RETOUCH - The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge

Hrvoje Bogunović, Freerk Venhuizen, Sophie Klimscha, Stefanos Apostolopoulos, Alireza Bab-Hadiashar, Ulas Bagci, Mirza Faisal Beg, Loza Bekalo, Qiang Chen, Carlos Ciller, Karthik Gopinath, Amirali K. Gostar, Kiwan Jeon, Zexuan Ji, Sung Ho Kang, Dara D. Koozekanani, Donghuan Lu, Dustin Morley, Keshab K. Parhi, Hyoung Suk Park, Abdolreza Rashno, Marinko Sarunic, Saad Shaikh, Jayanthi Sivaswamy, Ruwan Tennakoon, Shivin Yadav, Sandro De Zanet, Sebastian M. Waldstein, Bianca S. Gerendas, Caroline Klaver, Clara I. Sánchez, Ursula Schmidt-Erfurth

***Abstract*—Retinal swelling due to the accumulation of fluid is associated with the most vision-threatening retinal diseases. Optical coherence tomography (OCT) is the current standard of care in assessing the presence and quantity of retinal fluid and image-guided treatment management. Deep learning methods have made their impact across medical imaging and many retinal OCT analysis methods have been proposed. But it is currently not clear how successful they are in interpreting retinal fluid on OCT, which is due to the lack of standardized benchmarks. To address this, we organized a challenge RETOUCH in conjuction with MICCAI 2017, with eight teams participating. The challenge consisted of two tasks: fluid detection and fluid segmentation. It featured for the first time: all three retinal fluid types, with annotated images provided by two clinical centers, which were acquired with the three most common OCT device vendors from patients with two different retinal diseases. The analysis revealed that in the detection task, the performance on the automated fluid detection was within inter-grader variability. However, in the segmentation task, fusing the automated methods produced segmentations that were superior to all individual methods, indicating the need for further improvements in segmentation performance.**

***Index Terms*—Evaluation, Image segmentation, Image classification, Optical Coherence Tomography, Retina.**

I. Introduction

Macular edema is a swelling of the central retina caused by the leakage from the retinal capillaries and subsequent accumulation of the leaked fluid within the intercellular spaces of the retina. It causes sudden and severe loss of vision and it occurs secondary to a retinal disease such as age-related macular degeneration (AMD), retinal vein occlusion (RVO) or diabetic macular edema (DME). These three conditions constitute the most common cause of vision loss in developed countries, affecting a large number of people. In particular, AMD is the leading cause of blindness in developed countries affecting older patients [1], while RVO and DME are major causes of vision impairment in working age people.

An effective treatment for macular edema exists in the form of anti-vascular endothelial growth factor (anti-VEGF) therapy [2]. However the effectiveness of the treatment depends on frequent monitoring and an early detection of the disease. Furthermore, anti-VEGF drugs are expensive and have to be administered frequently for an extended period of time. Consequently, they pose a heavy socio-economic burden on both the patient and the healthcare system. Thus, several personalized treatment regimens have been developed, such as the *pro re nata* (PRN, “as needed") and *treat and extend* (T&E). For these regimens, injection decisions are guided by the re-occurrence of retinal bleeding or fluid accumulation, resulting in a lower number of injections while keeping the visual benefits comparable to those achieved with more frequent, monthly injections [3], [4].

Accumulated fluid causing macular edema can be readily imaged using optical coherence tomography (OCT) [5], [6].

Manuscript received November 13, 2018; revised February 13, 2019; accepted February 17, 2019. Date of publication XXXX XX, 2019. *(Corresponding author: Hrvoje Bogunović)*

H. Bogunović, F. Venhuizen, S. Klimscha, S.M. Waldstein, B.S. Gerendas, C. Klaver, C. I. Sánchez, and U. Schmidt-Erfurth organized the challenge. Participants are listed alphabetically and they provided short descriptions of their methods in the appendix. (e-mail: hrvoje.bogunovic@meduniwien.ac.at)

H. Bogunović, S. Klimscha, S. M. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth are with Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology, Medical University of Vienna, Vienna, Austria. They acknowledge the financial support by the Christian Doppler Research Association, the Austrian Federal Ministry for Digital and Economic Affairs, and the National Foundation for Research, Technology and Development.

F. Venhuizen and C. I. Sánchez are with Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, The Netherlands. They acknowledge the foundations: MD fonds, LSBS fonds and OOG fonds that contributed through UitZicht.

C. Klaver is with Department of Ophthalmology, Erasmus University Medical Center, Rotterdam, The Netherlands.

S. Yadav, K. Gopinath and J. Sivaswamy are with IIIT Hyderabad, Hyderabad, India.

S. H. Kang, H. S. Park and K. Jeon are with National Institute for Mathematical Sciences, Daejeon, Korea. They acknowledge National Institute for Mathematical Sciences (NIMS) grant funded by the Korean government (No. B18130000).

Q. Chen, Z. Ji and L. Bekalo are with Nanjing University of Science & Technology, Nanjing, China.

S. Apostolopoulos, C. Ciller and S. De Zanet are with RetinAI Medical GmbH, Bern, Switzerland.

R. Tennakoon, A. K. Gostar and A. Bab-Hadiashar are with School of Engineering, RMIT University, Melbourne, Australia.

D. Lu, M. Sarunic and M. F. Beg are with School of Engineering Science, Simon Fraser University, Burnaby, Canada.

D. Morley, S. Shaikh and U. Bagci are with University of Central Florida, Orlando, US

A. Rashno is with Department of Computer Engineering, Engineering Faculty, Lorestan University, Khorramabad, Iran.

K. K. Parhi is with Department of Electrical and Computer Engineering, and D. D. Koozekanani is with Department of Ophthalmology and Visual Neuroscience, University of Minnesota, Minneapolis, US.

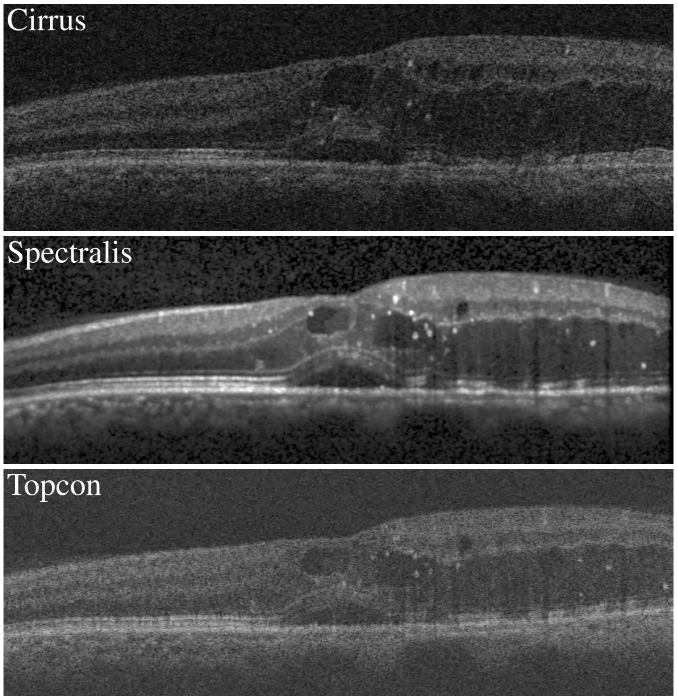


Fig. 1. Retina with macular edema imaged with OCT scanners: Cirrus, Spectralis, and Topcon, from three different vendors. The slices (B-scans) are of the same patient and approximately at the same anatomical position.

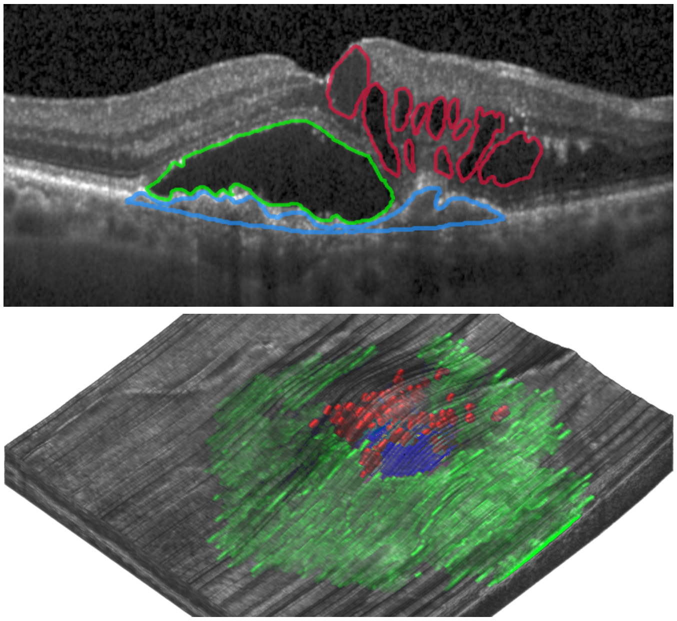


Fig. 2. The three fluid types on a 2D B-scan (above) and as a 3D volume rendering (below): IRF (red), SRF (green) and PED (blue).

Even though it is a recently introduced modality, OCT has already become a standard of care impacting the treatment of millions of people every year [7]. It obtains a high resolution 3D image of the retina in a fast and noninvasive manner by acquiring a series of cross-sectional slices (B-scans) [8], [9]. However, OCT volumes are prone to eye motion artifacts and to low signal-to-noise ratio (SNR) due to speckle. To address this, device vendors have to decide on a tradeoff between achieved SNR, image resolution and the scanning time. Thus, the acquired image quality and scan density can vary widely between different OCT device vendors as demonstrated in Fig. 1. It can be observed that the highest image quality was produced by Spectralis scanner, which reduces the noise by averaging multiple B-scans of the same anatomical location at the expense of acquiring fewer B-scans.

Three types of fluid (Fig. 2) are clinically distinguishable on OCT images and are considered relevant imaging biomarkers for visual acuity and retreatment indication.

a) Intraretinal Fluid (IRF): It consists of contiguous fluid-filled spaces containing columns of tissue as the arrangement of such spaces is determined by the Müller fibres, which are vertical. These spaces when viewed on OCT may appear as separated hyporeflective cystoid pockets hence sometimes such fluid is also referred to as *cystoid fluid*.

b) Subretinal Fluid (SRF): It corresponds to the accumulation of a clear or lipid-rich exudate in the subretinal space, i.e., between the neurosensory retina and the underlying retinal pigment epithelium (RPE) that nourishes photoreceptors.

c) Pigment Epithelial Detachment (PED): It represents detachment of the RPE along with the overlying retina from the remaining Bruch's membrane (BM) due to the accumulation of fluid or material in sub-RPE space. It is specific to AMD as it is associated with the choroidal neovascularization which originates beneath RPE. PED is composed of three clinical subtypes: *serous*, *fibrovascular*, and *drusenoid*.

Clinical studies have found these fluids to have important prognostic values: IRF represents one of the most important variables associated with vision loss, SRF is associated with a possibly favorable visual prognosis in AMD [10]–[12], and PED is considered the primary indicator for progressive disease activity [13]. In addition, quantifying the extent of SRF, the change of PED volume and area may prove to be useful in retreatment decision making [14]–[16]. Therefore it is becoming very important to quantify the exact amount, and measure the increase or the decrease of each fluid type to guide different retreatment regimens.

At present, only qualitative assessment of such fluid lesions in the form of detecting their presence and evaluating their extent is incorporated in the clinical workflow to guide anti-VEGF retreatment decisions. However modern OCT devices acquire large volumes of information that is difficult to manage manually and interpret qualitatively. In addition, individual subjectivity interferes with qualitative estimates of "fluid stability" or "fluid change". Thus, automated OCT image analysis tools that can quantify fluid are expected to be a driving force behind personalized and predictive medicine in the anti-VEGF treatment of macular edema [15], [16].

There has been a lot of effort in the recent years from the medical imaging community to develop automated methods for retinal OCT fluid quantification which could augment the clinical workflow. Nevertheless, it is currently not clear what is the state-of-the-art performance and how it varies across OCT images from different vendors and across retinas with different macular diseases. This is mainly due to the lack of standardized evaluation frameworks and annotated datasets.

To address the above points and investigate to what extent current deep learning methods are ready for automated interpretation of retinal OCT, we invited the medical imaging community to participate in a challenge by developing and testing automated OCT fluid detection and segmentation methods. In this paper we introduce a benchmark and report the analysis of the challenge RETOUCH. The challenge was a multi-group collaborative effort and featured data annotations from two clinical centers, images acquired with the three most common OCT device vendors and from patients with two different retinal diseases, which allowed for the first time to make an analysis of method performance and multicenter grading agreement for all three retinal fluid types and along those components of variability. The 112 OCT volumes (11,334 B-scans) and manual annotations for training are publicly available as an ongoing benchmarking resource at https://retouch.grand-challenge.org, the largest such available dataset to date.

II. Related prior Work

OCT was developed in 1990s [8] but only after the implementation of spectral domain SD-OCT, which became commercially available in 2006 and permitted faster signal acquisition, did the quality and resolution of images allow successful use of quantitative image analysis. Thus, retinal image analysis of OCT is a young field with the automated segmentation of retinal layers [17] being one of the earliest applications in healthy and mildly diseased retinas followed by the fluid segmentation in retinas with macular edema [18].

The first work on fluid segmentation in OCT was a semi-automated 2D approach based on active contours [19]. To segment IRF and SRF, one user interaction per lesion and B-scan was necessary to initialize the contour. A similar level set approach but using a fast Split Bregman solver was later used in [20] to generate all candidate fluid regions automatically which were then manually discarded or selected. However, semi-automated approaches are labor intensive and hence of very limited clinical use, devoting the recent work exclusively to a fully-automated segmentation.

An early automated approach addressed the problem as a local anomaly detection based on retinal texture and thickness properties [21], and was applied to determine 2D en-face footprints of fluid-filled regions. A fully 3D approach based on voxel classification followed by a graph-cut based segmentation was presented in [22]. A retinal layer-specific voxel-classification approach was proposed in [23]. A combination of a fuzzy C-means for initialization followed by an evolution of level set contour in three orthogonal OCT planes was proposed in [24]. Methods based on neutrosophic sets coupled with graph based methods were also proposed recently [25], [26].

A duality of retinal layer and fluid segmentation problems was recognized early on. Because IRF, SRF and PED are layer-specific, their segmentation would greatly benefit from an accurate retinal layer segmentation and vice versa. To jointly segment both the intraretinal layers and the fluid a voxel classification followed by dynamic programming based segmentation was used in [27]. Loosely coupled level sets were developed in [28] to jointly segment fluid and retinal layers by modeling the fluid as an additional space-variant layer. A purely data-driven model with minimum hard constraints was proposed in [29], based on auto-context and graph-cut segmentation to simultaneously segment retinal fluid and layers by learning their mutual interaction.

Segmentation of PED was initially approached as a layer segmentation problem, due to its deformed shape, having various subtypes, and its loose definition of sub-RPE fluid and non-fluid material. In [30], the authors use 2D anomaly footprint of [21] as a prior to adjust the multi-layer graph-search segmentation method. In [31], a multi-surface segmentation using graph-search is developed with different smoothness constraints for RPE and BM surfaces. In [32], BM surface was first estimated from the convex hull of RPE, followed by a shape-constrained graph-cut. In [33], [34], they segment SRF and PED fluid pockets by building a 3D fluid probability map from voxel-level texture, intensity and thickness scores, followed by a continuous max-flow segmentation.

Since the deep learning showed its great promise in 2012 [35], soon it made its way to retinal OCT analysis. A multi-scale convolutional neural network (CNN) was first proposed in [36] for patch-based voxel classification and it was able to differentiate between IRF and SRF fluid in both a fully supervised and a weakly supervised setting. CNN showed success also in segmenting retinal layer boundaries [37] and classifying vertical columns of OCT A-scans [38]. Currently, fully convolutional neural nets (FCNN [39] and U-net [40]), trained end-to-end are the state of the art. They were used in [41]–[44] to segment IRF, and in [45] to segment both the retinal layers and the fluid. Large validation studies were recently performed in [42], [46], which showed that FCNN can segment fluid across OCT devices and macular diseases.

Automated detection of fluid presence has received much less attention despite the presence being part of many clinical guidelines for retreatment. The fluid segmentation result of [23] was also extended to the task of fluid detection, achieving area under the curve (AUC) of 0.8 and 0.92 for two expert annotations, respectively. A method validated in [47] achieved an accuracy of 91% for detecting the presence of fluid when compared to the majority grading by three retinal specialists, but it is part of a commercial system so few methodological details were provided. As part of the IRF and SRF segmentation validation study, direct application of the number of segmented voxels in the detection task was demonstrated in [46]. Recently, presence of intraretinal fluid considered to be clinically relevant was detected in [48], using intensity and texture-based features combined with a classifier learned from representative samples only.

An overview of the selection of the discussed fluid segmentation algorithms, comparing them across a number of properties is shown in Table I. One can observe the variability in utilizing 3D context as some methods run segmentations only in 2D on a B-scan level. This is partly due to very anisotropic resolution of OCT and possible motion artifacts across B-scans. Furthermore most methods do not discriminate between different types of fluid but this leads to limited clinical applicability due to different associated prognostic

TABLE I
OVERVIEW OF THE RELATED WORK.

Reference	Fluid type	OCT vendor	Disease	Test set [# volumes]	Evaluation metric	2D/3D
Xu et al. (2015) [23]	any	Topcon	AMD	30	TPR/TNR	3D
Chiu et al. (2015) [27]	any	Spectralis	DME	10	DSC	2D
Sun et al. (2016) [32]	PED	Topcon	AMD	25	DSC	3D
Wang et al. (2016) [24]	IRF, SRF	RTVue	DME	10	DSC	2.5D
Montuoro et al. (2017) [29]	IRF, SRF	Cirrus/Spectralis	RVO/DME	100/10	DSC	3D
Roy et al. (2017) [45]	any	Spectralis	DME	10	DSC	2D
Lee et al. (2017) [41]	IRF	Spectralis	AMD, RVO, DME	30	DSC	2D
Novosel et al. (2017) [28]	SRF	Spectralis	CSR, DME	25	TPR/FPR, DSC	3D
Wu et al. (2017) [33]	SRF, PED	Cirrus	CSR	37	TPR/FPR, DSC	3D
Gopinath et al. (2018) [43]	IRF	Cirrus, Spectralis, Topcon, Nidek, Optovue	AMD, RVO, DME	35	DSC	3D
Girish et al. (2018) [44]	IRF	Cirrus, Spectralis, Topcon, Nidek	AMD, RVO, DME	15	DSC	2D
Venhuizen et al. (2018) [42]	IRF	Cirrus, Spectralis, Topcon, Nidek	AMD, RVO, DME	114	DSC, ICC	2D
Schlegl et al. (2018) [46]	IRF, SRF	Cirrus, Spectralis	AMD/RVO/DME	212/110/32	DSC	2D
OPTIMA Cyst Segmentation Challenge (2015) [49]	IRF	Cirrus, Spectralis, Topcon, Nidek	AMD, RVO, DME	15	DSC	
RETOUCH (2017)	**IRF, SRF, PED**	**Cirrus, Spectralis, Topcon**	**AMD, RVO**	**42**	**DSC, AVD**	

Abbreviations: Absolute volume difference (AVD), Dice Score (DSC), True positive rate (TPR, sensitivity), True negative rate (TNR, specificity), False positive rate (FPR), Intraclass correlation coefficient (ICC), Central serous retinopathy (CSR), Diabetic macular edema (DME).

significance. None of the methods were able to segment and discriminate between all three fluid types: IRF, SRF, and PED.

In summary, numerous methods have been published in the past but the methods were evaluated on different types of retinal images and using a different reference standard to report the accuracy. This makes an independent and fair comparison of their performance difficult. The only publicly available dataset, which some methods used for evaluation, was a data set from Duke containing 110 B-scans of 10 DME patients acquired with Spectralis OCT and annotated by two experts [27].

III. Challenge Setup

RETOUCH challenge aimed at creating a representative benchmark which can be used for evaluating algorithms for detecting and segmenting all of the three fluid types across retinal diseases and OCT vendors. This addressed significantly the current lack of large representative publicly available datasets and enabled method performance comparison. It goes substantially beyond the only prior fluid segmentation benchmark and the only OCT-based one, the OPTIMA Cyst Segmentation Challenge (OCSC) [49]. Other opthalmic image analysis benchmarks were addressing diabetic retinopathy (DR) and were based on color fundus images, namely the Retinopathy Online Challenge (ROC) that benchmarked algorithms for automatic detection of microaneurysms [50] and Diabetic Retinopathy Detection Kaggle challenge [51] aimed at automatically diagnosing DR disease stages.

RETOUCH extends substantially the OCSC in multiple ways (Table I). First, as it became evident that different fluid types have different clinical roles, all three fluid types were included in the challenge. Second, as the deep learning methods are the state of the art, a large training set was made available to enable training of such data-intensive models. Third, annotations and scans came from different clinical centers: Medical University of Vienna (MUV) in Austria, Erasmus University Medical Center (ERASMUS) and Radboud University Medical Center (RUNMC) in The Netherlands. Fourth, a total of 112 OCT volumetric scans are made available, the largest to date by a large margin, and were split into training and test set with $\approx 60\% - 40\%$ ratio. Fifth, the challenge evaluates the performance of the algorithms on two different tasks, both of high clinical significance: (1) fluid detection, and (2) fluid segmentation.

The challenge was launched in April 2017 by releasing the training data set on Grand Challenges in Biomedical Image Analysis hosting platform under https://retouch.grand-challenge.org. The challenge was announced on MICCAI 2017 and `grand-challenge.org` websites. The groups with a publishing track record in the field were invited personally over email. In addition, the challenge was advertised on mailing lists with wide international visibility and during the ARVO 2017 conference, the largest venue for vision research. This resulted in 64 teams signing the form and downloading the data over the course of the challenge. Out of those, nine teams submitted papers describing their work by the end of July 2017 deadline. One submission was rejected due to insufficient description of the method and eight were finally accepted to participate in the challenge. Overview of the methods used by the eight participating groups is shown in Table II.

The test set was released in Aug. 2017 and the participants had to submit the results by the end of the month. In case of multiple submissions, the last one was considered for the challenge. The individual results were not revealed during that month to avoid tuning on the test set. The two organizing groups (MUV and RUNMC), although having already published segmentation algorithms [42], [46] did not participate

TABLE II
CHALLENGE PARTICIPATING TEAMS IN ALPHABETICAL ORDER AND THE SUMMARY OF THEIR METHODS.

Team	Network	Data aug.	Layer seg.	Post-process	2D/3D
Helios	U-net	-	-	morphologic	2D
MABIC	U-net	x	-	U-net	2D
NJUST	Faster R-CNN	-	x	3D smooth	2.5D
RetinAI	U-net	x	x	-	2D
RMIT	U-net + adversarial	x	-	median filt.	3D
SFU	U-net	x	x	rand. forest	2D
UCF	ED-ResNet	x	-	graph-cut	2.5D
UMN	CNN	-	x	morphologic	2D

'x' denotes the use of data augmentation or retinal layer segmentation

due to direct conflicts of interest. The results of the evaluation on the test set were presented at a satellite workshop of MICCAI in Sep. 2017 in Quebec City, Canada.

IV. EVALUATION FRAMEWORK

A. *OCT Imaging Dataset*

We collected and anonymized a total of 112 macula-centered OCT volumes of 112 patients from MUV and ERASMUS. Half of the patients had macular edema secondary to AMD and half of them had edema secondary to RVO. OCT volumes were acquired with spectral-domain SD-OCT devices from three different vendors: Cirrus HD-OCT (Zeiss Meditec), Spectralis (Heidelberg Engineering), and T-1000/T-2000 (Topcon). The distribution of OCT volumes across the three vendors were uniform ($\approx$38 each). The device vendor was made known but the underlying retinal disease was not revealed to the participants.

Each Cirrus OCT consisted of 128 B-scans with a size of 512×1024 pixels. OCT acquired with the Spectralis device consisted of 49 B-scans with 512×496 pixels. OCT acquired with Topcon devices consisted of 128 B-scans with a size of $512{\times}885$ (T-2000) or $512{\times}650$ (T-1000) pixels. All the OCT volumes were covering a macular area of $6{\times}6$ mm^2 with axial resolutions of: 2 µm (Cirrus), 3.9 µm (Spectralis), and 2.6/3.5 µm (Topcon T-2000/T-1000).

The training set consisted of a set of 70 OCT volumes, with 24, 24, and 22 volumes acquired with Cirrus, Spectralis, and Topcon, respectively. The test set consisted of a set of 42 OCT volumes, with 14 volumes corresponding to each of the three device vendors. The properties of the training and test sets are summarized in Fig. 3. The distributions between the two sets matched well.

B. *Reference Standard*

The reference standard was obtained from manual voxel-level annotations of the IRF, SRF and PED fluid lesions in each of the B-scans of each individual OCT volume (a total of 11,334 B-scans). Fluid of a particular type was considered to be absent if none of the voxels of the OCT volume were annotated as such, and present otherwise. Manual annotation tasks were distributed to human graders from two clinical centers: (1) MUV, where 4 graders were supervised by one ophthalmology resident, all trained by two retinal specialists. (2) RUNMC, where 2 graders were supervised by one retinal specialist. All annotations were performed on a B-scan plane; however, orthogonal planes were also viewed and considered in case of doubt. The supervisors reviewed all annotations and corrected any errors. The annotations guideliness were agreed upon in advance between the two centers before the annotations started and they stated that boundaries had to be visible for fluid regions to be annotated and that the annotated regions can not contain holes. For the annotation repeatability of each center, we refer to the evaluations reported in [42], [49], [52].

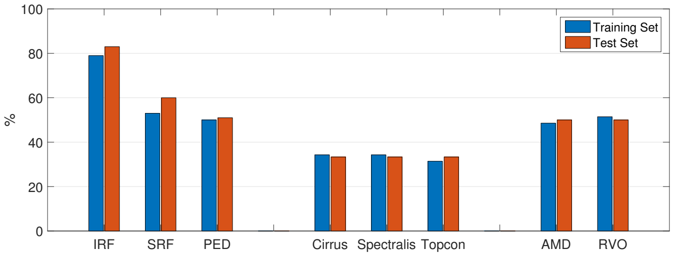


Fig. 3. Overview of the training and test set properties. Percentage of scans containing a particular fluid type, OCT vendor and retinal disease.

To keep the annotation effort feasible and maximize the total amount of annotated OCT volumes, training set was annotated only once, with OCT volumes from Cirrus and Spectralis devices annotated by MUV and Topcon volumes annotated by RUNMC. All the test set OCT volumes were annotated twice, by graders from both MUV and RUNMC to account for inter-center variability in the evaluation.

Test set annotation aggregation: A single reference standard is created for the test set using consensus, i.e., a strict combination of the annotations from the two centers. A fluid was determined to be present or absent in a scan only when both centers agreed. Similarly, only voxels for which the two centers' annotations agreed in the label were considered valid for segmentation evaluation. Voxels with inter-center disagreement were masked for exclusion. Such voxel-level annotation aggregation is illustrated in Fig. 4.

C. *Evaluation and Ranking*

Evaluation consisted of forming two main leaderboards corresponding to the two tasks: detection and segmentation. For each leaderboard, a *dense ranking* was used where teams with equal scores receive the same ranking number, and the next team receives the immediately following number without creating gaps in the ranking. The average rank across the two leaderboards determined the final ranking of the RETOUCH challenge. In case of a tie, better segmentation score had the priority.

1) Detection task: The teams submitted for each case probability of presence of each fluid type. These were compared to the manual grading of fluid presence. The cases with inter-center disagreement in the presence of a particular fluid type were excluded from evaluation of automated detection performance of that fluid type. For each of the three fluid types,

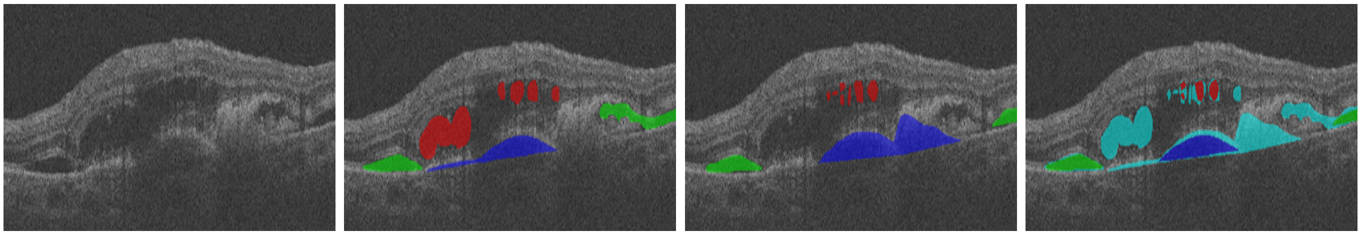


Fig. 4. Illustration of a reference standard on a specially chosen case with *extreme* disagreement. From left to right: OCT B-scan, annotation of center #1, annotation of center #2, and the consensus reference standard with the exclusion mask in cyan. Fluid: IRF (red), SRF (green), and PED (blue).

receiver operating characteristics (ROC) curve was created across the test set images with inter-center agreement, and an area under the curve (AUC) was calculated. Each team received a rank for each of the fluid types based on its AUC value. The detection score was determined by adding the three ranks. The detection leaderboard was created by ranking the teams by their detection scores.

2) Segmentation task: The teams submitted for each case a volume containing the segmentation results with each fluid type represented with its voxel label. The results were compared to the reference standard, where voxels masked for exclusion due to inter-center disagreement were ignored. For each OCT volume and fluid type the similarity of two samples, segmentation (X) and reference (Y), was measured using:

- Dice score (DSC), which measures voxel overlap of the two samples:

$$\mathrm{DSC} = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \qquad (1)$$

  where $|\cdot|$ denotes the number of voxels in the sample. TP marks the amount of true positives, FP the false positives, and FN the false negatives.

- Absolute volume difference (AVD) in [mm^3] between the two samples, which represents a clinically relevant parameter:

$$\mathrm{AVD} = \mathrm{abs}(|X| - |Y|). \qquad (2)$$

We chose to rank the participating teams in a number of categories and then sum the individual ranks, as done similarly in BRATS [53] and MRBrainS [54] challenges. Due to the large image quality variability between OCT device vendors, segmentation results were additionally summarized per each vendor separately. Thus, teams were ranked for each combination of: *fluid type − OCT vendor − evaluation measure*, based on the mean evaluation measure over the corresponding subset of test images. The segmentation score was then determined by adding the 18 individual ranks (3 fluids × 3 vendors × 2 measures). Finally, the segmentation leaderboard was created by ranking the teams by their segmentation scores.

To test for statistically significant differences in inter-center agreement and automated method performance across fluid types, OCT vendors and retinal diseases, a nonparametric Wilcoxon-Mann-Whitney test was applied. To account for triple comparisons, Bonferroni correction was applied which adjusted the significance level from 5% to 2%. Similarly, we tested for significant difference in performance between participant methods by comparing the distributions of their obtained DSC values.

TABLE III
DETECTION TASK: INTER-CENTER AGREEMENT RATE FOR A FLUID TYPE, OCT VENDOR, AND RETINAL DISEASE.

	IRF	SRF	PED	All
Cirrus	1.00	1.00	0.86	0.95
Spectralis	1.00	0.86	0.93	0.93
Topcon	0.86	1.00	1.00	0.95
AMD	0.90	0.90	0.95	0.92
RVO	1.00	1.00	0.90	0.97
All	0.95	0.95	0.93	

To evaluate the sensitivity to the quantity of fluid volume present in the scan, we analyze the performance over equal groups of fluid volume divided into fifths (quintiles), each corresponding to 20% of the volume range. The following were the four quintile values denoting the 20th, 40th, 60th, 80th-percentile, respectively. IRF: 0.0487, 0.1612, 0.2587, 0.5945 mm^3, SRF: 0.0040, 0.0888, 0.2121, 0.7425 mm^3, and PED: 0.1879, 0.3890, 0.7501, 1.5381 mm^3.

V. RESULTS

A. Fluid Detection Task

1) Inter-center agreement: The agreement per fluid was high, with a rate of: 0.95 (40/42), 0.95 (40/42), and 0.93 (39/42) for IRF, SRF, and PED, respectively. Inter-center detection agreement for each fluid type across OCT vendors and retinal diseases is shown in Table III. Agreement in detecting IRF and SRF in scans of patients with RVO was noticeably better than in patients with AMD. There was no noticeable difference in agreement across OCT vendors.

The performance of a center's gradings by using the other center as the ground truth is analyzed in Fig. 5. The two centers' operating points were favoring either maximizing sensitivity (center 1 for SRF, and center 2 for IRF and PED) or maximizing specificity (center 1 for IRF and PED, and center 2 for SRF).

2) Automated method performance: Average performance of teams measured using AUC for each fluid type across device and disease subgroups is summarized in Table IV. The few case-fluid combinations with disagreement were excluded from evaluation of team performances. IRF was found to be the most difficult to detect and PED the easiest. Topcon scans were found to be the most challenging. However no statistically significant differences (paired test) were observed

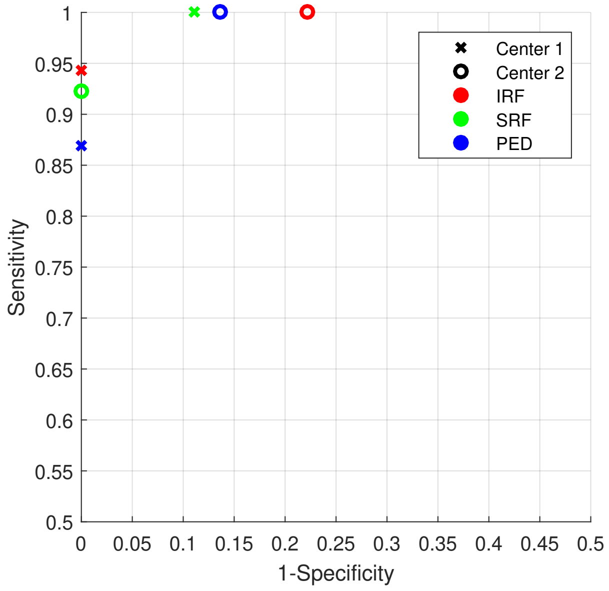


Fig. 5. Detection task: inter-center agreement. Operating points for each fluid type when taking the other center's gradings as the ground truth.

TABLE IV
DETECTION TASK: MEAN (STANDARD DEVIATION) AUC ACROSS TEAMS FOR A FLUID TYPE, OCT VENDOR, AND RETINAL DISEASE.

	IRF	SRF	PED	All
Cirrus	0.94 (0.18)	0.89 (0.10)	0.99 (0.03)	0.94 (0.08)
Spectralis	0.89 (0.12)	0.92 (0.11)	0.97 (0.04)	0.93 (0.05)
Topcon	0.84 (0.19)	0.95 (0.08)	0.93 (0.12)	0.90 (0.08)
AMD	0.88 (0.11)	0.91 (0.10)	N/A	0.89 (0.08)
RVO	N/A	0.93 (0.08)	N/A	0.93 (0.08)
All	0.88 (0.12)	0.92 (0.08)	0.96 (0.06)	

'N/A' denotes *not applicable* when the fluid is always present or absent.

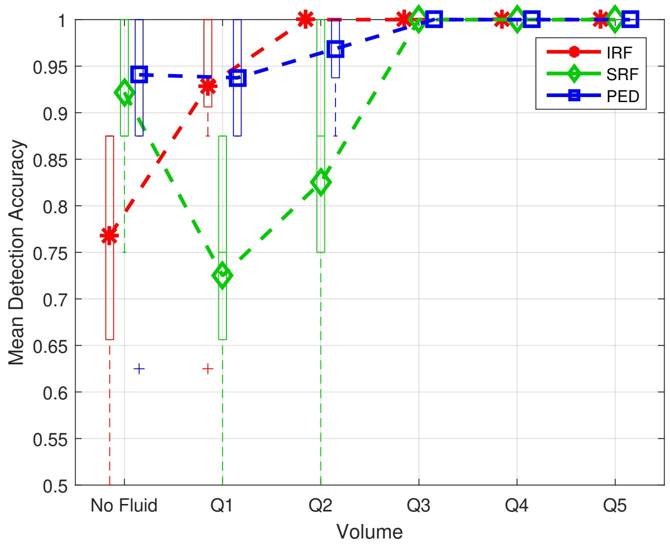


Fig. 7. Detection task: Box-plots of the mean team detection accuracy across cases with different fluid volume quintiles.

TABLE V
DETECTION TASK LEADERBOARD WITH AUC VALUES FOR EACH FLUID TYPE.

Rank Sum	Team	IRF	SRF	PED
3	SFU	1.00	1.00	1.00
6	UCF	0.94	0.92	1.00
8	Helios	0.93	1.00	0.97
10	MABIC	0.86	1.00	0.97
10	RMIT	0.71	0.92	1.00
11	RetinAI	0.99	0.78	0.82
11	UMN	0.91	0.92	0.95
13	NJUST	0.70	0.83	0.98
	Majority Vote	1.00	1.00	1.00

in team performance between devices and diseases, partly due to small sample size (8 teams). Examples of selected difficult cases are displayed in Fig. 6.

Mean team detection accuracy across cases with different volume quintiles are shown in Fig. 7. For this evaluation, operating points were selected corresponding to Youden index, i.e., maximal sensitivity+specificity. The methods clearly struggled in detecting smaller fluid quantities. Only the detection of PED performed with a high mean accuracy (> 0.90) across all different volumes.

Taking into account individual team AUC performance for each fluid, the detection leaderboard is shown in Table V. For further comparison, a result of an ensemble is created, where a majority vote was obtained by summing the scores using the supplied detection probabilities. The majority vote produced a perfect AUC of 1.0 for all three fluid types. Thus, we can conclude that in the detection task automated methods were able to achieve the performance comparable to human graders.

B. Fluid Segmentation Task

1) Inter-center agreement: The annotation agreement had an overall DSC mean (standard deviation) of 0.73 (0.17). Correlation in annotated volume sizes in log scale was: 0.98, 0.98, and 0.90, for IRF, SRF, and PED, respectively (Fig. 8). Overlap for each fluid type across OCT vendors and retinal diseases is summarized in Table VI. Overlap in segmenting PED was significantly higher ($p = 0.01$) than in segmenting IRF but no statistically significant difference in agreement was found between different devices and diseases. Limits of agreement and bias in annotated fluid volumes are presented in Fig. 9. It can be observed that SRF annotations were the most consistent and that for SRF and PED annotations the bias was either very small in size or not statistically significant. The main discrepancies were in IRF annotations where one center was annotating more conservatively.

2) Automated method performance: Mean team performance across device and disease subgroups is summarized in Table VII. No statistically significant differences could be observed in mean team DSC values across the subgroup cases. However, the eight teams did perform significantly better on RVO cases than on AMD ones ($p = 0.01$). They also performed substantially although not significantly better on Cirrus cases compared to Spectralis cases ($p = 0.07$), and on segmenting PED compared to SRF ($p = 0.11$). A series of qualitative examples of successful and unsuccessful results

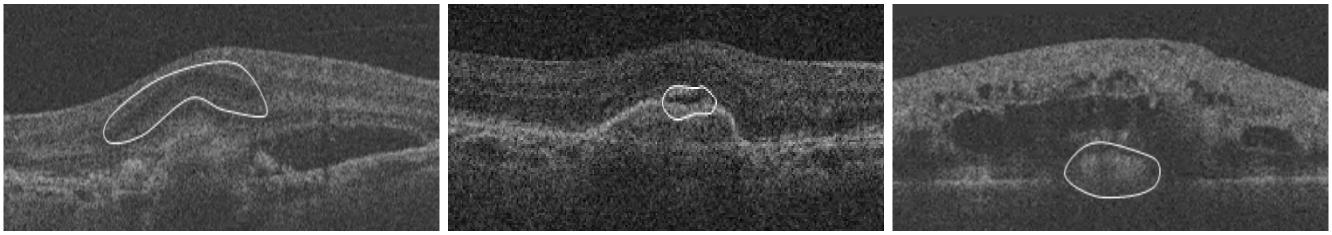


Fig. 6. Detection task: Examples of difficult cases (from left to right): Topcon scan where IRF was wrongly detected by 4/8 teams. Cirrus scan where SRF was missed by 5/8 teams. Topcon scan where PED was wrongly detected by 3/8 teams. Regions which were creating difficulties are denoted in white.

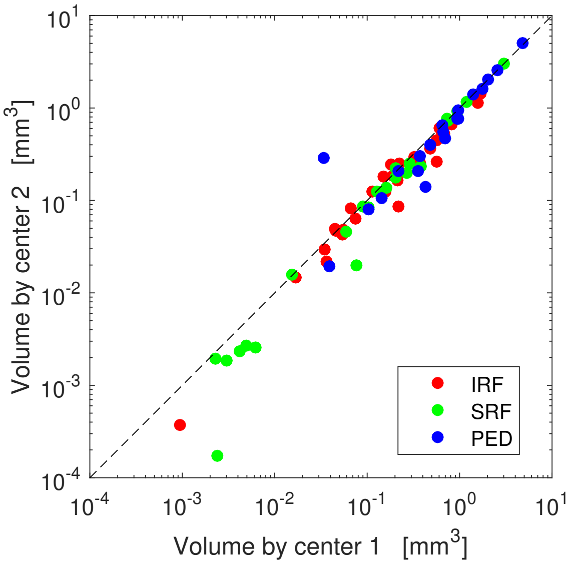


Fig. 8. Segmentation task: Annotated fluid volumes across the two centers.

TABLE VI
SEGMENTATION TASK: MEAN (STANDARD DEVIATION) INTER-OBSERVER DSC FOR A FLUID TYPE, OCT VENDOR, AND RETINAL DISEASE.

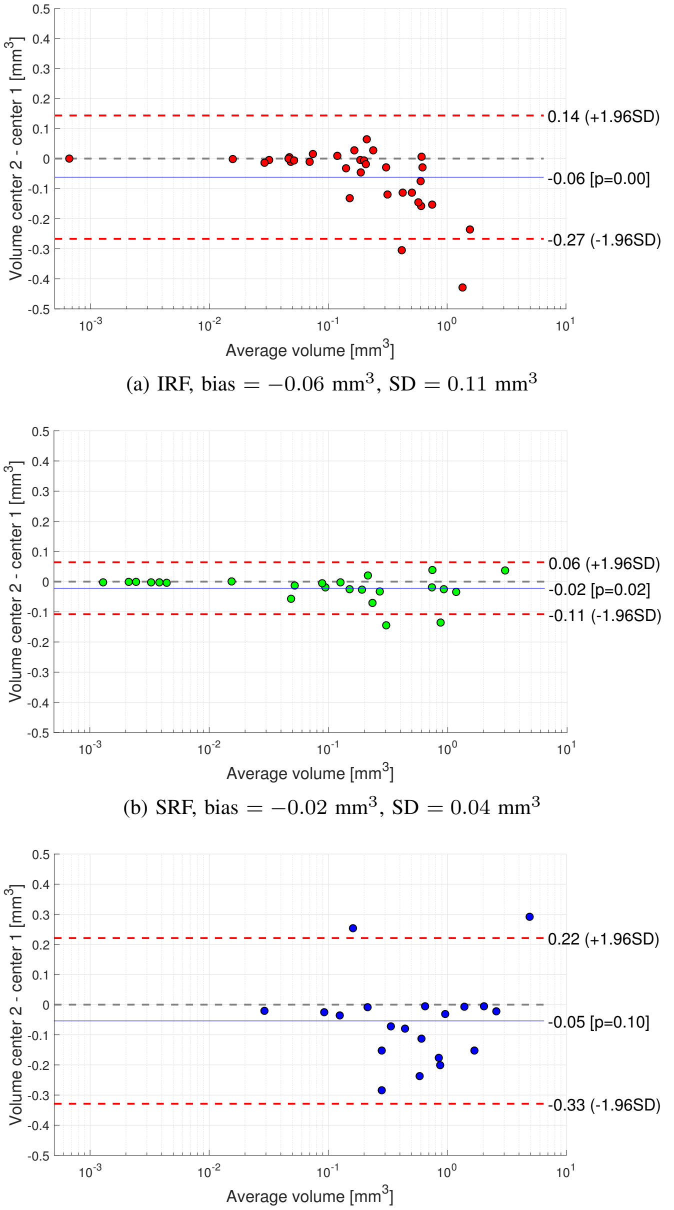
	IRF	SRF	PED	All
Cirrus	0.74 (0.07)	0.74 (0.16)	0.78 (0.11)	0.75 (0.10)
Spectralis	0.74 (0.09)	0.66 (0.29)	0.80 (0.15)	0.73 (0.18)
Topcon	0.63 (0.15)	0.79 (0.14)	0.70 (0.31)	0.71 (0.21)
AMD	0.66 (0.13)	0.71 (0.26)	0.76 (0.21)	0.72 (0.21)
RVO	0.74 (0.09)	0.75 (0.12)	N/A	0.75 (0.10)
All	0.71 (0.11)	0.73 (0.20)	0.76 (0.21)	

'N/A' denotes *not applicable* due to the absence of the fluid type.

across all three OCT vendors are shown in Fig. 10. We can observe that the loss of OCT signal due to shading of vessels and fluid pockets, as well as large pathological deformation of the retina, were the main sources of difficulties for automated methods.

Mean team DSC values across cases with different volume quintiles are shown in Fig. 11. Similar to detection performance, the methods had consistently worse performance in segmenting smaller fluid quantities, but DSC metric may favor the detection of larger fluid volumes [55].

The obtained segmentation leaderboard is shown in Table VIII, and the corresponding box-plots are shown in Fig. 12. Comparing the teams across all the 18 individual ranks,



(a) IRF, bias $= -0.06$ mm^3, SD $= 0.11$ mm^3

(b) SRF, bias $= -0.02$ mm^3, SD $= 0.04$ mm^3

(c) PED, bias $= -0.05$ mm^3, SD $= 0.14$ mm^3

Fig. 9. Bland-Altman plots showing the bias (blue line), the standard deviation (SD) and the limits of agreement (red lines) for the measured fluid volumes between the two centers.

TABLE VII
SEGMENTATION TASK: MEAN (STANDARD DEVIATION) MEAN TEAM DSC FOR A FLUID TYPE, OCT VENDOR, AND RETINAL DISEASE.

	IRF	SRF	PED	All
Cirrus	0.73 (0.13)	0.63 (0.39)	0.74 (0.18)	0.70 (0.24)
Spectralis	0.69 (0.11)	0.57 (0.31)	0.68 (0.30)	0.65 (0.23)
Topcon	0.61 (0.26)	0.75 (0.16)	0.66 (0.22)	0.67 (0.22)
AMD	0.61 (0.13)	0.60 (0.36)	0.69 (0.23)	0.64 (0.25)
RVO	0.73 (0.18)	0.72 (0.18)	N/A	0.72 (0.18)
All	0.68 (0.17)	0.66 (0.29)	0.69 (0.23)	

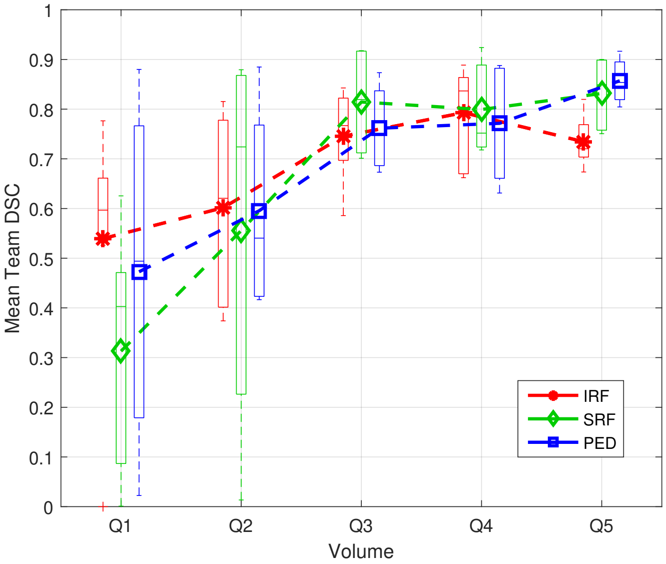


Fig. 11. Segmentation task: Box-plots of the mean team DSC across cases with different fluid volume quintiles.

the first team was substantially better than the second team ($p = 0.07$) and significantly better than the rest ($p < 0.01$). Second and third teams were significantly better than the sixth and onwards ($p = 0.02$). Unlike detection performance, the automated segmentations were not within the inter-center agreement. Finally, we also include ensemble results obtained by fusing the team segmentation labels using majority voting. Such label fusion already outperformed all of the methods, showing that there is a potential for further performance improvement.

VI. DISCUSSION

A. Main Findings

With RETOUCH we introduced a dataset that serves as benchmark and evaluation framework with standardized training and test sets of OCT images. We have performed an extensive analysis of the results submitted by the eight teams that participated in the challenge. With the analysis we aimed at answering the following set of questions:

a) How good is the automated method performance?: In the detection task one of the methods already achieved perfect score and the same is achieved with majority voting method fusion. Thus we can conclude that the automated detection at expert level is possible, i.e., it operates within inter-center grading variability. However, this is not the case with segmentation task where the best methods achieved DSC $= 0.7 - 0.8$ on the consensus reference and the majority voting label fusion outperformed all of the methods, showing that there is ample space for improvement. The winning method (team SFU) was proved the best in both detection and segmentation tasks. Interestingly, some methods (teams UMN and UCF) did very well on one task but poorly on the other.

b) How did the automated methods differ?: All the methods were deep learning based (Table II), with most teams implementing a variant of a fully convolutional network [39], [40], [56]. Thus the differences were in the extent of pre/post-processing, usage of auxiliary retinal layer segmentation and training details. In particular, the size and complexity of the networks varied hence different regularization approaches were used to prevent overfitting. Furthermore, teams employed different strategies in the training to compensate for class imbalance. All the teams applied their networks in 2D, i.e., per B-scan. To benefit from 3D context, one team (RMIT) trained their models to utilize the two neighboring B-scans, and two teams (NJUST, UCF) utilized the 3D context in the post-processing stage. Most teams standardized the image size and intensities across OCT devices but the two top teams were the only ones that trained a neural network for each of the three devices separately. Finally, the top team was the only one that combined all the main elements: layer segmentation, data augmentation, and extensive post-processing.

c) How good is the inter-center annotation agreement?: The fluid detection agreement was high (≈ 0.95) but fluid segmentation agreement had a mean DSC $= 0.73$. This is similar to the human inter-rater mean DSC $= 0.75$ reported in [41], and reflects the difficulty of the manual fluid annotation task. In comparison to other image analysis challenges, segmentation agreement falls in-between brain tumor segmentation (BRATS) [53] where DSC $= 0.74 - 0.85$ were reported, and ischemic stroke lesion segmentation (ISLES) where DSC $= 0.7$ [57], both also very difficult tasks for even manual annotation.

d) How do the motion artifacts and noise impact performance?: All the teams, with the exception of RMIT, perform the OCT segmentation in 2D, B-scan per B-scan. In such an approach, the impact of motion artifacts is avoided at the cost of losing the benefit of 3D context. To tackle the speckle noise, the preprocessing approaches utilized traditional denosing techniques, ranging from Gaussian (UCF) and median filtering (RMIT) to bilateral filtering [58] (NJUST) and total variation-based [59] denoising (Helios). However, we observed that CNNs even without any denoising showed good robustness to noise (Fig. 10) and the degradation of image quality caused by retinal abnormalities attenuating OCT signal had a stronger detrimental impact than the speckle.

e) Is there a difference in performance between fluid types?: There was no observed difference in inter-center detection agreement but agreement in segmentation of PED was better than for SRF and IRF. That is somewhat expected as PED is easier to visually spot as it consists of layer detachment. It is also encouraging that graders from the two centers had the same notion of what constitutes a PED, given that it

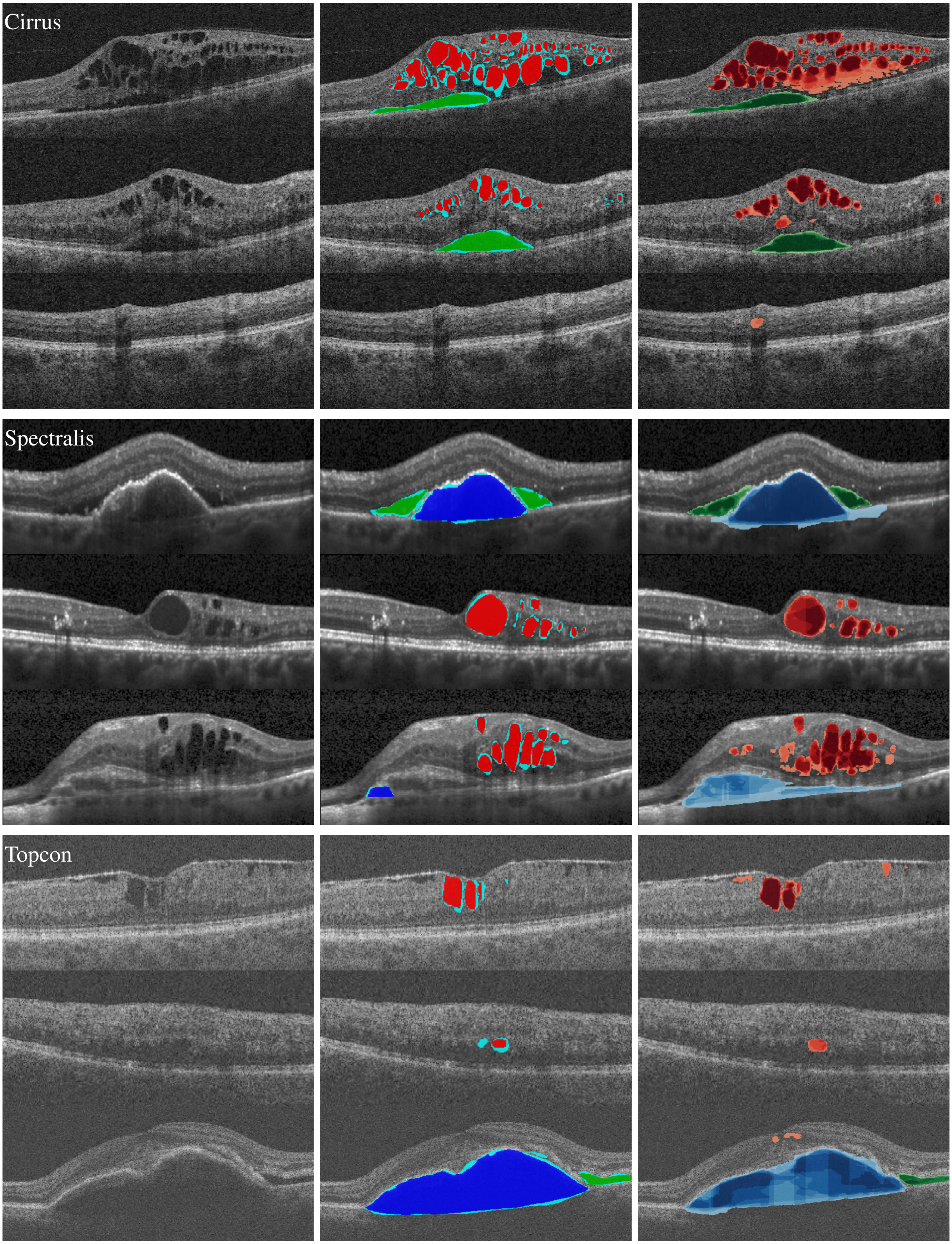


Fig. 10. Segmentation task. Qualitative results (rows) grouped by OCT vendors: Cirrus, Spectralis, and Topcon. Columns: (Left) OCT B-scan, (Middle) reference standard with the exclusion mask (cyan), (Right) the distribution of team segmentations with darker hue denoting better agreement of the 8 teams. The bottom example of each vendor group was found to be especially challenging to segment. Fluid: IRF (red), SRF (green) and PED (blue).

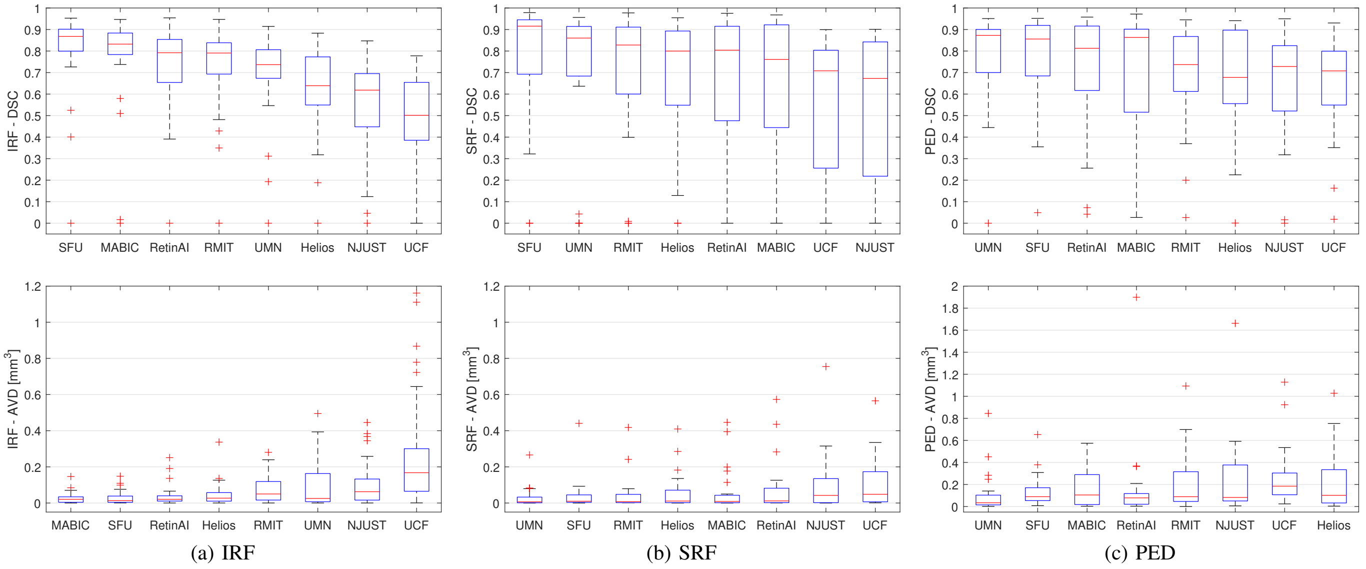


Fig. 12. Segmentation task: Box-plots illustrating the DSC (top row) and AVD (bottom row) performance of the teams across the three fluid types, sorted by the mean DSC and AVD, respectively.

TABLE VIII
SEGMENTATION TASK LEADERBOARD. MEAN AND STANDARD DEVIATION OF VALUES ACROSS CASES FOR DICE SCORE (DSC) AND ABSOLUTE VOLUME DIFFERENCE (AVD) IN [MM^3].

Rank Sum	Team	IRF		SRF		PED	
		DSC	AVD	DSC	AVD	DSC	AVD
39	SFU	0.82 (0.19)	0.030 (0.036)	0.75 (0.30)	0.041 (0.089)	0.74 (0.24)	0.140 (0.154)
59	UMN	0.70 (0.20)	0.088 (0.131)	0.71 (0.33)	0.032 (0.058)	0.77 (0.23)	0.119 (0.207)
64	MABIC	0.78 (0.22)	0.027 (0.032)	0.66 (0.32)	0.064 (0.123)	0.70 (0.29)	0.167 (0.169)
73	RMIT	0.73 (0.20)	0.078 (0.079)	0.70 (0.31)	0.046 (0.094)	0.69 (0.25)	0.245 (0.290)
74	RetinAI	0.74 (0.19)	0.039 (0.054)	0.67 (0.33)	0.079 (0.147)	0.71 (0.29)	0.189 (0.416)
88	Helios	0.63 (0.19)	0.048 (0.064)	0.68 (0.30)	0.059 (0.103)	0.66 (0.26)	0.297 (0.503)
120	NJUST	0.57 (0.21)	0.107 (0.124)	0.53 (0.34)	0.103 (0.169)	0.63 (0.27)	0.253 (0.380)
130	UCF	0.49 (0.20)	0.276 (0.319)	0.54 (0.33)	0.112 (0.140)	0.63 (0.24)	0.280 (0.285)
	Majority Vote	0.83 (0.17)	0.027 (0.036)	0.79 (0.31)	0.027 (0.048)	0.80 (0.24)	0.095 (0.110)

appears in different forms and subtypes. Method performance in detecting PED was also observed to be higher than SRF, which was furthermore higher than IRF. In segmentation task, similar method performance was observed among all three fluid types.

f) Is there a difference in performance between OCT vendors?: One could expect that automated method performance on Spectralis scans would be the highest due to their superior SNR (Figs 1, 10). However they contained smaller number of B-scans: 49 compared to 128 in Cirrus and Topcon, effectively making the Spectralis training set more than 2 times smaller. In addition, their larger spacing between the B-scans may hinder the ability to exploit the 3D context. Those factors may partly explain why the mean automated segmentation performance on Spectralis scans was the lowest. In general, the variability of fluid lesion manifestation, which directly affects the difficulty of a task, is assumed to be larger and more dominating factor than the variability of image quality across device vendors.

g) Is there a difference in performance between retinal diseases?: One would expect that scans of retinas suffering from RVO would be easier to quantify as AMD is a more morphologically damaging disease. Indeed, our analysis found that inter-center agreements and automated method performances were consistently better for macular edemas secondary to RVO, compared to AMD, in both detection and segmentation.

B. Strengths and Limitations

The challenge dataset goes substantially beyond what has been available before in both its size (112 OCTs with 11,334 B-scans) and variability. For the first time, all three retinal fluid types have been addressed simultaneously and they were present with a wide range of volumes, from tiny pockets to large regions. The scans were assembled from multiple clinical centers. Even though acquisition patterns vary in practice, on our dataset we kept the scanning patterns from the same OCT vendor homogeneous. All Spectralis volumes consisted of 49 B-scans, while Cirrus and Topcon had 128. The B-scan resolutions were similarly kept homogeneous. This limited the preprocessing and normalization needed while still capturing the inter-vendor scan variability.

Clinical information was purposely not provided to the participants. RVO is a disease of retinal vasculature, which lies within the retina and it almost always produces IRF but rarely PED. On the other hand, wet AMD originates from the choroid below the retina and it almost always produces some form of PED. In practice, knowing the underlying disease could be used as strong prior for fluid presence.

Despite its large overall size, the sizes of individual fluid−device−disease categories were found to be of low statistical power to capture differences in performances. Therefore, it would be desirable to increase in the future the number of test cases. This would also better capture the large variability of fluid lesion phenotypes occurring in practice and assess better the generalization capabilities of automated methods. However, increasing the training set substantially may not be necessary as fusion label techniques, e.g., majority voting, were already able to boost the performance.

The annotations were performed at two different medical centers from two countries. This allows to evaluate grading variability that includes the inter-center variability. The two annotations were used to generate the test set reference standard by a strict agreement, i.e., voxel-level consensus, because adjudication or including a third medical center was not considered feasible given the physical distance betweem the centers and the amount of annotations performed (4270 B-scans in the test-set). In such a scenario, the automated quantification would ideally achieve a perfect score as there is by definition no inter-center variability in the consensus. The training set was also labeled by the two medical centers but each case was labeled by a single center. Thus inconsistencies in training labels may be present. However, this was compensated by having larger total amount of cases available for training, which should allow the methods to overcome potential inter-center inconsistencies.

The performed analysis was focused on the DSC measure which is known to be overly sensitive to cases with small fluid volume. To compensate for this effect, we introduced an additional measure AVD into the leaderboard rankings, which directly reflects volume size discrepancies.

Finally, the eight teams that competed represent only a small subset of the teams working in this field as many recognized research groups (section II) were not able or decided not to take part.

C. Path Forward

While fluid detection has clear clinical relevance for screening where high sensitivity is paramount, the level of fluid segmentation performance needed for clinical use is currently not clear and it likely depends on the application. In fact, the results of automated fluid segmentations have already been used for building successful predictive models of treatment responders [60], retreatment need [16], [61] and longer term treatment requirements [62], as well as analyzing and predicting visual acuity outcomes under anti-VEGF treatment [63]. Thus, a potential path forward for future retinal OCT challenges would be to focus on directly predicting future clinical outcomes, analogous to TADPOLE challenge [64] for predicting Alzheimer's disease progression from brain scans. Such use in predictive modeling is especially beneficial as it would provides objective prognosis in contrast to currently subjective and variable ones provided by clinicians.

VII. Conclusion

RETOUCH is a benchmark and evaluation framework for automated detection and segmentation of retinal fluid from OCT. We thoroughly analyzed the results of the eight teams that participated in the corresponding MICCAI 2017 challenge. The winning method performed the best across both detection and segmentation tasks and hence was a clear winner of the challenge. However it did not win in all the categories considered. All of the analyzed methods were deep learning based. In particular most relied on U-net [40], a popular fully convolutional network architecture for medical image segmentation. The recent development of newer semantic segmentation algorithms, which further exploit the image context without loosing the spatial resolution [65]–[67], might push the future method performance even further.

The performance on the automated fluid detection was high, which makes it a very promising technology for real world deployment in the clinic. This would already be very helpful to clinicians as it would provide them with additional, objective "pair of eyes" in detecting the presence of retinal fluid. Automated fluid segmentation task was shown to still be a difficult challenge even for human experts as observed by a large inter-center grading variability. Large variability of fluid lesion phenotypes seemed to have dominated the variability in devices or underlying retinal diseases. However, fusing several methods using a majority vote produced segmentations that ranked above all individual methods, indicating opportunities for further improvements in methodology and consequently in performance.

Appendix A
Participating Methods

Below are short descriptions of each of the methods. More detailed information regarding the methods and the underlying techniques are available within the MICCAI challenge workshop proceedings [68].

A. Helios [69]

We propose a fully automated Generalized Motion Pattern (GMP) based segmentation method using a cascade of fully convolutional networks for detection and segmentation of retinal fluids. The GMP which is derived by inducing motion (rotation and translation) to an image to suppress the background. The segmentation and detection task are accomplished by providing GMP images as an input to a Fully Convolutional Network (FCN) U-Net based architecture. The detection is achieved by introducing fully connected layers at the end of first cascaded stage (the bottleneck).

We use binary cross entropy and dice coefficient based loss function for the detection and segmentation task respectively. Since abnormalities are 3D structures prevailing in multiple slices, considering k-neighboring slices for predicting and segmenting the retinal fluid aids in accurate detection and helps in eradicating false positives. The proposed method is parallelizable and handles inter-scanner variability efficiently.

B. MABIC [70]

We propose a two-step neural network for detection and segmentation of the retinal OCT fluid. The first network performs detection and segmentation of fluids, while the second network performs post-processing on the output images of the former network to enhance robustness of the former network. More specifically, both networks basically adopt the U-net architecture to segment the fluids.

In the first network, U-net is combined with one fully connected layer for fluid detection. The second network again trains the segmented fluids taking both OCT image and corresponding segmented image generated from former network as input data with two channels. The second network is separately trained for each type of fluid (i.e., IRF, SRF, and PED). To avoid the overfitting, dropout layer and maxout activation (instead of ReLU) are added. Binary cross entropy is used as a loss function for both detection and segmentation, and it is minimized using the AdaDelta optimizer.

As a pre-processing step, all OCT images acquired from 3 different devices are resized to 512×512 pixels (B-scans), and each image is normalized to the range [0,1]. Images are classified into one of eight classes (all combinations of IRF, SRF, and PED presence). Data augmentation is performed using translation along the horizontal and vertical axes ($\leq 15\%$), rotation ($\leq 5°$), scaling ($\leq 15\%$), and reflection along the horizontal axis. The compensation for class imbalance is not considered.

C. NJUST [71]

A novel segmentation method by combining Faster R-CNN [56], region growing and effective layer segmentation is proposed. Segmented fluid served directly as the detection result.

(1) Faster R-CNN for IRF segmentation: Faster R-CNN is a unified, deep learning-based real-time object detection method in computer vision. By sharing convolutional features with the down-stream detection network, the region proposal step is nearly cost-free. The learned region proposal network improves region proposal quality and thus the overall object detection accuracy. IRF lesion region always appears between ILM and IS/OS lines.During the training phase, we construct a smallest rectangular box containing IRF region. Then, we input training data, labels, and rectangular boxes to the Faster R-CNN. For each testing B-scan, the network would search the possible IRF region in the ROI obtained by effective layer segmentation and label the possible lesion area with rectangular box. Each rectangular has a score from Faster R-CNN. Finally, the IRF lesion region can be detected and segmented when the corresponding score is higher than the pre-setting threshold. Data augmentation: we used 11 different scales from [0.5,1.5] with a step of 0.1.

(2) 3D region growing for SRF segmentation: Between IS/OS and RPE, 3D region growing is applied to segment the SRF area. In the first step of 3D region growing, Faster R-CNN is utilized to generate the initial seed candidate set which is then clustered into two classes by k-means. The cluster centers are regarded as the seeds. The criterion for 3D region growing is that the intensity of the candidate voxel must be within one standard deviation from the mean of 26 neighboring points in 3D around the seeds.

(3) RPE layer segmentation for PED segmentation: A novel retinal layer segmentation method based on reflectivity distribution characteristics of retinal images was proposed for the OCT images with serious retinal diseases. PED will result in the arching of RPE, so we can segment the PED area directly by computing the thickness of RPE and BM after layer segmentation.

D. RetinAI [72]

The simultaneous classification and segmentation of the three fluid types is based on a modified version of the deep learning approach proposed by Apostolopoulos et al. [73] This CNN was designed as an encoder-decoder configuration, where each input image is processed by a series of convolutional blocks and contracting operations (encoder layers), followed by a series of convolutional blocks and expanding operations (decoder layers). In the new approach, the same architecture was reused and hyperparameters were set according to Apostolopoulos et al. [73]. Given an OCT image and an indexed five layer segmentation image as input, this CNN outputs an indexed map of the same dimensions, with a range of [0, 3], representing background and the three fluid types. The same input was fed to a RetiNet which detects the presence of the three fluid types as three binary outputs. The latter consists of a CNN, trained in an extreme learning fashion, which has shown good results for classification of both B-scans and C-scans [74]. Finally, the output of the classifier is used as a gate to remove false positive segmentations.

The classification and segmentation outputs were learned jointly, relying on stochastic gradient descent with an initial learning rate of 10^{-3}, which was halved every 20 iterations, and continued until the loss function converged. While fluid segmentation and classification is performed per B-scan, the diagnosis of a patient has to be performed on the full C-scans volume. Again, we use the RetiNet network, specifically the RetiNet-C configuration, which processes every B-scan in a single pass, using the RetiNet-B configuration as a feature extractor. Because C-scans from different devices have different resolutions (mm / pixel), they are first resampled to a common resolution using cubic spline interpolation.

E. RMIT [75]

The proposed fluid segmentation method uses a neural network model that consists of a modified U-net linked to an adversarial network. The modifications to the U-net include: batch normalisation following each convolutions/deconvolution block, dropout regularization at each skip connection, and multi-scale feature aggregation at the final convolutional layer. The adversarial network, trained to differentiate between predicted segmentation masks and ground-truth segmentations, helps encode higher order information between image regions and eliminates the need for a post-processing step. The output of the last feature layer of the segmentation U-net across all slices in the volume was aggregated and a second shallow network was trained to predict the presence of a particular fluid type in a given OCT volume.

The segmentation CNN was trained using the loss-function given below

$$\mathcal{L} = \mathcal{L}_{cross} + \lambda_d \mathcal{L}_{dice} + \lambda_a \mathcal{L}_{adv} + \lambda_w \left\| w \right\|_2 .$$

The first term, $\mathcal{L}_{cross}$, is the class-balanced cross-entropy loss that normalizes the general cross-entropy to account for class imbalance in each slice. The second term is a smoothed version of dice coefficient between pathologies and normal tissue. The third term, $\mathcal{L}_{adv}$, guides the network towards generating a segmentation mask that can confuse the trained adversarial network. The overall network was trained for 100 epochs using the Adam optimizer.

To minimize the variations between images from different vendors each OCT volume was subjected to a set of pre-processing steps. First, the voxel values were normalized to range $[0, 255]$ followed by histogram matching using a randomly selected OCT scan as template. Next, a median filter along the z-dimension (size 5) is applied and the volumes are re-sampled to give approximately the same voxel spacing across all vendors. Finally, entropy of voxels is used to identify the region of a slice that contain the retinal layers. During the training phase fixed size overlapping sub-volumes ($[256 \times 128 \times 3]$) are extracted from the interesting region identified above and was augmented using random rotations and horizontal flipping. During the test phase, the CNN is applied to each pre-processed OCT sub-volume separately. The resulting segmentations are stacked together and a median filter was applied across B-scans to produce the the overall segmentation. The code is available at https://github.com/RuwanT/retouch.

The detection network takes as input the last layer features of the segmentation network and supplies them to a convolutional and a global average pooling layer followed by 3-dense layers with soft-max activations (one for each fluid type).

F. SFU [76]

The proposed segmentation framework consisted of three steps:

1) Preprocessing: To reduce the effect of speckle, motion-corrected intensity B-scans in each volume were smoothed by 3D Bounded variation (BV). Then, 3D graph-cut based algorithm was applied to segment the internal limiting membrane (ILM) and the Bruch's membrane (BM).

2) Multi-class fluid pixel segmentation: A 2D fully convolutional neural network, which shared a similar structure as the standard U-net, was proposed to segment each pixel into background, IRF, SRF and PED. Because the proposed network determined the class of each pixel solely by the intensities of its neighbors, it can hardly recognize different types of fluids with only raw intensity B-scan as input. Therefore, relative distance map was concatenated to the intensity of each B-scan as the second channel of input image. For a pixel (x, y), its intensity in the relative distance map is defined as: $I(x, y) = \frac{y - Y_1(x)}{Y_1(x) - Y_2(x)}$, where $Y_1(x)$ and $Y_2(x)$ represent ILM and BM, respectively. Softmax cross entropy was applied as loss function with only true positive and false positive pixels used for calculation to avoid class imbalance between background and fluid pixels. To increase the number of training samples, three processes - flip, rotation and zooming - were applied for data augmentation with rotation degree from -25^o to 25^o and maximum zooming ratio 0.5.

3) Post-processing: To prevent over-segmentation, potential fluid pixels with 8-connectivity in each B-scan were considered as a candidate region. For each region, a 16-dimensional shape and intensity feature was extracted and used to train a random forest classifier to rule out false positive regions.

With the fluid segmentation result from previous steps, the presence of fluid in each volume was determined based on the assumption that fluid usually existed within multiple B-scans. The probability of fluid presence in each B-scan was defined as the highest probability of all candidate fluid regions within the scan from random forest classifier, and the probability of a volume was calculated as the mean of the 10 highest probabilities overall all B-scans in this volume.

G. UCF [77]

A deep encoder-decoder ResNet (ED-ResNET) CNN was used for pixelwise segmentation of individual OCT slices. The CNN contains a total of 43 convolutional layers, 32 of which are on the encoder side of the network. The network was trained with cross entropy loss and class imbalance was handled by applying class-specific weights to the loss at each pixel. Specifically, the weight for each class was proportional to the percentage of pixels not belonging to that class.

Pre-processing consisted of smoothing the 3D OCT volumes before slice extraction, resizing, and cropping to the retina area. Post-processing consisted of a graph cut algorithm and some knowledge-guided morphological operations for refinement. In training, a novel data augmentation technique, called myopic warping, was proposed, in which an image of the retina was warped to look more myopic (curvier retina). Myopic warping was used jointly with rotation data augmentation to greatly increase the size of the training set.

Detection probabilities were computed by dividing the total segmented fluid volume by a constant.

H. UMN [78]

The pre-processing step in fluid segmentation involves the segmentation of ILM and RPE layers as a region of interest for the IRF and SRF. For this task, the graph is constructed from each OCT B-scan by mapping each pixel in the image to one node in a graph. Layer segmentation is carried out by graph shortest path method. Weights in graph, computed by the proposed methods, guarantee that pixels located in layer boundaries have the minimum weights and consequently are the best candidates to be selected by graph shortest path methods.

For IRF and SRF segmentation, a supervised method based on CNN is trained in which ROI is present in the training and testing phases. Here, fluid (IRF and SRF) and tissue pixels are labeled as 1 and 0, respectively. Therefore, the CNN is trained for binary classification with a quadratic loss function which minimizes the quadratic sum of distance of each point to the regression line. The procedure to select training pixels for CNN to handle class imbalance is as follows. Consider a fluid pixel in training set; three properties including pixel intensity, average and standard deviation of the intensity of neighboring pixels are considered for this pixel. In the next step, all fluid pixels whose properties are similar to this pixel are removed from the training set. Finally, a tissue pixel set with the same size of fluid pixel set is selected from tissue pixels randomly. This approach decreases the size of the training set significantly since the majority of fluid pixels have the same behavior with respect to the three mentioned properties. CNN is applied to each B-scan separately. For this task, each pixel is represented by a [10,10] window of its neighbors. Then, these windows are input to the CNN. Therefore, each pixel is labeled as fluid or tissue by its 100 neighbors. Architecture of CNN is summarized as follows. First Layer: Convolutional layer with number of feature maps=10, size of kernels=[3,3], activation function=ReLU. Second Layer: Pooling layer with subsample rate=2, subsample method=mean. Third layer: Fully-connected layer with number of nodes=150, activation function=$\tanh$. Fourth layer: Fully-connected layer with number of nodes=2, activation function=$\tanh$. The post-processing step for fluid segmentation is to ignore small fluid regions with sizes under a threshold.

For fluid detection, in each B-scan, probabilities of the existence of IRF, SRF and PED are computed. These probabilities are computed for each OCT volume. Then, the probability of IRF, SRF and PED are computed by thresholding. Therefore, a binary value is assigned for each B-scan which means that this B-scan may or may not contain fluid.

References

[1] L. S. Lim, P. Mitchell, J. M. Seddon, F. G. Holz, and T. Y. Wong, "Age-related macular degeneration," The Lancet, vol. 379, no. 9827, pp. 1728–1738, may 2012.

[2] P. J. Rosenfeld, D. M. Brown, J. S. Heier, D. S. Boyer, P. K. Kaiser, C. Y. Chung, R. Y. Kim, and MARINA Study Group, "Ranibizumab for neovascular age-related macular degeneration." N. Engl. J. Med., vol. 355, no. 14, pp. 1419–31, Oct. 2006.

[3] A. C. Ho, B. G. Busbee, C. D. Regillo, M. R. Wieland, S. A. Van Everen, Z. Li, R. G. Rubio, and P. Lai, "Twenty-four-month efficacy and safety of 0.5 mg or 2.0 mg ranibizumab in patients with subfoveal neovascular age-related macular degeneration." Ophthalmology, vol. 121, no. 11, pp. 2181–92, nov 2014.

[4] R. Silva, A. Berta, M. Larsen, W. Macfadden, C. Feller, J. Monés, and TREND Study Group, "Treat-and-Extend versus Monthly Regimen in Neovascular Age-Related Macular Degeneration," Ophthalmology, vol. 125, no. 1, pp. 57–65, jan 2018.

[5] U. Schmidt-Erfurth and S. M. Waldstein, "A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration." Progr. Retin. Eye. Res., vol. 50, pp. 1–24, Jan. 2016.

[6] U. Schmidt-Erfurth, S. Klimscha, S. M. Waldstein, and H. Bogunović, "A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration," Eye, vol. 31, no. 1, pp. 26–44, Jan. 2017.

[7] E. A. Swanson and J. G. Fujimoto, "The ecosystem that powered the translation of OCT from fundamental research to clinical and commercial impact," Biomed. Opt. Express, vol. 8, no. 3, p. 1638, mar 2017.

[8] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, and C. A. Puliafito, "Optical coherence tomography," Science, vol. 254, no. 5035, pp. 1178–81, Nov. 1991.

[9] C. K. Hitzenberger, "Optical measurement of the axial eye length by laser Doppler interferometry." Invest. Ophthalmol. Vis. Sci., vol. 32, no. 3, pp. 616–24, Mar. 1991.

[10] S. M. Waldstein, C. Simader, G. Staurenghi, N. V. Chong, P. Mitchell, G. J. Jaffe, C. Lu, T. A. Katz, and U. Schmidt-Erfurth, "Morphology and Visual Acuity in Aflibercept and Ranibizumab Therapy for Neovascular Age-Related Macular Degeneration in the VIEW Trials," Ophthalmology, vol. 123, no. 7, pp. 1521–1529, jul 2016.

[11] S. M. Waldstein, A.-M. Philip, R. Leitner, C. Simader, G. Langs, B. S. Gerendas, and U. Schmidt-Erfurth, "Correlation of 3-Dimensionally Quantified Intraretinal and Subretinal Fluid With Visual Acuity in Neovascular Age-Related Macular Degeneration." JAMA Ophthalmol., vol. 134, no. 2, pp. 182–90, Feb. 2016.

[12] M. G. Maguire, D. F. Martin, G.-S. Ying, G. J. Jaffe, E. Daniel, J. E. Grunwald, C. A. Toth, F. L. Ferris, and S. L. Fine, "Five-Year Outcomes with Anti-Vascular Endothelial Growth Factor Treatment of Neovascular Age-Related Macular Degeneration: The Comparison of Age-Related Macular Degeneration Treatments Trials." Ophthalmology, apr 2016.

[13] U. Schmidt-Erfurth, S. M. Waldstein, G. G. Deak, M. Kundi, and C. Simader, "Pigment epithelial detachment followed by retinal cystoid degeneration leads to vision loss in treatment of neovascular age-related macular degeneration," Ophthalmology, vol. 122, no. 4, pp. 822–832, Apr. 2015.

[14] J. J. Arnold, C. M. Markey, N. P. Kurstjens, and R. H. Guymer, "The role of sub-retinal fluid in determining treatment outcomes in patients with neovascular age-related macular degeneration - a phase IV randomised clinical trial with ranibizumab: the FLUID study," BMC Ophthalmology, vol. 16, no. 1, p. 31, dec 2016.

[15] F. M. Penha, G. Gregori, C. A. d. A. Garcia Filho, Z. Yehoshua, W. J. Feuer, and P. J. Rosenfeld, "Quantitative changes in retinal pigment epithelial detachments as a predictor for retreatment with anti-VEGF therapy." Retina, vol. 33, no. 3, pp. 459–66, mar 2013.

[16] E. W. Chan, M. Eldeeb, G. Lingam, D. Thomas, M. Bhargava, and C. K. Chee, "Quantitative Changes in Pigment Epithelial Detachment Area and Volume Predict Retreatment in Polypoidal Choroidal Vasculopathy," Am. J. Ophthalmol., vol. 177, pp. 195–205, May 2017.

[17] M. K. Garvin, M. D. Abramoff, 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. Imag., vol. 28, no. 9, pp. 1436–47, Sep. 2009.

[18] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," IEEE Rev. Biomed. Eng., vol. 3, pp. 169–208, Jan. 2010.

[19] D. C. Fernández, "Delineating fluid-filled region boundaries in optical coherence tomography images of the retina," IEEE Trans. Med. Imag., vol. 24, no. 8, pp. 929–945, Aug. 2005.

[20] Y. Zheng, J. Sahni, C. Campa, A. N. Stangos, A. Raj, and S. P. Harding, "Computerized Assessment of Intraretinal and Subretinal Fluid Regions in Spectral-Domain Optical Coherence Tomography Images of the Retina," Am. J. Ophthalmol., vol. 155, no. 2, pp. 277–286.e1, 2013.

[21] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, and M. Sonka, "Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in SD-OCT of the macula," IEEE Trans. Med. Imag., vol. 29, no. 6, pp. 1321–30, Jun. 2010.

[22] 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 Trans. Med. Imag., vol. 31, no. 8, pp. 1521–31, Aug. 2012.

[23] X. Xu, K. Lee, L. Zhang, M. Sonka, and M. D. Abramoff, "Stratified sampling voxel classification for segmentation of intraretinal and subretinal fluid in longitudinal clinical OCT data," IEEE Trans. Med. Imag., vol. 34, no. 7, pp. 1616–1623, Jul. 2015.

[24] J. Wang, M. Zhang, A. D. Pechauer, L. Liu, T. S. Hwang, D. J. Wilson, D. Li, and Y. Jia, "Automated volumetric segmentation of retinal fluid on optical coherence tomography." Biomed. Opt. Express, vol. 7, no. 4, pp. 1577–89, apr 2016.

[25] A. Rashno, B. Nazari, D. D. Koozekanani, P. M. Drayna, S. Sadri, H. Rabbani, and K. K. Parhi, "Fully-automated segmentation of fluid regions in exudative age-related macular degeneration subjects: Kernel graph cut in neutrosophic domain," PLOS ONE, vol. 12, no. 10, p. e0186949, oct 2017.

[26] A. Rashno, D. D. Koozekanani, P. M. Drayna, B. Nazari, S. Sadri, H. Rabbani, and K. K. Parhi, "Fully automated segmentation of fluid/cyst regions in optical coherence tomography images with diabetic macular edema using neutrosophic sets and graph algorithms," IEEE Trans. Biomed. Eng., vol. 65, no. 5, pp. 989–1001, May 2018.

[27] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," Biomed. Opt. Express, vol. 6, no. 4, pp. 1172–1194, Apr. 2015.

[28] J. Novosel, K. A. Vermeer, J. H. de Jong, Z. Wang, and L. J. van Vliet, "Joint Segmentation of Retinal Layers and Focal Lesions in 3-D OCT Data of Topologically Disrupted Retinas," IEEE Trans. Med. Imag., vol. 36, no. 6, pp. 1276–1286, jun 2017.

[29] A. Montuoro, S. M. Waldstein, B. S. Gerendas, and U. Schmidt-Erfurth, "Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context," Biomed. Opt. Express, vol. 8, no. 3, pp. 182–190, mar 2017.

[30] L. Zhang, M. Sonka, S. Russell, J. Folk, and M. D. Abràmoff, "Quantifying disrupted outer retina-subretinal layer in SD-OCT images in choroidal neovascularization," Invest. Ophthalmol. Vis. Sci., vol. 55, no. 4, pp. 2329–35, Feb. 2014.

[31] F. Shi, X. Chen, H. Zhao, W. Zhu, D. Xiang, E. Gao, M. Sonka, and H. Chen, "Automated 3-D retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments," IEEE Trans. Med. Imag., vol. 34, no. 2, pp. 441–52, Feb. 2015.

[32] Z. Sun, H. Chen, F. Shi, L. Wang, W. Zhu, D. Xiang, C. Yan, L. Li, and X. Chen, "An automated framework for 3D serous pigment epithelium detachment segmentation in SD-OCT images." Sci. Rep., vol. 6, Feb. 2016, article number: 21739.

[33] M. Wu, W. Fan, Q. Chen, Z. Du, X. Li, S. Yuan, and H. Park, "Three-dimensional continuous max flow optimization-based serous retinal detachment segmentation in SD-OCT for central serous chorioretinopathy," Biomed. Opt. Express, vol. 8, no. 9, p. 4257, sep 2017.

[34] M. Wu, Q. Chen, X. He, P. Li, W. Fan, S. Yuan, and H. Park, "Automatic subretinal fluid segmentation of retinal sd-oct images with neurosensory retinal detachment guided by enface fundus imaging," IEEE Trans. Biomed. Eng., vol. 65, no. 1, pp. 87–95, Jan 2018.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25 (NIPS), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., 2012, pp. 1097–1105.

[36] T. Schlegl, S. M. Waldstein, W.-D. Vogl, U. Schmidt-Erfurth, and G. Langs, "Predicting semantic descriptions from medical images with convolutional neural networks," in Proc. Int. Conf. Inform. Process. Med. Imag. (IPMI), ser. Lect. Notes Comput. Sci., vol. 9123, 2015, pp. 437–48.

[37] 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, vol. 8, no. 5, pp. 2732–2744, may 2017.

[38] J. Loo, L. Fang, D. Cunefare, G. J. 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, vol. 9, no. 6, pp. 2681–2698, jun 2018.

[39] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, apr 2017.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Imag. Comput. & Comput. Assist. Interven. (MICCAI), ser. Lect. Notes Comput. Sci., vol. 9351, 2015, pp. 234–241.

[41] C. S. Lee, A. J. Tyring, N. P. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography." Biomed. Opt. Express, vol. 8, no. 7, pp. 3440–3448, jul 2017.

[42] F. G. Venhuizen, B. van Ginneken, B. Liefers, F. van Asten, V. Schreur, S. Fauser, C. Hoyng, T. Theelen, and C. I. Sánchez, "Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography," Biomed. Opt. Express, vol. 9, no. 4, p. 1545, apr 2018.

[43] K. Gopinath and J. Sivaswamy, "Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement," IEEE J. Biomed. Health Inform., 2018, in press.

[44] G. N. Girish, B. Thakur, S. R. Chowdhury, A. R. Kothari, and J. Rajan, "Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model," IEEE J. Biomed. Health Inform., 2018.

[45] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks." Biomed. Opt. Express, vol. 8, no. 8, pp. 3627–3642, aug 2017.

[46] T. Schlegl, S. M. Waldstein, H. Bogunovic, F. Endstraßer, A. Sadeghipour, A.-M. Philip, D. Podkowinski, B. S. Gerendas, G. Langs, and U. Schmidt-Erfurth, "Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning," Ophthalmology, vol. 125, no. 4, pp. 549–558, Apr. 2018.

[47] U. Chakravarthy, D. Goldenberg, G. Young, M. Havilio, O. Rafaeli, G. Benyamini, and A. Loewenstein, "Automated identification of lesion activity in neovascular age-related macular degeneration," Ophthalmology, vol. 123, no. 8, pp. 1731–1736, Aug. 2016.

[48] P. L. Vidal, J. de Moura, J. Novo, M. G. Penedo, and M. Ortega, "Intraretinal fluid identification via enhanced maps using optical coherence tomography images," Biomed. Opt. Express, vol. 9, no. 10, pp. 4730–4754, oct 2018.

[49] J. Wu, A.-M. Philip, D. Podkowinski, B. S. Gerendas, G. Langs, C. Simader, S. M. Waldstein, and U. Schmidt-Erfurth, "Multivendor Spectral-Domain Optical Coherence Tomography Dataset, Observer Annotation Performance Evaluation, and Standardized Evaluation Framework for Intraretinal Cystoid Fluid Segmentation," J. Ophthalmol., 2016, article ID 3898750.

[50] M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Qin Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M. D. Abramoff, "Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs," IEEE Trans. Med. Imag., vol. 29, no. 1, pp. 185–195, Jan. 2010.

[51] "Kaggle, Inc. Diabetic Retinopathy Detection," https://www.kaggle.com/c/diabetic-retinopathy-detection, 2015.

[52] E. Shahrian Varnousfaderani, J. Wu, W.-D. Vogl, A.-M. Philip, A. Montuoro, R. Leitner, C. Simader, S. M. Waldstein, B. S. Gerendas, and U. Schmidt-Erfurth, "A novel benchmark model for intelligent annotation of spectral-domain optical coherence tomography scans using the example of cyst annotation," Comput. Meth. Programs Biomed., vol. 130, pp. 93–105, jul 2016.

[53] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Trans. Med. Imag., vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[54] A. M. Mendrik, K. L. Vincken, H. J. Kuijf, M. Breeuwer, W. H. Bouvy, J. de Bresser, A. Alansary, M. de Bruijne, A. Carass, A. El-Baz, A. Jog, R. Katyal, A. R. Khan, F. van der Lijn, Q. Mahmood, R. Mukherjee, A. van Opbroek, S. Paneri, S. Pereira, M. Persson, M. Rajchl, D. Sarikaya, Ö. Smedby, C. A. Silva, H. A. Vrooman, S. Vyas, C. Wang, L. Zhao, G. J. Biessels, and M. A. Viergever, "MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans," Computational Intelligence and Neuroscience, pp. 1–16, 2015.

[55] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool." BMC Med. Imaging, vol. 15, p. 29, aug 2015.

[56] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, jun 2017.

[57] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, D. Christiaens, F. Dutil, K. Egger, C. Feng, B. Glocker, M. Götz, T. Haeck, H.-L. Halme, M. Havaei, K. M. Iftekharuddin, P.-M. Jodoin, K. Kamnitsas, E. Kellner, A. Korvenoja, H. Larochelle, C. Ledig, J.-H. Lee, F. Maes, Q. Mahmood, K. H. Maier-Hein, R. McKinley, J. Muschelli, C. Pal, L. Pei, J. R. Rangarajan, S. M. Reza, D. Robben, D. Rueckert, E. Salli, P. Suetens, C.-W. Wang, M. Wilms, J. S. Kirschke, U. M. Krämer, T. F. Münte, P. Schramm, R. Wiest, H. Handels, and M. Reyes, "ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI," Med. Image Anal., vol. 35, pp. 250–269, Jan. 2017.

[58] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 1998, pp. 839–846.

[59] A. Chambolle, "An algorithm for total variation minimization and applications," J. Math. Imaging Vis., vol. 20, no. 1-2, pp. 89–97, Jan. 2004.

[60] H. Bogunovic, L. Zhang, M. D. Abramoff, and M. Sonka, "Prediction of treatment response from retinal OCT in patients with exudative age-related macular degeneration," in MICCAI Workshop on Ophthalmic Medical Image Analysis (OMIA), Jan. 2014, pp. 129–136.

[61] W.-D. Vogl, S. M. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth, and G. Langs, "Predicting Macular Edema Recurrence from Spatio-Temporal Signatures in Optical Coherence Tomography Images," IEEE Trans. Med. Imag., pp. 1–1, May 2017.

[62] H. Bogunovic, S. M. Waldstein, T. Schlegl, G. Langs, A. Sadeghipour, X. Liu, B. S. Gerendas, A. Osborne, and U. Schmidt-Erfurth, "Prediction of Anti-VEGF Treatment Requirements in Neovascular AMD Using a Machine Learning Approach," Invest. Ophthalmol. Vis. Sci., vol. 58, no. 7, pp. 3240–3248, jun 2017.

[63] W.-D. Vogl, S. M. Waldstein, B. S. Gerendas, T. Schlegl, G. Langs, and U. Schmidt-Erfurth, "Analyzing and Predicting Visual Acuity Outcomes of Anti-VEGF Therapy by a Longitudinal Mixed Effects Model of Imaging and Clinical Data," Invest. Ophthalmol. Vis. Sci., vol. 58, no. 10, p. 4173, aug 2017.

[64] R. V. Marinescu, N. P. Oxtoby, A. L. Young, E. E. Bron, A. W. Toga, M. W. Weiner, F. Barkhof, N. C. Fox, S. Klein, D. C. Alexander, t. E. Consortium, and f. t. A. D. N. Initiative, "TADPOLE challenge: Prediction of longitudinal evolution in alzheimer's disease," ArXiv, Aug. 2018. [Online]. Available: http://arxiv.org/abs/1805.03909

[65] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018.

[66] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn. (CVPR), 2017.

[67] S. Jégou, M. Drozdzal, D. Vázquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn. (CVPR), 2017.

[68] "RETOUCH - RETinal Oct flUid CHallenge, a satellite event of MICCAI 2017," https://retouch.grand-challenge.org/workshop/, 2017.

[69] S. Yadav, K. Gopinath, and J. Sivaswamy, "A generalized motion pattern and FCN based approach for retinal fluid detection and segmentation," in Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH), Sep. 2017.

[70] S. H. Kang, H. S. Park, J. Jang, and K. Jeon, "Deep neural networks for the detection and segmentation of the retinal fluid in OCT images," in Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH), Sep. 2017.

[71] Q. Chen, Z. Ji, T. Wang, Y. Tand, C. Yu, O. I. Paul, and L. B. Sappa, "Automatic segmentation of fluid-associated abnormalities and pigment epithelial detachment in retinal SD-OCT images," in Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH), 2017.

[72] S. Apostolopoulos, C. Ciller, R. Sznitman, and S. D. Zanet, "Simultaneous classification and segmentation of cysts in retinal OCT," in Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH), Sep. 2017.

[73] S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological OCT retinal layer segmentation using branch residual U-shape networks," in Proc. Int. Conf. Med. Imag. Comput. & Comput. Assist. Interven. (MICCAI), ser. Lect. Notes Comput. Sci., vol. 10435, 2017, pp. 294–301.

[74] S. Apostolopoulos, C. Ciller, S. I. D. Zanet, S. Wolf, and R. Sznitman, "Retinet: Automatic AMD identification in OCT volumetric data," ArXiv, 2016. [Online]. Available: https://arxiv.org/abs/1610.03628

[75] R. Tennakoon, A. K. Gostar, R. Hoseinnezhad, and A. Bab-Hadiashar, "Retinal fluid segmentation and classification in OCT images using adversarial loss based CNN," in Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH), Sep. 2017. [Online]. Available: https://github.com/RuwanT/retouch

[76] D. Lu, M. Heisler, S. Lee, G. W. Ding, V. Vanzan, E. Navajas, M. V. Sarunic, and M. F. Beg, "Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network," Med. Image Anal., 2019, accepted. [Online]. Available: https://arxiv.org/abs/1710.04778

[77] D. Morley, H. Foroosh, S. Shaikh, and U. Bagci, "Simultaneous detection and quantification of retinal fluid with deep learning," in Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH), Sep. 2017.

[78] A. Rashno, D. D. Koozekanani, and K. K. Parhi, "Detection and segmentation of various types of fluids with graph shortest path and deep learning approaches," in Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH), Sep. 2017.