Simultaneous Classification and Segmentation of Cysts in Retinal OCT

Stefanos Apostolopoulos1*, Carlos Ciller1, Raphael Sznitman2, and Sandro De Zanet1*

1 RetinAI Medical GmbH, Bern, Switzerland
2 University of Bern, Switzerland

Abstract. The automatic segmentation of fluid deposits in OCT imaging enables clinically relevant quantification and monitoring of eye disorders over time. Eyes with late-stage diseases are particularly challenging to segment, as their shape is often highly warped. In this context, we propose a novel fully-Convolutional Neural Network (CNN) architecture which combines dilated residual blocks in an asymmetric U-shape configuration, and can simultaneously segment and classify cysts in pathological eyes. In here, we validate our approach on the Retouch Challenge with the Medical Image Computing and Computer-Assisted Intervention (MICCAI) '17 Conference dataset.

1 Introduction

Optical Coherence Tomography (OCT) is a non-invasive medical imaging modality that provides micrometer-resolution volumetric scans of biological tissue [1]. Since its introduction in 1991, OCT has seen widespread use in the field of ophthalmology, as it enables direct, non-invasive imaging of the retinal layers. As shown in Fig. 1, OCT enables the visualization of both healthy tissue and pathological biomarkers such as fluid pockets and hyper-reflective spots within and underneath the retinal layers. Critically, these have been linked to diseases such as Age-related Macular Degeneration (AMD), Diabetic Retinopathy (DR) and Retinal Vein Occlusion (RVO) [2,3].

Given the widespread occurrence of these diseases, which is estimated at over 300 million people worldwide, medical image analysis methods for OCT imaging have gained popularity in recent years. The automatic segmentation of fluid pockets is of particular interest as it allows for the quantification, characterization, early detection and monitoring of retinal disorders over time. This remains a challenging task, as fluid pockets can appear in a wide variety of sizes, shapes and locations within the retina. To this, we focus on providing more accurate fluid segmentation and classification in the retina.

A number of methods for fluid segmentation has been proposed in the literature. With a mask highlighting the retina and removing the choroid and vitreous humor, Oguz et al. constrain the fluids to the relevant area [4]. In a different work, Cysts are segmented with a “sheetness” measure adaption of the Frangi filter [5]. Using CNNs, Montuoro et al. present a fully automated 3D method which is able to segment fluid-filled regions and retinal cell layer structures simultaneously [6]. Wu et al., on the contrary, leverage an en-face fundus image prior to differentiate between different kinds of fluid types using

* These two authors contributed equally
fuzzy level set method [7]. Overall, most of these methods face difficulties in segmenting all fluids accurately, not only due to the inhomogeneity in disease appearances, but also due to the variability in imaging quality from different OCT devices.

To this end, we present a novel strategy to overcome the above limitations and provide accurate results in a wider variety of cases. Inspired by recent CNN approaches for semantic segmentation [8] and image classification [9], we use a variant of the CNN architecture used in layer segmentation with a Branch Residual U-Net (BRUNet), that learns to segment fluid pockets as a supervised classification problem [10]. This network combines residual building blocks with dilated convolutions into an asymmetric U-shape configuration, and can segment all three fluid types in highly pathological eyes at Bscan level. We also propose a Cscan (volume) level classifier, which detects the presence or absence of fluids and connects it with its associated fluid type. We construct this network based on the RetiNet [11], a CNN intended for the classification of AMD vs. healthy OCT volumes.

We demonstrate the performance of this method through an extensive evaluation on 4752 manually segmented Bscans from 70 volume scans suffering from fluid accumulation due to various diseases, such as AMD and DR. The dataset is used in the context of the MICCAI ’17 Retouch challenge. We evaluate the segmentation accuracy against a test set both on BScan and CScan level. Finally, we provide an overview of qualitative overview of the segmentation results of each method.

2 Methods

Our goal is to segment retinal fluid pockets in OCT images. The main challenge in this task stems primarily from the highly variable and irregular shape of pathological eyes, and secondarily from the variable image quality (i.e., signal strength and speckle noise)
of clinical OCT scans, depending on the patient’s eye condition and the device used to capture the volume. In here, we opt to segment the fluid pockets in 2D to compensate for the requirements of a global alignment for all cross-sections, or Bscan. This approach simplifies the segmentation problem and avoids the need for computationally intensive 3-dimensional convolutions [12] and volumetric pre-processing (i.e., registration and alignment). In a second step, we aim to provide the information of fluid pocket types present in a full CScan, i.e. volumetric data.

2.1 Preprocessing

We performed a number of preprocessing steps geared towards reducing the complexity of the problem. The three OCT devices acquire scans at different depth resolutions, which changes the anatomical shape representations from one device to the next. By resampling all scans to have an equal resolution, we normalize the appearance of the retinal structures, as well as that of the pathologies.

Since the type of cyst is highly correlated with retinal layer positions, we pre-segment the Bscan to reveal seven different retinal cell layers: Nerve Fibre Layer (NFL), Ganglion Cell Layer (GCL), Inner Plexiform Layer (IPL), Outer Plexiform Layer (OPL), Outer Nuclear Layer (ONL), Inner Nuclear Layer (INL), Retinal Pigment Epithelium (RPE) and Internal Limiting Membrane (ILM). To achieve this, we use the network configuration presented as BRUNet [10].

To ensure proper generalization, the dataset was augmented by altering Bscans with pathology and anatomy preserving transformations, such as horizontal flip, shear, rotation, shift and Gaussian noise. These transformations widely amplify the variability of the scans leading to a more versatile network.

![Fig. 2: Example of the inputs (first and second image from the left), the original Bscan and its corresponding layer segmentation, and the expected output (three channels combined into one image).](image)

2.2 Fluid segmentation

In our approach, we model the task of segmenting fluid pockets as a classification problem. Given a Bscan image, $I$, we wish to find a function $T : I \rightarrow F$, that maps each
pixel in \( \mathcal{I} \) to a label \( \mathcal{F} \in \{0, 1, 2, 3\} \) corresponding to the background and a fluid pocket type. We consider the following three types of cysts as per the Retouch challenge: (0) Background, (1) IRF, (2) SRF, (3) PED.

We leverage the BRUNet [10] architecture and its layer segmentation to build upon a successful segmentation network in OCT, and use the outcome of the network as the prior for the retinal layers. Each of the three output channels corresponds to the segmentation of one of the three fluid types. We then train the network to minimize the binary cross-entropy between the predicted segmentation and the ground truth. To conclude, the network parameters are updated via back-propagation and the Adam optimization process with the infinity norm [13].

![Diagram of the proposed joint segmentation and classification network.](image)

**Fig. 3:** Layout for the proposed joint segmentation and classification network. The original image and a layer-segmentation is fed to the segmentation network as well as to the classifier. Additionally, the output of the segmentation network is added to the input of the classifier. The segmentation and classification outputs are combined to give the final output. All outputs are marked with red outlines.

In our evaluations of the segmentation we have noticed a disproportionate number of false positives. To this end, the described fluid segmenter is coupled with a classifier (in our case is a RetiNet B [11]). The latter consists of a CNN, trained in an extreme learning fashion, which has shown good results for classification of Bscans in the past. We jointly learn the classification and refine the segmentation by back-propagating through both. An overview of the presented network can be observed in Fig. 3. The networks has two outputs: Bscan fluid type presence classification \( C = [c_0, c_1, c_2], c_i \in [0, 1] \) and pixel-wise fluid segmentation \( S = [s_0, s_1, s_2], s_i \in [0, 1]^2 \). Both outputs are trained on...
categorical cross-entropy. The output \( s \) is used as a gate to remove false positives to create a more accurate result \( s' \), by multiplying the outputs: \( s'_i = c_i \cdot s_i \).

2.3 Volumetric classification of retinal cysts

While fluid segmentation and classification is performed per Bscan, the diagnosis of a patient has to be performed on the full volume (Cscan). Given a Cscan volume, \( V \), we wish to find a function \( T : V \rightarrow T \), that maps the full volume \( V \) to a triplet of labels \( F \in \{1, 2, 3\} \) corresponding to the presence of the particular type of fluid: (1) IRF, (2) SRF, (3) PED. With classical deep learning this involves a classifier with three outputs. Again, we use the RetiNet network, specifically RetiNet C, which stacks all Bscans into a single image and reuses weights from RetiNet B. Due to the different number of Bscans per Cscan, missing Bscans are linearly interpolated to add up to 128 Bscans per Cscan.

3 Experimental Results

Dataset: The training set provided by the RETOUCH challenge consists of Cscans from four different devices:

- Cirrus (Zeiss Meditec) with 128 Bscans.
- Spectralis (Heidelberg Engineering) with 49 Bscans.
- T-1000 and T-2000 (Topcon) with 64 and 128 Bscans, respectively.

Each device exhibits different resolution and signal to noise ratio. For each Bscan in each volume, manually segmented ground truth fluid pockets were provided by ophthalmologists from two different reading centers. First we evaluate the ROC on the Cscan level.

We compare plain BRUNet classification with DenseNet [14] and Retinet B as an average over all Bscans and, as can be seen in Fig. 4. Retinet B achieves an Area Under the Curve (AUC) of 0.95, and outperforms direct classification with BRUNet, as well as DenseNet. Second, we compute pixelwise segmentation accuracy using the Dice Coefficient and the Absolute Volume Difference (AVD) metrics. Plain BRUNet achieves a Dice score of 0.51±0.36, 0.58±0.44 and 0.73±0.38 for IRF, SRF and PED, whereas the gated version
of BRUNet achieves 0.59±0.35, 0.55±0.45 and 0.77±0.35 respectively. The corresponding AVD values are 1.38e+05±2.29e+05, 1.33e+05±3.32e+05, 4.97e+04±1.17e+05 for the plain variant and 1.26e+05±2.25e+05, 6.97e+04±1.71e+05, 4.99e+04±1.23e+05 for the gated version. A qualitative comparison can be seen in Fig. 6.

Execution time for segmentation of Bscans amounts to 3-10 seconds (depending on the number and resolution of Bscans in a volume), with a AMD Ryzen 1700 CPU in combination with an NVidia 1080 Ti GPU.

Fig. 5: Segmentation results per cyst type and device: Dice Coefficient (left) and absolute volume difference (right). Top row displays segmentation results without gating; bottom row with. Color indicates device: Cirrus (blue), Spectralis (red) and T-1000/T-2000 (green).

4 Conclusion

We have presented a method for performing cyst segmentation and classification on OCT scans of highly pathological retinas. Inspired by recent advances in computer vision, we have designed a novel fully-convolutional CNN architecture that can segment and classify Bscans simultaneously. These results appear promising and warrant further development.
Fig. 6: Two examples of segmentations. From left to right: original Bscan, ground truth segmentation and ours.

References

3. Morgan, J.I.W.: The fundus photo has met its match: Optical coherence tomography and adaptive optics ophthalmoscopy are here to stay (2016)