-related macular degeneration (AMD), is the primary cause of blindness in people over 65 affecting around 1 in 4 adults over the age of 75. A projected 196 million people will experience AMD by the year 2020, according to estimates. By 2040, this figure is anticipated to reach 288 million [1, 2].

International retinal specialists used fundus imaging to classify AMD based on clinical features to determine the likelihood that a late AMD will develop [3]. Depending on drusen and pigmentary changes within two disk diameters of the fovea, eyes were classified as normal age-related changes, early, moderate and late AMD. As new therapy modalities emerge aiming to prevent the development of a late AMD, it crucial to precisely detect and follow up the changes in drusen and pigmentary changes, to predict the natural disease progression and thereby assess the therapeutic benefit.

The quantitative morphological evaluation of drusen and pigmentary changes is made possible by several established and emerging imaging technologies, including

¹ Corresponding Author: Kemal Yildirim, E-mail: kemal.yildirim@uni-muenster.de.

color fundus photography (CFP), fundus autofluorescence (FAF), infrared imaging (IR) and spectral domain optical coherence tomography (SD-OCT). SD-OCT imaging offers the most precise and prompt diagnosis. Compared to FAF and IR, which may show higher variability since the image intensities substantially fluctuate due to variations in illumination and corneal curvature, SD-OCT scans offer consistent anatomic landmarks for objective assessments.

OCT is a non-invasive imaging method that uses low-coherence light to produce cross-sectional images of the macula or optic nerve head [4]. OCT is a highly favored method by ophthalmologists for the evaluation of retinal diseases, such as AMD, due to its non-invasiveness and simplicity of image acquisition [5]. However, it takes a lot of time and effort for ophthalmologists to precisely examine several OCT cross-sections for each patient. Additionally, the chronic nature of AMD adds to the load on ophthalmologists and medical facilities. Therefore, the availability of a computer-aided diagnosis (CAD)-based screening tool that is automated could aid in prioritizing patients based on their conditions and lessen this load.

2. Methods

The open-source Computer Vision Annotation Tool (CVAT) [6] was used by a trained domain expert to label four types of biomarkers on 1200 OCT scans: drusen with 4181 labels, pseudodrusen with 108 labels, choroidal neovascularizations (CNV) with 2810 labels, and pigment epithelial detachments (PED) with 86 labels. To ensure proper capture of the structures of interest, the annotations were reviewed and the polygons were refined by two additional clinicians, if necessary. This was especially helpful in wet AMD with severe geographic atrophy, when the retinal layers were no longer easy to distinguish from each other. Nevertheless, to obtain a sample of clinical cases that is as representative as possible, cases with significant fibrotic or atrophic lesions were also included. Additionally, we have used a range of data augmentation techniques to address the extreme label imbalance and improve its ability to generalize to new data. Following transformations were applied to a copy of the original dataset, prior to the training process, ensuring that the input for all training iterations remained consistent: Gaussian noise, contrast adjustments, elastic deformations, grid/optical distortions and random affine transformations [7].

Our segmentation model is built on CNN with U-net and was trained on the data mentioned above. Different studies have demonstrated that U-net with an additional attention module and ResNet34 as encoder is an effective combination to handle datasets with imbalanced classes and minimize loss [8–10]. For comparison purposes, the encoder was initialized with ImageNet and custom weights from a public OCT-Dataset, which contains 108309 OCT scans [11]. This usually reduces the amount of data necessary for training, allowing models to learn task-specific features while utilizing the knowledge gained from broader datasets.

To calculate the weights of the custom dataset, a classifier based on the same ResNet34 encoder of our segmentation model was trained to distinguish between the given classes of diabetic macular edema (DME), drusen (dry AMD), CNV (wet AMD) and normal (healthy). Our ResNet34-Classifer was pretrained on ImageNet weights. A weighted cross-entropy loss with inverse class frequencies was used to adjust the relative contribution of each class to the loss computation. The weights of the classifier's encoder were then extracted and transferred to our segmentation model's encoder.

In addition, our segmentation model used an alpha-balanced compound loss with Dice and weighted Focal loss to train the network to successfully segment small pathogenic structures within the OCT scans [12]. Specific combinations of loss functions have shown to improve focus on relevant regions and make more precise predictions.

To measure the performance of our model, the Dice/F1 score was calculated for each class and epoch, as the structures for CAD must be captured not only completely but also precisely. This metric measures model accuracy by combining its precision and recall, indicating overlap between predicted segmentation and ground truth mask.

3. Results

Classifier (OCT). The classifier is built on top of ResNet34 and was initialized using ImageNet weights. It was trained over 15 epochs and by using 16 training examples per mini-batch from the public dataset. Initial learning rate was $1e-4$. At epoch 7, the weighted cross-entropy loss was at its minimum. With an 80:20 split for each class, the training accuracy reached 98.207% and the validation accuracy peaked at 97.697%. Rounded accuracies for each class were as follows: NORMAL 98.49%, DME 97.9%, DRUSEN 96.54%, CNV 95.23%.

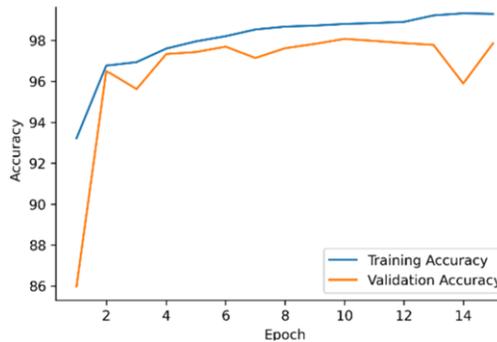


Figure 1. Training and validation accuracy per epoch

Segmentation Model. We have tested our segmentation model with three different weight initialization strategies: random initialization, pre-trained ImageNet weights and custom weights from our OCT classifier. Table 1 compares the results from the epochs with the highest F1 scores. Figure 2 exemplifies results on a separate test dataset.

We used a mini-batch size of 4, an initial learning rate of $1e-3$, and 40 epochs. The same preprocessing and parameters were used throughout all three trials. The compound loss was defined and balanced as described in Eq. (1).

$$Loss_{Compound} = \alpha \cdot Loss_{Focal} + (1 - \alpha) \cdot Loss_{Dice}, \quad \alpha = 0.6, \quad \gamma_{Focal} = 2 \quad (1)$$

Table 1. The F1 scores of the best epoch in each case for all three executions

	Epoch	Drusen	CNV	PED
Random Initialization	17/40	0.777	0.752	0.899
ImageNet	13/40	0.763	0.832	0.888
OCT	21/40	0.807	0.844	0.878

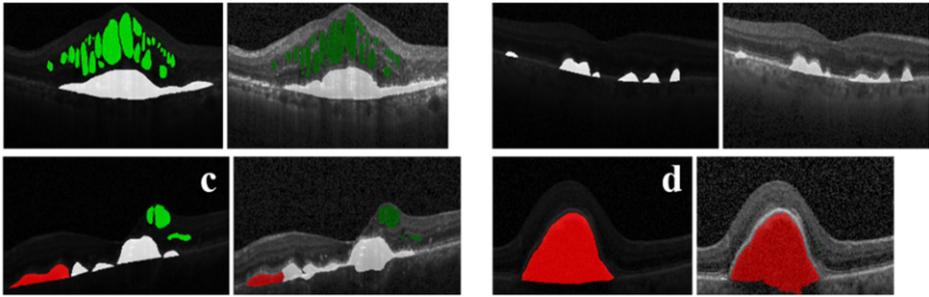


Figure 2. Ground truth and prediction with pretrained encoder. Drusen (white), CNV (green) and PED (red).

4. Discussion

This study's main objective is to analyze the feasibility of CAD and progress monitoring for patients with AMD. We therefore utilize a broader set of features than previous OCT segmentation research, which mostly focuses on single biomarkers such as fluid, layers, or drusen [13]. In addition, we investigate the impact of using weights from different datasets on the segmentation of relevant biomarkers collectively.

Given several AI-based systems achieving comparable image discrimination rates as that of retinal specialists [14], we primarily focus on improving the segmentation. The underlying hypothesis was that it may be improved by using transfer learning with weights from similar or unrelated datasets. Thus, three image segmentation models were trained and compared for the classes drusen, CNV, PED and pseudodrusen. Pseudodrusen had to be excluded due to the insufficient number of labels and the resulting bias in the loss computation.

As seen in the Table 1, the model trained using OCT-weights had the highest F1 score of 0.807 for drusen class. This suggests that training on the custom dataset may have improved the model's class segmentation. The comparably low value of drusenoid structures is likely since the pigments within these deposits are similar to those found in the choroid beneath. This makes it difficult to accurately label and segment lower parts of drusen, as it is based on the assumption of how the retinal pigment epithelium was prior to the formation of a druse. This becomes especially challenging in cases of wet AMD with large and overlapping deposits.

The model trained with our custom weights had the highest F1 score for CNV, 0.844. This suggests that our encoder provided adequate training for this class. ImageNet alone resulted in the highest F1 score for PED, 0.888. OCT-weighted model placed second with 0.878. Transfer learning may have improved performance, but our encoder was unable to produce any additional improvement. The higher F1 score seen in CNV and PED could be related to the hypopigmentation of these structures, which allows a better differentiation from the surrounding tissue. The background class was disregarded from Table 1 due to many easy positives resulting in a F1 score near to 1.

The precision of segmentation results is often influenced by the characteristics of the datasets used. We have included a substantial number of wet AMD cases (600 out of 1300), which may have a diminishing but important impact on the research results.

It is also essential to note that no systematic hyperparameter tuning was performed. Instead, previously established parameters from similar cases were used for our classifier

and segmentation model, including those for loss computation. Our initial tests have demonstrated that the compound and balance of loss functions used was sufficient but with obvious room for enhancement.

5. Conclusion and Outlook

Using image segmentation models, this study evaluated the viability of CAD for patients with AMD. The results demonstrated that transfer learning with weights from similar or unrelated datasets can improve the performance of models for specific classes. Since our classifier has been trained on a publicly available dataset, it can therefore be tested on various diagnoses and biomarkers in future research. In our case, additional data labelling for existing and new AMD biomarkers, preprocessing, and augmentation, as well as hyperparameter tuning, may improve the results. We intend to create interpolated 3D volumetric scans of the OCT images to extract additional properties such as volume, diameter and other criteria that are necessary for clinical classification of AMD [3].

References

- [1] Friedman DS, O'Colmain BJ, Muñoz B, Tomany SC, McCarty C, DeJong, Paulus T. V. M., et al. Prevalence of Age-Related Macular Degeneration in the United States. *Archives of Ophthalmology*. 2004;122:564–72. doi:10.1001/archophth.122.4.564.
- [2] Laurence S Lim, Paul Mitchell, Johanna M Seddon, Frank G Holz, Tien Y Wong. Age-related macular degeneration. *The Lancet*. 2012;379:1728–38. doi:10.1016/S0140-6736(12)60282-7.
- [3] Ferris FL3, Wilkinson CP, Bird A, Chakravarthy U, Chew E, Csaky K, Sarda SR. Clinical classification of age-related macular degeneration. *Ophthalmology*. 2013;120:844–51. doi:10.1016/j.ophtha.2012.10.036.
- [4] van Velthoven, Mirjam E J, Faber DJ, Verbraak FD, van Leeuwen TG, Smet MD de. Recent developments in optical coherence tomography for imaging the retina. *Prog Retin Eye Res*. 2007;26:57–77. doi:10.1016/j.preteyeres.2006.10.002.
- [5] Vineeta Das, Samarendra Dandapat, Prabin Kumar Bora. Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images. *Biomedical Signal Processing and Control*. 2019;54:101605. doi:10.1016/j.bspc.2019.101605.
- [6] CVAT.ai Corporation. Computer Vision Annotation Tool (CVAT); 2022.
- [7] Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: Fast and Flexible Image Augmentations. *Information* 2020. doi:10.3390/info11020125.
- [8] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015: arXiv. doi:10.48550/ARXIV.1505.04597.
- [9] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition 2015: arXiv. doi:10.48550/ARXIV.1512.03385.
- [10] Roy AG, Navab N, Wachinger C. Recalibrating Fully Convolutional Networks with Spatial and Channel 'Squeeze & Excitation' Blocks 2018: arXiv. doi:10.48550/ARXIV.1808.08127.
- [11] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*. 2018;172:1122-1131.e9. doi:10.1016/j.cell.2018.02.010.
- [12] Yeung M, Sala E, Schönlieb C-B, Rundo L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation 2021: arXiv. doi:10.48550/ARXIV.2102.04525.
- [13] Liefers B, Taylor P, Alsaedi A, Bailey C, Balaskas K, Dhingra N, et al. Quantification of key retinal features in early and late age-related macular degeneration using deep learning. *American Journal of Ophthalmology*. 2021;226:1–12.
- [14] Sheng B, Chen X, Li T, Ma T, Yang Y, Bi L, Zhang X. An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. *Frontiers in Public Health*. 2022;10.