IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 February 2024, accepted 21 February 2024, date of publication 26 February 2024, date of current version 4 March 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3369913*

RESEARCH ARTICLE

# Performance Evaluation of Retinal OCT Fluid Segmentation, Detection, and Generalization Over Variations of Data Sources

**NCHONGMAJE NDIPENOCH, (Student Member, IEEE), ALINA MIRON, (Member, IEEE), AND YONGMIN LI, (Senior Member, IEEE)**
Department of Computer Science, Brunel University London, UB8 3PH Uxbridge, U.K.
Corresponding author: Yongmin Li (yongmin.li@brunel.ac.uk)

**ABSTRACT** Retinal Optical Coherence Tomography (OCT) is a non-invasive cross-sectional scan of the eye that provides qualitative 3D visualization of the retinal anatomy. It is used to study the retinal structure and the presence of pathogens. The advent of retinal OCT has transformed ophthalmology and is currently paramount for the diagnosis, monitoring, and treatment of many eye diseases, including macular edema, which impairs vision severely, and glaucoma, which can cause irreversible blindness. However, the quality of OCT images can vary among device manufacturers. Deep learning methods have been successful in the medical image segmentation community, but it is not yet clear if the level of success can be generalized across images collected from different device vendors. In this study, we provide a comprehensive review of current deep learning segmentation methods applied to OCT images. Furthermore, to investigate the problem of variant of data sources from OCT device vendors, we analyse a selection of the most representative methods to address this problem, including those on the top of the RETOUCH competition such as nnUNet and its variant nnUNet_RASPP, SAM and its variant SAMedOCT, IAUNet_SPP_CL, alongside other state-of-the-art algorithms. The algorithms were validated on the RETOUCH challenge dataset, which was acquired from three device vendors across three medical centers from patients suffering from two retinal disease types. Experimental results show that for several tasks of segmentation, detection and generalisation performance from the retinal images, while fine-tuned large foundation models such as SAMedOCT have demonstrated promising performance, the specifically designed and trained models such as nnUNet and nnUNet_RASPP still offer a slight advantage overall. Also, the nnUNet_RASPP obtained the best performance of 82.3% of mean Dice score for fluid segmentation.

**INDEX TERMS** Medical imaging, segmentation, optical coherence tomography, retinal, convolutional neural network, deep learning, nnUNet, residual connection, Atrous spatial pyramid pooling.

## I. INTRODUCTION

Edema is an eye condition that occurs when there is a leakage of blood vessels into part of the retinal called Macular (central of the eye at the back where the vision is sharpest) and hence impairing the vision severely. There are many eye diseases that can cause this including, age-related macular degeneration (AMD) and diabetic macular edema (DME). Recent study [42] indicates that there's a rise of retinal diseases in Europe with more than 34 and 4 million people affected with AMD and DME respectively in the continent. AMD, is mostly common among older people (50-years and above). The early stage of AMD is asymptomatic and slows to progress to the late stage which is more severe and less common. DME, is the thickening of the retinal caused by the accumulation of intraretinal fluid in the macular and

The associate editor coordinating the review of this manuscript and approving it for publication was Thomas Canhao Xu.

it's mostly common among diabetic patients. Currently there is no cure for these diseases and anti-vascular endothelial growth factor (Anti-VEGF) therapy is the main treatment. This requires constant administering of injections which are expensive and hence a socio-economic burden to most patients and the healthcare system. Therefore early diagnostic and active monitoring the progress of these diseases is vital because the doctors can give some behavioral advice like change of diet or doing regular exercise which will help slow down the progress or in some cases prevents the diseases from getting into a later stage. As of today this is mostly done manually which is laborious, time intensive and prone to error. Therefore an automatic and reliable tool is very crucial in this process and to further exploit the qualitative features of the retinal OCT modality efficiently. Also, the presence of eye motion artifacts in OCT lower the signal-to-noise ratio (SNR) due to speckle noise. To circumvent this problem device manufacturers have to find a balance between achieving high SNR, image resolution and the scanning time. Hence the quality of the images varies among device vendors and hence the need to develop an automate tool with high performance that can generalise across images from all the device vendors.

To address the above issues, in this work we investigate the most representative methods to address this problem, in particular those that have appeared in the leading positions of the RETOUCH challenge [11], on three problems of segmentation, detection and generalisation over multiple data sources.

Our main contributions are as follows: 1) We provide a comprehensive review of representative models addressing the problem of retina fluid segmentation and detection from OCT images, evaluating their generalization performance across variations in data sources. 2) Through the use of a blinded evaluation, we demonstrate that large foundation models, including SAM and its variants SAMed and SAMedOCT, show promising performance in the segmentation task following a routine fine-tuning process for this specific problem. 3)We illustrate that, at the current stage, specifically designed and trained deep networks such as nnUNet and its variant nnUNet_RASPP maintain a slight advantage in the performance of segmentation, detection, and generalization across all tasks for this particular problem.

The rest of the paper is organized as follows. A brief review of the previous studies is provided in Section II. Section III presents the main methods included in this study, mostly from those that have been submitted to the RETOUCH competition [11] and achieved the leading position on the performance table. The experiment with results and visualisation are presented in Section IV. Finally, the conclusion with our contributions is described in Sections V.

## II. BACKGROUND

OCT was first developed in the early 1990s but only became commercially available in 2006 and rapidly became popular due to its high image quality resolution [21]. Before the era of deep learning, research on retinal OCT image analysis have been ongoing for many years and can be grouped into various categories including probabilistic modelling [34], [35], graph-cut [33], [63], [65], [66], Markov Random Fields [64], [75], level set [21], [22]. Some of the comprehensive reviews and studies of retinal OCT images conducted in the past include: [14], [38], [69], [74]

Recently, deep learning models have provided enhanced solutions to address these problems. UNet [60] is one of such models for medical image segmentation. Its architecture has a U-shape and consists of an encoder, bottleneck and a decoder block. It's an end to end framework in which the encoder is use to extract features from the input images/maps, and the decoder is used for pixels localisation. At the end of the decoder path is a classification layer that classifies each pixel into one of the segmented classes. Also, between the encoder and the decoder paths is a bottleneck to ensure the smooth transition from the encoder to the decoder. The encoder, decoder and bottleneck are made up of a series of convolutional layers arranged in a special order.

The DA-PSPNet, introduced in [77], is designed for the automatic layering of retinal OCT images. It utilizes a dual attention mechanism to segment seven retinal layers in retinal OCT B-scan images. The architecture is comprised of a ResNet [29] backbone with dilated convolution to extract shallow features, a self-attention mechanism for capturing contextual information, and residual blocks. Additionally, the authors integrated a pyramid pooling module at varying scales into the network's architecture to gather global information. Evaluation of the algorithm was conducted on the 2014_BOE_Srinivasan dataset [68], published by Duke University. Experimental results demonstrate that the proposed architecture surpasses SOTA algorithms in the segmentation of the seven OCT layers.

The RFS-Net, presented in [27], is composed of a VGG-16 [67] backbone. It incorporates an atrous spatial pyramid pooling (ASPP) module with four parallel branches featuring varying dilating ratings ranging from 1 to 7, aimed at capturing global information. The model addresses the vanishing gradient problem through the use of residual connections. Inspired by GoogLeNet [70], it employs an inception module for dimensionality reduction and includes an expanding module to recover high-level features from the preceding modules. Evaluation of the model was conducted on three datasets: Retouch [11], Gholami [26], and Kermany [37]. Experimental results indicate that the RFS-Net demonstrates performance comparable to SOTA algorithms.

Various deep learning methods for automatic choroidal segmentation in OCT images are outlined in [41]. The employed architectures include the Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and the UNet. The CNN and RNN are categorized as patch-based networks, while the UNet [60] is classified as a semantic segmentation network. The algorithm underwent evaluation on a private dataset comprising spectral domain

OCT (SD-OCT) images collected from 101 children during four different visits over an 18-month period. Experimental results demonstrate that the algorithms achieved performance comparable to SOTA methods.

Retinal fluids segmentation using volumetric deep neural networks on OCT scans is introduced in [7], employing the 3D UNet [60] architecture with five blocks in both the encoder and decoder paths. Each block comprises of two convolution layers with $(3 \times 3 \times 3)$ convolutions, followed by a rectified linear unit (ReLU). Batch normalization precedes the ReLU unit, and a Max-pooling of $(2 \times 2 \times 2)$ is applied. The algorithm underwent validation on the Retouch dataset, achieving a Dice score of 0.79 for intraretinal fluid (IRF), which was the most identified fluid.

An automated quantification of pathological fluids in neovascular Age-Related Macular Degeneration (nAMD) using fully connected neural networks (FCNN) is presented in [48]. The architecture encompasses an encoder path for capturing image features, a decoder path for synthesizing information from the feature maps to make pixel predictions, a classification layer for pixel classification, a residual block to address the vanishing gradient problem, and a squeeze-excite block for capturing global contextual features. The network has a depth of five, corresponding to four downsampling steps, and a kernel size of $7 \times 7$. The evaluation was conducted on a private dataset comprising of two subsets: set 1 includes 107 SD-OCT volumes (49 B-scans, $6 \times 6$ mm cube examination) from patients with nAMD, extracted from the Heidelberg Spectralis device; set 2 was acquired from 42 eyes of 40 patients during routine nAMD check-ups. Experimental results indicate that the algorithm achieved an area under the curve of 0.97, 0.95, and 0.99 for the detection of intraretinal fluid (IRF), subretinal fluid (SRF), and pigment epithelial detachment (PED), respectively, along with corresponding Dice scores of 0.73, 0.67, and 0.82.

The biomarker-infused global-to-local network (Bio-Net) is presented in [81] for the automatic segmentation and visualization of the choroid in OCT through knowledge-infused deep learning. Bio-Net comprises a shadow localization and an elimination pipeline. The authors employed the UNet for generating the shadow location mask and the Deshadow-Net for eliminating shadows. The architecture of Deshadow-Net draws inspiration from the two-stage generative adversarial network (GAN) presented in [54]. Initially, the region of interest (ROI), retinal pigment epithelium layer, and choroid regions are extracted. The UNet is utilized for ROI segmentation, and the Deshadow-Net is employed to eliminate unwanted pixels or shadows. The algorithm underwent evaluation on a private dataset consisting of 30 OCT volumes from 30 subjects. Experimental results demonstrate that the model surpassed current state-of-the-art algorithms by a significant margin.

An ensemble learning approach named DelNet, utilizing a combination of Convolutional Neural Network (CNN) models based on fully convolutional networks (FCN), is introduced in [8] for the segmentation of retinal layers in OCT images. The architecture of DelNet draws inspiration from ensemble stack generalization [76] in deep learning methods. The authors aggregate outputs from a series of FCNs (base networks) into a single output for image segmentation, aiming to construct a robust model by combining knowledge from weak learners. This approach is employed to address the high variability among retinal layers. The method underwent validation on the Duke Cyst DME dataset [18], and experimental results indicate that the algorithm performs is comparable to SOTA algorithms.

An automated approach for fluid segmentation in retinal OCT images utilizing attention mechanisms is introduced in [45]. The method comprises of an encoding path and two decoding paths. The encoder is responsible for feature extraction, the first decoder extracts semantic information, and the other decoder predicts distance maps. Both the encoder and decoder paths consist of three blocks, each containing a convolutional layer with a padding of 1, a batch normalization layer, and a ReLU layer. In the encoding path, a max-pooling layer with strides of 2 is employed for down-sampling. For image reconstruction in the decoder path, skip connections, similar to those used in [44], are employed. Attention mechanisms inspired by [10] and [32] are incorporated into the network's architecture. The method underwent evaluation on the Multivendor dataset [8], comprising 500 scans from four OCT scan devices (Cirrus, Nidek, Spectralis, Topcon), with the respective number of scans being 57, 159, 53, and 231. Experimental results demonstrate that the proposed approach significantly outperformed the baseline methods.

The cascade dual-branch deep neural networks for retinal layer and fluid segmentation in OCT, incorporating a relative positional map, is introduced in [47]. The method comprises two UNet architectures stacked in series. The first UNet is dedicated to segmenting the Region of Interest (ROI), defined as the area between the inner limiting membrane (ILM) and the Bruch membrane (BM). The second UNet is employed for segmenting six retinal layers and fluid within the ROI. A Random Forest classifier is used at the classification layer for pixel classification. Both the encoder and decoder paths consist of four convolutional blocks with feature sizes of 64, 128, 256, and 512. The kernel size is set to $3 \times 3$, and a $2 \times 2$ max-pooling layer with stride of 2 is utilized. The method underwent validation on a private dataset comprising 58 3D volumes acquired using the Zeiss Cirrus 5000 HD-OCT. Each volume consists of 245 B-scans, resulting in a total of 14,210 B-scans. Experimental results demonstrate that the model's performance is comparable to SOTA algorithms.

The DeepRetina for layer segmentation in OCT images is presented in [43]. The method consists of three main components: Xception65 [19] which serves as the backbone to extract and learn the characteristics of retinal layers,

the ASPP module, with a dilating rate of 2, 6, 12, 18, to capture global information, and an Encoder-Decoder module [9], [17] that further enhances the segmentation map by eliminating misclassified pixels. The algorithm is evaluated on two datasets. The first dataset is a private dataset from the Shenzhen Eye Hospital affiliated with Jinan University and the Shenzhen University School of Medicine, consisting of 280 volumes, resulting in a total of 11,200 images (40 B-scans per volume). The second dataset is the Duke dataset with 110 OCT B-scan images spanning 10 different patients. Experimental results demonstrate that the method's performance is superior to the baseline models.

The Deep-ResUNet++ is presented in [55] for simultaneous segmentation of layers and fluids in retinal OCT images. The approached incorporated residual connections, ASPP blocks and Squeeze and Exciting blocks into the traditional 2D UNet [60] architecture to simultaneously segment 3 retinal layers, 3 fluids and 2 background classes from 1136 B-Scans from 24 patients suffering from wet AMD. The algorithm is validated on the Annotated Retinal OCT Images (AROI) [51] which is publicly available.

The CoNet (Coherent Network) is presented in [56] for the simultaneous segmentation of 7 layers, 2 backgrounds and 1 fluid in retinal OCT images using the Duke DME dataset [18] obtaining a mean Dice Score of 88%. The CoNet uses standard UNet as the backbone with a reduced depth and incorporates an atrous spatial pyramid pooling (ASPP) block at the input layer to capture global contextual features. The ASSP block uses 4 parallel filters with different frequencies or dilating rates.

A clinical application for diagnosis and referral of retinal diseases is proposed in [25] in which 14,884 OCT B-Scans collected from 7,621 patients are trained on a framework consisting of two main parts: The segmentation model (3D Unet [20]) and the classification model.

A combination of Convolutional Neural Networks (CNN) and Graph Search (GS) method is presented in [24]. The framework aims to validate nine layers boundaries from 60 retinal OCT volumes (2915 B-scans, from 20 human eyes) obtained from patients suffering from dry AMD. CNN is used for the extraction of the layer boundaries features while the GS is used for the pixels classification.

In [57] another Deep Learning approach for retinal OCT segmentation combining a FCN for segmentation with Gaussian Processes for post processing is proposed. The method is validated on the University of Miami dataset [73] which consists of 50 volumes from 10 patients suffering from diabetic retinopathy. Their approach is divided into two main steps which are the pixel classification using the FCN and the post processing using Gaussian Processes.

Another CNN-based approach for the simultaneous segmentation of layers and fluid is presented in [61]. They presented a 2D UNet like architecture with a reduced depth for the segmentation of 10 classes consisting of 8 layers, 1 background and 1 fluid from 10 patients suffering from Diabetic Macular Edema (DME). The Duke DME dataset [18] is used to validate the algorithm.

The nnUNet, a self-configuring framework is introduced in [31]. The framework aims to eliminate the problem of manual parameters setting ''trial and error'' by using the dataset's demographic features to determine and automatically set some of the model's key parameters like the batch size. The framework uses the standard UNet [60] and is evaluated on 11 biomedical image segmentation challenges consisting of 23 datasets for 53 segmentation tasks.

An extended version of the nnUNet [31] is presented by McConnell et al in [49] by integrating residual, dense, and inception blocks into the network for the segmentation of medical imaging on multiple datasets. The algorithm is evaluated on eight datasets consisting of 20 target anatomical structures.

ScSE nnU-Net, another extended version of the nnUNet [31] is presented in [78] for the segmentation of head and neck cancer tumors. It extends the original nnUNet by incorporating spatial channels with squeeze and excitation blocks into the network's architecture. The algorithm uses nnUNet to extract features from the input images/maps and then the squeeze and excitation blocks to further suppress the weaker pixels. The method was validated on the HECKTOR 2020 dataset consisting of 201 cases and a test set of 53 cases.

The exploration of advanced architectural variations of the nnUNet is detailed in [50] by McConnell et al. They investigated eight advanced variants of the nnUNet, evaluating the methods across eight datasets encompassing 20 anatomical structures. The architectural variations include residual, dense, inception, and attention gates, resulting in six novel nnUNet variations: Residual-nnUNet, Dense-nnUNet, Inception-nnUNet, Spatial-Single-Attention-nnUNet, Spatial-Multi-Attention-nnUNet, and Channel-Spatial-Attention-nnUNet. Their findings indicate that no single architecture universally performs best for all medical image segmentation problems. Therefore, the selection of a network's architecture should be tailored to the specific problem at hand. Additionally, the study concluded that the Standard-nnUNet and Baseline-nnUNet are optimal for problems or datasets featuring a singular anatomical region, while the other architectures excel in scenarios involving spatially imbalanced datasets or problems with multiple anatomical regions.

The Segment Anything Model (SAM) is introduced in [39] by Kirillov et al., drawing inspiration from the concept of zero-shot and few-shot generalization in natural language processing (NLP). Developed by researchers at Meta, SAM is positioned as a foundational model for image segmentation. It undergoes training on an extensive dataset consisting of 1 billion masks and 11 million images, utilizing the Vision Transformer (ViT) architecture [23]. Researchers and developers have the flexibility to fine-tune the SAM model

and leverage the pre-trained model for specific segmentation tasks during training.

The MA-SAM, an adaptation of SAM for volumetric images, is introduced in [15]. Originally, SAM was developed and trained on 2D images. The authors of MA-SAM incorporate critical third-dimensional or temporal knowledge during fine-tuning to effectively adapt SAM and leverage the volumetric structure inherent in medical imaging. This adaptation involves integrating a series of 3D adapters into the 2D transformer blocks within the SAM architecture. These adapters are employed to extract essential volumetric or temporal insights necessary for medical image analysis. The algorithm's performance is evaluated on four medical image segmentation tasks across 10 public datasets, encompassing CT, MRI, and surgical video data.

The nnFormer (notanother transFormer), a combination of the transformer and nnUNet, is introduced in [83]. The authors integrated the transformer architecture into the nnUNet pipeline, capitalizing on pre-processing and self-parametrization techniques. The network's architecture comprises of an encoder, decoder, a bottleneck in-between, and skip attention blocks at every convolutional layer. The algorithm is evaluated on three public medical datasets: BraTS [52], Synapse [1], ACDC [2]. The authors drew the following conclusions: 1) nnFormer outperformed other transformer-based architectures significantly, and 2) nnFormer and nnUNet demonstrate a high level of complementarity.

The DconnNet [80], which employs a directional connectivity modeling scheme for segmentation, was assessed on the training (subset) dataset of the RETOUCH Grand Challenge using 3-fold cross-validation, achieving a dice score of 87.7.

### A. LIMITATIONS AND BENEFITS

While deep learning methods have demonstrated enhanced performance in medical image segmentation, detection, and classification tasks, their reliance on large datasets poses a significant challenge. The limited availability of public data, driven by privacy concerns, exacerbates this issue. Additionally, the manual segmentation/annotation of datasets is labor-intensive and time-consuming. Furthermore, the Retouch challenge is not publicly available, and each participant team can only submit a maximum of three models for evaluation, thus limiting our ability to assess other SOTA algorithms like [4], [5], [6], and [59]. Moreover, a notable challenge in the field is the lack of consistent evaluation metrics. While some authors employ metrics like Dice Similarity (DS) and Intersection over Union (IoU), others use alternative measures such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and mean unsigned error. On the positive side, a significant advantage lies in the ability to fine-tune large foundation models with limited computational resources, thereby enhancing the generalization performance of the models. Also, another benefit is the availability of several annotated small datasets from different anatomic regions of the body.

## III. METHODS

The RETOUCH grand challenge [11] is a competition focused on the segmentation and detection of three retinal fluids from retinal OCT images. The training dataset comprises 70 raw images with corresponding masks, while the testing dataset includes 40 raw images without their corresponding masks. To ensure fairness in comparison, the organizers employed a blinded evaluation by retaining the masks or ground truth of the testing dataset, and participants can submit their predictions via email for evaluation. In adherence to competition requirements, each submission must be accompanied by a written paper explaining the methods employed. The results of the submission are communicated to the teams via email and are also published on the organizer's website alongside the accompanying papers. The RETOUCH challenge, initially organized in conjunction with MICCAI 2017 in Quebec, Canada, featured the participation of eight teams. Subsequently, the competition transitioned to an online format, and it remains ongoing, continuing to accept submissions [3].

In this section we provide details of the methods that are in the leading positions from the RETOUCH competition, including the nnUNet, nnUnet_RASPP, SAMedOCT, IAUNet_SPP_CL, as well as the standard UNet and other methods.

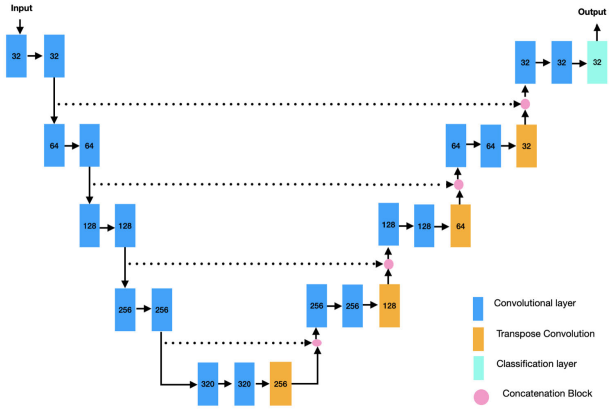


**FIGURE 1. An illustration of the standard UNet architecture used in nnUNet.**

### A. UNet

The UNet is an end to end architecture for medical image segmentation. It consists of 3 main parts: the encoder, the decoder and bottleneck between the encoder and decoder. The encoder captures contextual information (or features extraction) and reduces the size of the feature map by half after every convolutional block as we move down the encoding path by implying strided convolutions. Pixels localisation is done at the decoder. As we move up the decoder path the size of the feature map is doubled

**TABLE 1. A summary of previous work including the authors, year, segmentation, dataset and evualation matrics: Dice Score (DS), Mean average error (MAE), Mean Average Difference (MAD) receiver operating characteristic curve (ROC), Root Mean Squared Error (RMSE), and Intersection over Union(IoU).**

| Author | Year | Segmentation | Dataset | Modality | Task | Evaluation Metric |
|---|---|---|---|---|---|---|
| [65] | 2010 | Choroid | Private | OCT | Segmentation | DS / MAE |
| [66] | 2011 | Fluid | DIARETDB1/DRIVE | Fundus | Segmentation | MAD |
| [64] | 2012 | Fluid | Kermany_S_Dataset | Fundus | Classification | ROC |
| [34] | 2013 | Blood vessels | STARE | Fundus | Segmentation | Accuracy |
| [35] | 2014 | Blood vessels | STARE/DRIVE | Fundus | Segmentation | Accuracy |
| [63] | 2014 | Blood vessels | STARE/DRIVE | Fundus | Segmentation | Accuracy |
| [33] | 2015 | Layers | Private | OCT | Segmentation | DS / RMSE |
| [75] | 2017 | Choroid | Private | OCT | Segmentation | DS |
| [24] | 2017 | Layers | Private | OCT | Segmentation | Mean Difference |
| [61] | 2017 | Fluid / Layers | Duke DME | OCT | Segmentation | DS |
| [46] | 2017 | Fluid | Retouch | OCT | Segmentation | DS / AVD |
| [58] | 2017 | Fluid | Retouch | OCT | Segmentation | DS / AVD |
| [36] | 2017 | Fluid | Retouch | OCT | Segmentation | DS / AVD |
| [62] | 2017 | Fluid | Retouch | OCT | Segmentation | DS / AVD |
| [72] | 2018 | Fluid | Retouch | OCT | Segmentation | DS / AVD |
| [65] | 2018 | Fluid | Private | OCT | Classification | ROC |
| [25] | 2019 | Layers | Private | OCT | Segmentation | DS / RMSE |
| [21] | 2019 | Layers | Private | OCT | Segmentation | DS / RMSE |
| [22] | 2019 | Layers | Private | OCT | Segmentation | DS / RMSE |
| [41] | 2019 | Choroid | Private | OCT | Segmentation | Dice/MAE |
| [57] | 2019 | Fluid / Layers | Uni. of Miami | OCT | Segmentation | Mean Unsigned Error |
| [7] | 2020 | Fluid | RETOUCH | OCT | Segmentation | DS |
| [81] | 2020 | Layers | Private | OCT | Segmentation | DS / IoU |
| [8] | 2020 | Layers | Duke Cyst DME | OCT | Segmentation | DS |
| [47] | 2020 | Layers | Private | OCT | Segmentation | DS |
| [43] | 2020 | Layers | Private / Duke Cyst DME | OCT | Segmentation | DS |
| [77] | 2021 | Layers | Duke DME/AMD | OCT | Segmentation | IoU |
| [27] | 2021 | Fluid | RETOUCH / OCTID | OCT | Segmentation | IoU / DS |
| [48] | 2021 | Fluid | Private | OCT | Detection / Segmentation | DS |
| [45] | 2021 | Fluid | Multivendor | OCT | Segmentation | DS |
| [31] | 2021 | Multiple | Multiple | Multiple | Segmentation | DS |
| [78] | 2021 | Tumour | Hecktor | PT/ CT | Segmentation | DS |
| [83] | 2021 | Multiple | BraTS / Synapse / ACDC | MRI / CT | Segmentation | DS |
| [49] | 2022 | Multiple | Multiple | Multiple | Segmentation | DS |
| [55] | 2022 | Layers / Fluid | AROI | OCT | Segmentation | DS |
| [79] | 2022 | Fluid | Retouch | OCT | Segmentation | DS /AVD |
| [40] | 2022 | Fluid | Retouch/Private | OCT | Segmentation | DS |
| [55] | 2023 | Layers | Duke DME | OCT | Segmentation | DS |
| [50] | 2023 | Multiple | Multiple | Multiple | Segmentation | DS |
| [15] | 2023 | Multiple | Multiple | MRI / CT | Segmentation | DS |
| [39] | 2023 | Multiple | Multiple | Multiple | Segmentation | mIoU |
| [39] | 2023 | Fluid | Retouch | OCT | Segmentation | DS |
| [12] | 2023 | Fluid | Retouch | OCT | Segmentation | DS /AVD |
| [82] | 2023 | Multi-Organs | Synapse | CT | Segmentation | DS |

after every convolutional block by implying transposed convolutions, and for the reconstruction process features maps are concatenated to the corresponding map in the encoder path using up-sampling operations. The bottleneck serves as a bridge, linking the encoding and decoding paths together. It consists of a convolutional block that ensures a smooth transition from the encoder path to the decoder path. At the encoding path, decoding path and bridge layer each convolutional block consists of a convolutional layer that converts the pixels of the receptive field into a single value before passing it to the next operation followed by an instance normalisation to prevent over-fitting during training. This is followed by a LeakyReLU activation function to diminish vanishing gradient. A high level diagram to illustrate the architectural structure of the standard UNet is shown in FIGURE 1.

### B. LOWERCASENNUNLOWERCASEET AND LOWERCASENNUNLOWERCASEET_RASPP

The nnUNet is a self-configuring and automatic pipeline for medical image segmentation with the ability to automatically determine and choose the best model hyper-parameters given the data and the hardware availability, thus alleviating the problem of trial and error of manual parameters setting. Given a training data the framework extracts the ‘‘data-fingerprint‘‘ such as modality, shape, and spacing and base on the hardware (GPU memory) constraints the network topology, image re-sampling methods, and input-image patch sizes are determined. After training is complete, the framework determines if post-processing is needed. The framework uses the standard UNet as the network's architecture. Please refer to the original publication [31] for more information.

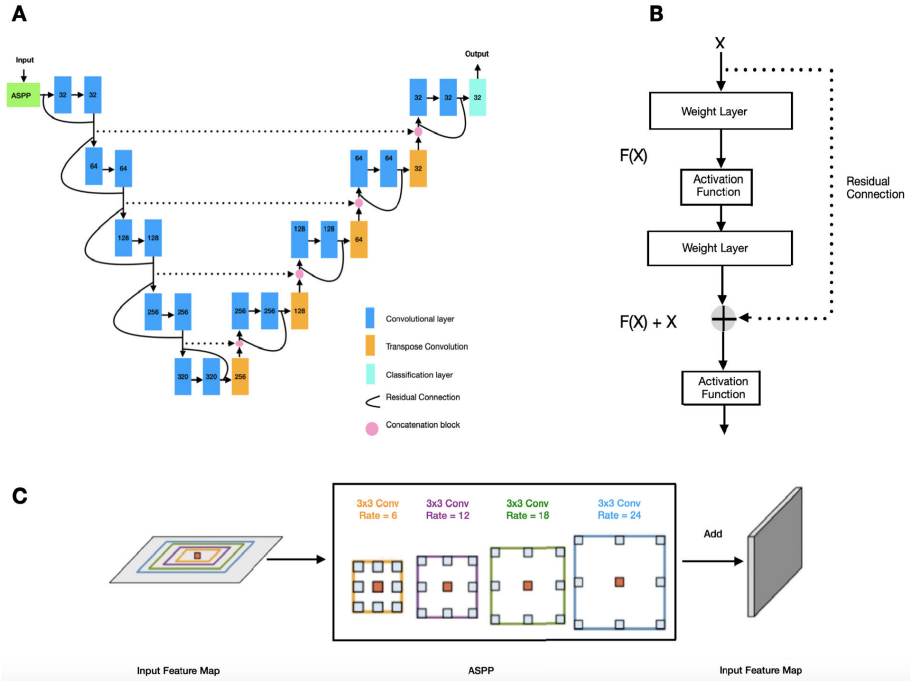


**FIGURE 2.** A high level illustration of nnUNet_RASPP architecture with B, a residual connection block to address the vanishing gradient problem where X is an input and F(X) is a function of X; and C, an ASPP block of multiple parallel filters at different dilating rates or frequencies to capture global information.

Inspired by the success of nnUNet we have introduced an enhanced architecture nnUNet_RASPP.[1,2] by incorporating residual connections and an ASPP block in the network's architecture to solve the problem of data source variation. Residual connections were incorporated in every convolutional layer at both the encoding and decoding paths to reduce the training error rate, further learn complex features, and combat the problem of vanishing gradients. The ASPP was incorporated at the input layer of the encoding path to mitigate the problem of fluid variance. Incorporating these techniques into the standard nnUNet improves the overall performance of the network. The diagram of nnUNet_RASPP is demonstrated in FIGURE 2.A.

Residual connection is a technique used to combat the problem of vanishing gradient developed in [28]. The UNet architecture uses the chain rule for back propagation during training. This process can sometimes lead to vanishing gradient and one of the ways to circumvent this, is to introduce residual connection into the network's architecture. As indicated in [28], residual connections reduce the training error rate as we increase the depth of the network. nnUNet automatically determines the depth of the network, and the introduction of residual connections to the network's architecture further reduces the training error rate and allows the network to learn complex features, thus improving the overall performance. The diagram of residual connection is demonstrated in FIGURE 2.B.

ASPP [16] is a technique used to extract or capture global contextual features by applying paralleling filters with different frequencies or dilating rate for a given input filter. ASPP can be used to solve problems with high variability. To the best of our knowledge, this is the first time that ASPP has been integrated into nnUNet to solve this particular problem. The diagram of the ASPP block is demonstrated in FIGURE 2.C.

### C. SAMeDOCT

The SAMedOCT [12] is inspired and adpated from the Segment Anything Model (SAM) [39]. It is a " foundation model " for image segmentation developed by researchers at Meta. SAM gained prominence for its ability to enable zero-shot transfer to various segmentation tasks, having been trained on over 1 billion masks from 11 million diverse images. Due to its extensive training dataset, SAM demonstrates the capability to generalize to new tasks beyond those encountered during training. SAM comprises of three main components:

1) An Image Encoder built from the Vision Transformer (ViT) [23], which preprocesses high-resolution inputs and runs once per image;
2) A Prompt Encoder embedding dense prompts (i.e., masks) using convolutions, which are then summed element-wise with the image embedding;

[1]The RETOUCH online competition is still ongoing. At the time of writing, our nnUNet_RASPP is currently ranked first among 216 participants from both online and offline submissions. Details of the competition, including the leaders table, number of participants and other statistics, are availabe at: https://retouch.grand-challenge.org/statistics/

[2]We have made the source code of the implementation publicly available at https://github.com/ndipenoch/nnUNetRASPP.git.

3) A Mask Decoder that efficiently maps the image embedding, prompt embeddings, and an output token to a mask.

Focal and dice loss are employed during training. SAMed, a variant of SAM adapted for medical segmentation, is introduced in [82]. SAMed is derived from SAM by freezing the image encoder and adopting a low-rank-based fine-tuning strategy (LoRA) [30]. This strategy approximates the low-rank update of the parameters in the image encoder and fine-tunes the lightweight prompt encoder and the mask decoder of SAM. SAMed was evaluated on the Synapse multi-organ segmentation dataset, achieving remarkable results. Building on the success of SAMed, SAMedOCT was adapted from SAMed to address the challenges posed by the RETOUCH grand challenge.

### D. IAUNet_SPP_CL

IAUNet_SPP_CL, a combination of a graph-theoretic method, a fully convolutional neural network (FCN), and curvature regularization loss function is presented in [79]. The graph-theoretic method is employed in the preprocessing stage to delineate layers and regions of interest (ROI), while the FCN is utilized for fluid segmentation, employing the standard attention UNet as the backbone. The authors enhanced the architecture by introducing spatial pyramid pooling (SPP) modules with four pooling maps at different scales in parallel, concatenating the original input after bilinear interpolation to enhance the network's capability to segment multi-scale objects. The curvature regularization loss function is applied to smooth boundaries and eliminate unnecessary holes within the predicted fluid lesions.

### E. SFU

The SFU, a 3-part CNN-based and Random Forest (RF) framework is developed by [46]. The first part of the framework is used for pre-processing of the images, the second part consists of a 2D UNet architecture for the extraction of features and a RF classifier to classify the pixels at the third part. At the segmentation layer, axial motion between scans was corrected using cross-correlation by applying bounded variation 3D smoothing. This correction aimed to reduce the effect of speckle while preserving and enhancing the boundaries between retinal layers. To prevent overfitting during training, a dropout layer was introduced before the 1 to 1 convolutional layer. Additionally, to address data limitations, data augmentation techniques such as flipping, rotation, and zooming were applied during preprocessing.

### F. UMN

The UMN, a combination of CNN and graph-shortest path (GSP) method is presented in [58]. The CNN is used for the segmentation of region of interest (ROI) and the GSP is further used for the segmentation of the layers and fluid from the ROI. B-scans were extracted from the 3D volumes for training. At the segmentation layer, the initial step involved segmenting the layers as ROI to efficiently detect the presence of fluids. Extracting the ROI helped reduce training time, as training the network on the entire image would be more time-consuming. The GSP was employed for pixel classification, mapping each pixel in the image to one node in the graph. Only local relationships between pixels were considered, and an 8-regular graph was constructed using the 8 neighbors of each pixel.

### G. MABIC

The MABIC, a standard double-UNet architecture, is proposed in [36]. The method utilizes two UNet architectures connected in series, where the output of the first UNet serves as an input to the second UNet. The initial part takes raw images as input to extract the ROI. Additionally, in this initial part, dropout and maxout activation are applied at each layer to enhance accuracy and prevent overfitting. The subsequent part takes the extracted ROI and the segmentation mask as input. Importantly, there are no fully connected layers between encoding and decoding layers in the latter part.

### H. RMIT

The RMIT, an approach using a combination of deep neural network and adversarial loss function is presented in [72]. The authors adapted the architecture from the standard UNet by incorporating a batch normalization layer in each block of convolutions. They introduced dropout at each skip connection to prevent overfitting and incorporated an adversarial loss function to estimate the loss during training.

### I. REtinAI

The RetinAI, introduced in [62], is a standard 2D UNet with residual connections. The network was trained on B-scans. As part of the preprocessing, all the B-scans were normalized to the same resolutions, and horizontal flip, shear, rotation, shift, and Gaussian noise were applied for data augmentation. Categorical cross-entropy was used as the loss function during training.

### J. SVDNA

A noise adaptation approach based on singular value decomposition (SVDNA) [40] is introduced as an unsupervised technique for noise transfer in the domain adaptation of retinal OCT images. The pipeline comprises of two phases. In the first phase, SVDNA is employed to generate masks, which are subsequently used to train a supervised segmentation network in the second phase. The model's performance was evaluated online, achieving a mean DS of 0.71 on the hidden test dataset. The authors didn't publish the AVD scores.

## IV. EXPERIMENTS AND PERFORMANCE EVALUATION

### A. DATASET

The methods were validated on the MICCAI 2017 RETOUCH grande challenge dataset [11]. The dataset is publicly available and it consists of 112 OCT volumes of patients suffering with early AMD and DME collected from 3 device manufacturers: Cirrus, Spectralis and Topcon from 3 clinical centers: Medical University of Vienna (MUV) in Austria, Erasmus University Medical Center (ERASMUS) and Radboud University Medical Center (RUNMC) in the Netherlands. Examples of the dataset are shown in FIGURE 3.

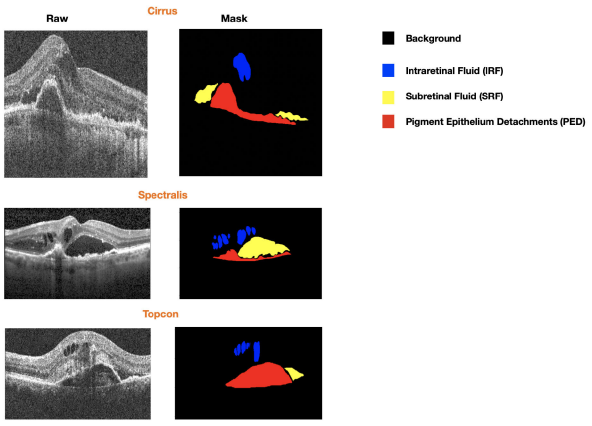


**FIGURE 3.** B-scan examples of raw (column 1) and their corresponded annotated mask (column 2) of OCT volumes taken from the 3 device vendors (rows): Cirrus, Spectralis and Topcon. The classes are coloured as follows: Black for the background, blue for the Intraretinal Fluid (IRF), yelow for the Subretinal Fluid (SRF) and red for the Pigment Epithelium Detachments (PED).

The dimensions of the OCT volumes per vendor machine are as follows: Each volume of the Cirrus consists 128 B-Scans of 512 × 1024 pixels, Spectralis consists of 49 B-scans of 512 × 496 pixels and 128 B-Scans of 512 × 885 (T-2000) or 512 × 650 (T-1000) pixels for Topcon.

The training set consists of 70 volumes of 24, 24, and 22 acquired with Cirrus, Spectralis, and Topcon, respectively. Both the raw and annotated mask of the training set are made available to the public. The testing set consists of 42 OCT volumes of 14 volumes per device vendor. The raw or input of the testing set is available publicly but their corresponding annotated masks are held by the organizers of the challenge. Submission and evaluation of prediction on the testing dataset is arranged privately with the organizers and the results are sent to the participants.

Manual annotation was done by 6 grader experts from 2 medical centers: MUV (4 graders supervised by an ophthalmology resident), and RUNMC (2 graders supervised by a retinal specialist). The dataset is annotated for 4 classes of 1 background labelled as 0 and 3 fluids which are: Intraretinal Fluid (IRF) labelled as 1, Subretinal Fluid (SRF) labeled as 2 and Pigment Epithelium Detachments (PED) labelled as 3.

**TABLE 2.** Segmentation table of the Dice Scores (DS) by segment classes (columns) and teams (rows) for training on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set.

| Methods | IRF | SRF | PED | Mean |
|---|---|---|---|---|
| nnUNet_RASPP | 0.84 | **0.80** | **0.83** | **0.823** |
| nnUNet | **0.85** | 0.78 | 0.82 | 0.817 |
| SFU | 0.81 | 0.75 | 0.74 | 0.78 |
| SAMedOCT [12] | 0.77 | 0.76 | 0.82 | 0.78 |
| IAUNet_SPP_CL [79] | 0.79 | 0.74 | 0.77 | 0.77 |
| UMN | 0.69 | 0.70 | 0.77 | 0.72 |
| MABIC | 0.77 | 0.66 | 0.71 | 0.71 |
| SVDNA [40] | 0.80 | 0.61 | 0.72 | 0.71 |
| RMIT | 0.72 | 0.70 | 0.69 | 0.70 |
| RetinAI | 0.73 | 0.67 | 0.71 | 0.70 |
| Helios | 0.62 | 0.67 | 0.66 | 0.65 |
| NJUST | 0.56 | 0.53 | 0.64 | 0.58 |
| UCF | 0.49 | 0.54 | 0.63 | 0.55 |

The RETOUCH dataset is particularly interesting because of its high level of variability. It was collected using multiple device vendors, the sizes and number of B-Scans varies per device vendor, and it was collected and annotated in multiple clinical centers. Also, for fair comparison the annotated testing set is held by the organizers and submission is curbed to a maximum of 3 per participating team.

### B. TRAINING AND TESTING

Training was done on the 70 OCT volumes of the training set (both raw and mask volumes). The estimated probabilities and predicted segmentation of the testing set (42 raw volumes) were submitted to the challenge organizers for blinded evaluation on the ground truth or masks.

Also, to further evaluate the robustness and generalisability of the methods, the predicted segmentation of the algorithm was evaluated on OCT volumes from two vendor devices and tested on the third. In this case OCT volumes from the third vendor device weren't seen during training. For this experiment, two sets of weights were generated which are: (1) Training on 46 OCT volumes from both Spectralis (24 OCT volumes) and Topcon (22 OCT volumes) and evaluated on 14 OCT volumes from the Cirrus testing set and (2) training on 48 OCT volumes from both Cirrus (24 OCT volumes) and Spectralis (24 OCT volumes) and evaluated on 14 OCT volumes from the Topcon testing set. Again the same environmental settings were used to conduct all the experiments.

In the detection task the estimated probabilities of presence of each fluid type is plotted using the receiver operating characteristics (ROC) curve. The area under the curve (AUC) which measures the ability of a binary classifier to distinguish between classes is used as the evaluation matrice. The AUC gives a score between 0 and 1 with 1 being the perfect score and 0 is the worst. For the segmentation task, two evaluation matrices are used to measure the performance of the algorithms:

1) The Dice Score (DS) [13], [53], [71] which is twice the intersection, divided by the union. It measures the

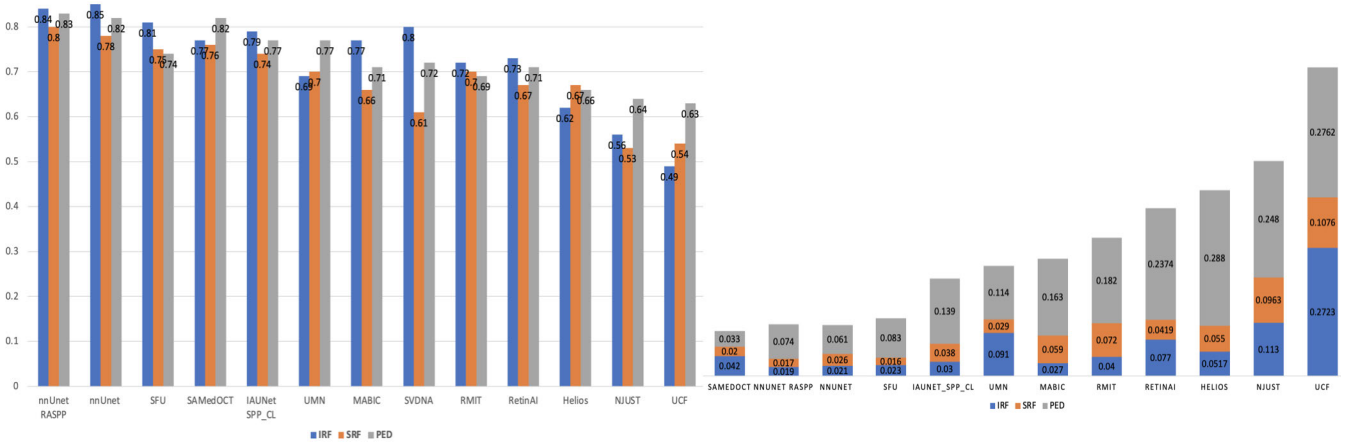


**FIGURE 4.** Segmentation performance comparison by DS on the right and AVD on the left of the nnUNet_RASPP and baseline nnUNet, together with the SOTA algorithms grouped by the segment classes when trained on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set.

**TABLE 3.** Segmentation table of the Absolute Volume Difference (AVD) by segment classes (columns) and teams (rows) for training on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set.

| Methods | IRF | SRF | PED | Mean |
|---|---|---|---|---|
| SAMedOCT [12] | 0.042 | 0.020 | **0.033** | **0.032** |
| nnUNet | **0.019** | 0.017 | 0.074 | 0.036 |
| IAUNet_SPP_CL [79] | 0.021 | 0.026 | 0.061 | 0.036 |
| nnUNet_RASPP | 0.023 | **0.016** | 0.083 | 0.041 |
| SFU | 0.030 | 0.038 | 0.139 | 0.069 |
| UMN | 0.091 | 0.029 | 0.114 | 0.078 |
| MABIC | 0.027 | 0.059 | 0.163 | 0.083 |
| RMIT | 0.040 | 0.072 | 0.1820 | 0.098 |
| RetinAI | 0.077 | 0.0419 | 0.2374 | 0.118 |
| Helios | 0.0517 | 0.055 | 0.288 | 0.132 |
| NJUST | 0.1130 | 0.0963 | 0.248 | 0.153 |
| UCF | 0.2723 | 0.1076 | 0.2762 | 0.219 |

**TABLE 4.** Detection table of the Area Under the Curve (AUC) by segment classes (columns) and teams (rows) for training on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set.

| Methods | IRF | SRF | PED | Mean |
|---|---|---|---|---|
| nnUNet | **1.0** | **1.0** | **1.0** | **1.0** |
| SFU | **1.0** | **1.0** | **1.0** | **1.0** |
| nnUNet_RASPP | 0.93 | 0.97 | **1.0** | 0.97 |
| Helios | 0.93 | **1.0** | 0.97 | 0.97 |
| UCF | 0.94 | 0.92 | **1.0** | 0.95 |
| MABIC | 0.86 | **1.0** | 0.97 | 0.94 |
| UMN | 0.91 | 0.92 | 0.95 | 0.93 |
| RMIT | 0.71 | 0.92 | **1.0** | 0.88 |
| RetinAI | 0.99 | 0.78 | 0.82 | 0.86 |
| NJUST | 0.70 | 0.83 | 0.98 | 0.84 |

overlapping of the pixels in the range from 0 to 1 with 1 being the perfect score and 0 being the worst.

2) The Absolute Volume Difference (AVD) [71] which is the absolute difference between the predicted and the ground truth. The value ranges from 0 to 1 with 0 being the best result and 1 being the worst.

The equation to calculate the DS is shown on Eqn (1) and that for AVD in Eqn (2). Where X is the raw input or raw image, Y is the ground truth, or mask, ∩ is the intersection and | | is the absolute value.

$$DS = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

$$AVD = |X| - |Y| \tag{2}$$

### C. RESULTS

In this section we report the performance for the detection task measured by the Area Under the Curve (AUC), and the segmentation task measured by the Dice Score (DS) and Absolute Volume Difference (AVD) for the nnUNet_RASPP, and baseline nnUNet. We also compare our results to the current state-of-the-arts (SOTA) architectures.

The segmentation performance grouped by segment classes per algorithm measured in DS is illustrated in Table 2 with the corresponding diagram in FIGURE 4, and that measured in AVD is illustrated in Table 3 with corresponding diagram in FIGURE 4. Here we noticed the following:

1) The nnUNet_RASPP and nnUNet outperform the current SOTA architectures by a clear margin with a mean DS of 0.823 and 0.817 respectively. Also, obtaining a mean AVD of 0.036 for nnUNet and 0.041 for nnUNet_RASPP.
2) Enhancing the nnUNet improved the performance. The SRF class was the most difficult to segment with nnUNet_RASPP (the enhanced version of nnUNet) obtaining the best DS of 0.80 which is 2% higher than the standard nnUNet and and 5% higher than the best SOTA architecture. The nnUNet_RASPP also obtained the best SRF AVD of 0.016 compare to 0.017 of the baseline nnUNet or 0.026 of the best SOTA models.
3) The best mean AVD score of 0.032 is achieved by SAMedOCT.
4) The nnUNet and nnUNet_RASPP possess the second and third-best mean AVD scores, but they exhibit better

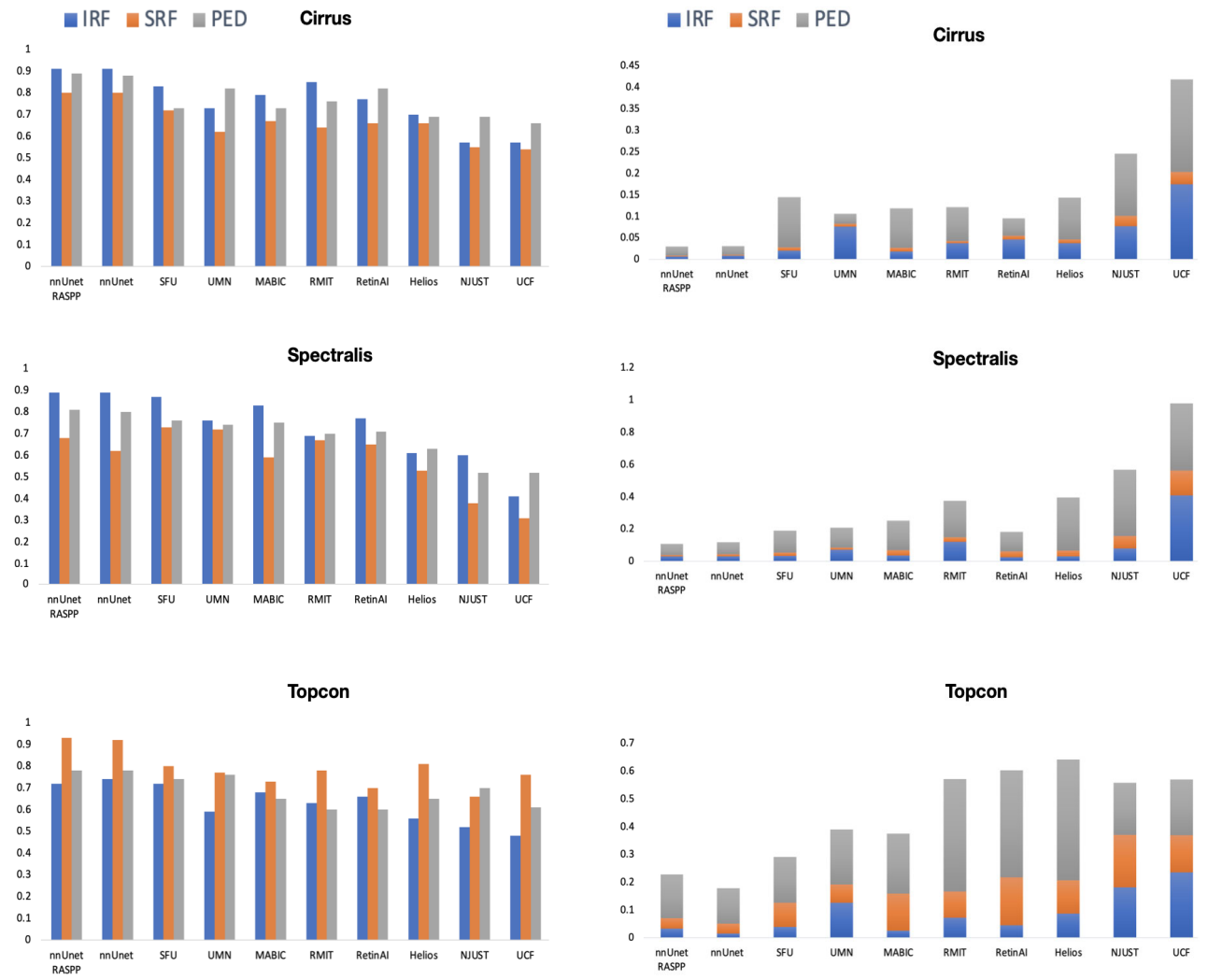


**FIGURE 5.** Segmentation performance comparison by DS on the right and AVD on the left of the nnUNet_RASPP and baseline nnUNet, together with the SOTA algorithms grouped by the segment classes when trained on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set per device.

IRF (0.019 and 0.021 compared to 0.042) and SRF (0.017 and 0.016 compared to 0.020) AVD scores than SAMedOCT.

5) Apart from the IRF class, the nnUNet_RASPP has the best DS in every single class when compare to the other models/teams.
6) IAUNet_SPP_CL and nnUNet_RASPP jointly achieve the second-best mean AVD score of 0.036.
7) We observed that, overall, the CNN/DNN models exhibit slightly better performance than the foundational model (SAMedOCT [12]). We believe this is because SAMedOCT is constructed with ViT as a backbone, and ViTs are more data-hungry than CNNs due to their ability to model long-range dependencies, as explained in [23].

A detail break down of the DS and AVD per vendor device trained on the entire 70 volumes and tested on the holding 42 cases of the testing set is shown in Table 5 with the corresponding diagrams in FIGURE 5. We noticed the following:

1) nnUNet_RASPP outperformed the baseline nnUNet and the state-of-the-arts models in two (Cirrus and Spectralis) of the 3 devices in both DS and AVD. The nnUNet RASPP model came in second place on the third device (Topcon) with a marginal difference from the baseline model, nnUNet.
2) The nnUNet_RASPP and nnUnet were the only two algorithms to maintain constant high level performance and generalisability across all classes and data sources in both DS and AVD. Both models constantly occupied the top 2 spots in performance per segment classes and vendor devices.

Table 6 with its corresponding diagrams in FIGURE 6 shows the results when trained on 2 vendor devices from the training set and tested on the third device from the holding testing set measured in DS and AVD. In this case

**TABLE 5.** Segmentation table of the Dice Score (DS) and Absolute Volume Difference (AVD) by segment classes (columns) and teams (rows) for training on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set per device.

| Cirrus | | | | | | |
|---|---|---|---|---|---|---|
| Methods | IRF | | SRF | | PED | |
| nnUNet_RASPP | **0.91** | **0.00670** | **0.80** | **0.00190** | **0.89** | 0.021700 |
| nnUNet | **0.91** | 0.00850 | **0.80** | **0.00190** | 0.88 | **0.02060** |
| SFU | 0.83 | 0.020388 | 0.72 | 0.008069 | 0.73 | 0.116385 |
| UMN | 0.73 | 0.076024 | 0.62 | 0.007309 | 0.82 | 0.023110 |
| MABIC | 0.79 | 0.018695 | 0.67 | 0.008188 | 0.73 | 0.091524 |
| RMIT | 0.85 | 0.037172 | 0.64 | 0.005207 | 0.76 | 0.079259 |
| RetinAI | 0.77 | 0.046548 | 0.66 | 0.008857 | 0.82 | 0.040525 |
| Helios | 0.70 | 0.038073 | 0.66 | 0.008313 | 0.69 | 0.097135 |
| NJUST | 0.57 | 0.077267 | 0.55 | 0.024092 | 0.69 | 0.144518 |
| UCF | 0.57 | 0.174140 | 0.54 | 0.028924 | 0.66 | 0.215379 |
| Spectralis | | | | | | |
| Methods | IRF | | SRF | | PED | |
| nnUNet_RASPP | **0.89** | **0.030100** | 0.68 | **0.008400** | **0.81** | **0.068600** |
| nnUNet | **0.89** | 0.031400 | 0.62 | 0.012600 | 0.80 | 0.073600 |
| SFU | 0.87 | 0.033594 | **0.73** | 0.020017 | 0.76 | 0.135562 |
| UMN | 0.76 | 0.072541 | 0.72 | 0.013499 | 0.74 | 0.121404 |
| MABIC | 0.83 | 0.036273 | 0.59 | 0.033384 | 0.75 | 0.181842 |
| RMIT | 0.69 | 0.121642 | 0.67 | 0.026377 | 0.70 | 0.228323 |
| RetinAI | 0.77 | 0.026921 | 0.65 | 0.036062 | 0.71 | 0.120528 |
| Helios | 0.61 | 0.030149 | 0.53 | 0.035625 | 0.63 | 0.330431 |
| NJUST | 0.60 | 0.080740 | 0.38 | 0.076071 | 0.52 | 0.412231 |
| UCF | 0.41 | 0.407741 | 0.31 | 0.155769 | 0.52 | 0.414739 |
| Topcon | | | | | | |
| Methods | IRF | | SRF | | PED | |
| nnUNet_RASPP | 0.72 | 0.032500 | **0.93** | 0.037800 | **0.78** | 0.157300 |
| nnUNet | **0.74** | **0.015900** | 0.92 | **0.036300** | **0.78** | **0.127700** |
| SFU | 0.72 | 0.039515 | 0.80 | 0.085907 | 0.74 | 0.164926 |
| UMN | 0.59 | 0.125454 | 0.77 | 0.066680 | 0.76 | 0.197794 |
| MABIC | 0.68 | 0.025097 | 0.73 | 0.134050 | 0.65 | 0.215687 |
| RMIT | 0.63 | 0.072609 | 0.78 | 0.094004 | 0.60 | 0.404842 |
| RetinAI | 0.66 | 0.045674 | 0.70 | 0.171808 | 0.60 | 0.385178 |
| Helios | 0.56 | 0.086773 | 0.81 | 0.119888 | 0.65 | 0.435057 |
| NJUST | 0.52 | 0.181237 | 0.66 | 0.188827 | 0.70 | 0.187733 |
| UCF | 0.48 | 0.235298 | 0.76 | 0.134283 | 0.61 | 0.200602 |

**TABLE 6.** Generalisation table of the DS and AVD by segment classes (columns) and teams (rows) trained on 48 OCT volumes from 2 device sources and evaluated on 14 OCT volumes from the testing set on the third device that wasn't seen at training.

| Cirrus | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Teams | IRF | | SRF | | PED | | Mean | |
| nnUNet_RASPP | **0.90** | **0.0122** | **0.78** | **0.0031** | **0.89** | **0.019** | **0.86** | **0.0114** |
| SFU | 0.83 | 0.0204 | 0.72 | 0.0081 | 0.73 | 0.1164 | 0.76 | 0.0483 |
| UMN | 0.73 | 0.0760 | 0.62 | 0.0073 | 0.82 | 0.0231 | 0.72 | 0.0355 |
| MABIC | 0.79 | 0.0187 | 0.67 | 0.0082 | 0.73 | 0.0915 | 0.73 | 0.0395 |
| RMIT | 0.85 | 0.0372 | 0.64 | 0.0052 | 0.76 | 0.0793 | 0.75 | 0.0406 |
| RetinAI | 0.77 | 0.0466 | 0.66 | 0.0089 | 0.82 | 0.0405 | 0.75 | 0.0320 |
| Helios | 0.70 | 0.0381 | 0.66 | 0.0083 | 0.69 | 0.0971 | 0.68 | 0.0478 |
| SVDNA [40] | 0.61 | – | 0.66 | – | 0.74 | – | 0.67 | – |
| NJUST | 0.57 | 0.0773 | 0.55 | 0.0241 | 0.69 | 0.1446 | 0.60 | 0.0820 |
| UCF | 0.57 | 0.1741 | 0.54 | 0.0289 | 0.66 | 0.2154 | 0.59 | 0.1395 |
| Topcon | | | | | | | | |
| Teams | IRF | | SRF | | PED | | Mean | |
| nnUNet_RASPP | 0.72 | **0.0201** | **0.93** | **0.0298** | **0.78** | **0.2119** | **0.81** | **0.0873** |
| SFU | 0.72 | 0.0395 | 0.80 | 0.0859 | 0.74 | 0.1649 | 0.75 | 0.0968 |
| UMN | 0.59 | 0.1255 | 0.77 | 0.0667 | 0.76 | 0.1978 | 0.71 | 0.1300 |
| SVDNA [40] | 0.61 | – | 0.80 | – | 0.72 | – | 0.71 | – |
| MABIC | 0.68 | 0.0251 | 0.73 | 0.1341 | 0.65 | 0.2157 | 0.69 | 0.1250 |
| RMIT | 0.63 | 0.0726 | 0.78 | 0.0940 | 0.60 | 0.4048 | 0.67 | 0.1905 |
| RetinAI | 0.66 | 0.0457 | 0.70 | 0.1718 | 0.60 | 0.3852 | 0.65 | 0.2009 |
| Helios | 0.56 | 0.0868 | 0.81 | 0.1199 | 0.65 | 0.4351 | 0.67 | 0.2139 |
| NJUST | 0.52 | 0.1812 | 0.66 | 0.1888 | 0.70 | 0.1877 | 0.63 | 0.1859 |
| UCF | 0.48 | 0.2353 | 0.76 | 0.1343 | 0.61 | 0.2006 | 0.62 | 0.1900 |

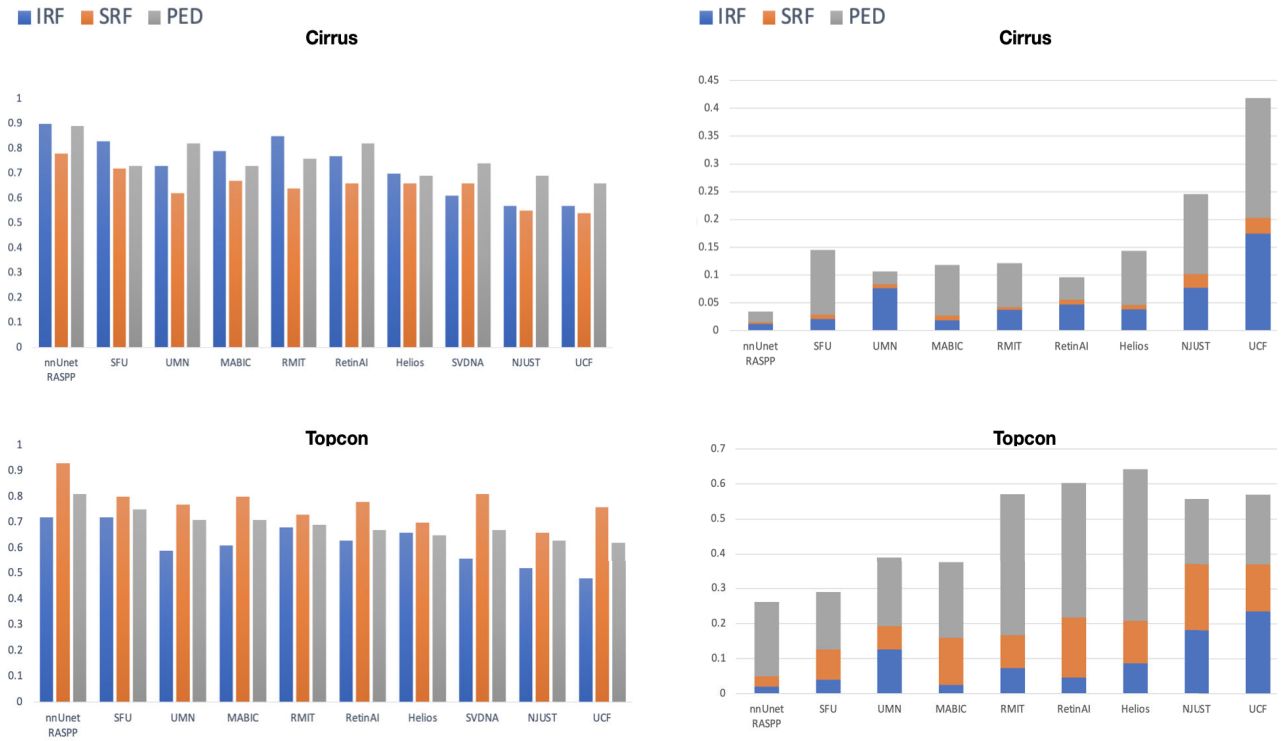


**FIGURE 6.** Segmentation performance comparison by DS on the right and AVD on the left of the nnUNet_RASPP and baseline nnUNet, together with the SOTA algorithms grouped by the segment classes when trained on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set per device.

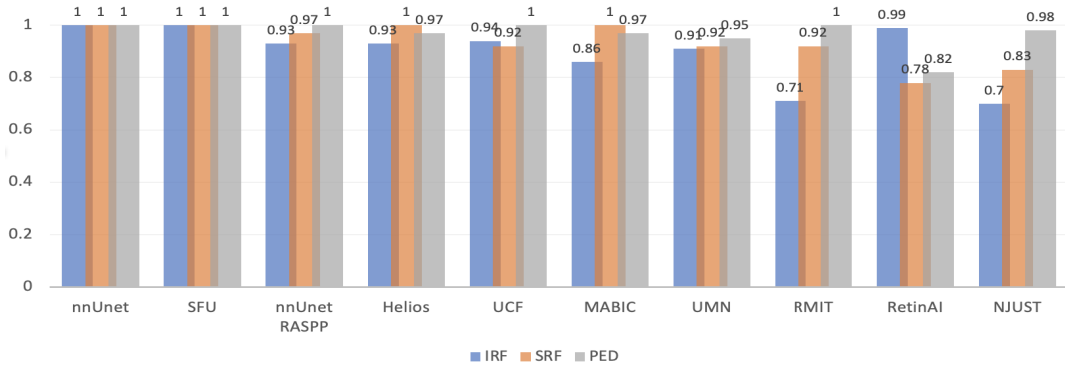


**FIGURE 7.** Detection performance comparison by DS of the nnUNet_RASPP and baseline nnUNet, together with the state-of-the-arts algorithms grouped by the segment classes when trained on the entire 70 OCT volumes of the training set and tested on the holding 42 OCT volumes from the testing set.

because of the constraint of the evaluation submission (curb to 3 maximum per team) of the predicted segmentation on the testing set, results for nnUnet are unavailable. Here we noticed that

1) nnUNet_RASPP outperformed the current SOTA architecture scoring a mean DS of 0.86 (10% higher than the second best) on the Cirrus device and 0.81 (6% higher than the second best) on the Topcon device;
2) nnUNet_RASPP also obtained the best AVD scores, scoring a mean of 0.0114 and 0.0878 on the Cirrus and Topcon devices respectively;
3) nnUNet_RASPP still maintained its high level of robustness and generalisability with a consistently high level of performance measure in DS and AVD.

The detection performance grouped by segment classes per algorithm measured by the AUC is illustrated in Table 4 with the corresponding diagram in FIGURE 7. Here the nnUNet obtained a perfect AUC score of 1 for all three fluid classes and nnUNet_RASPP obtained an AUC score of 0.93, 0.97, and 1.0 for the IRF, SRF, and PED respectively.

The visualizations using orange arrows to highlight the fine details capture by nnUNet_RASPP when trained on

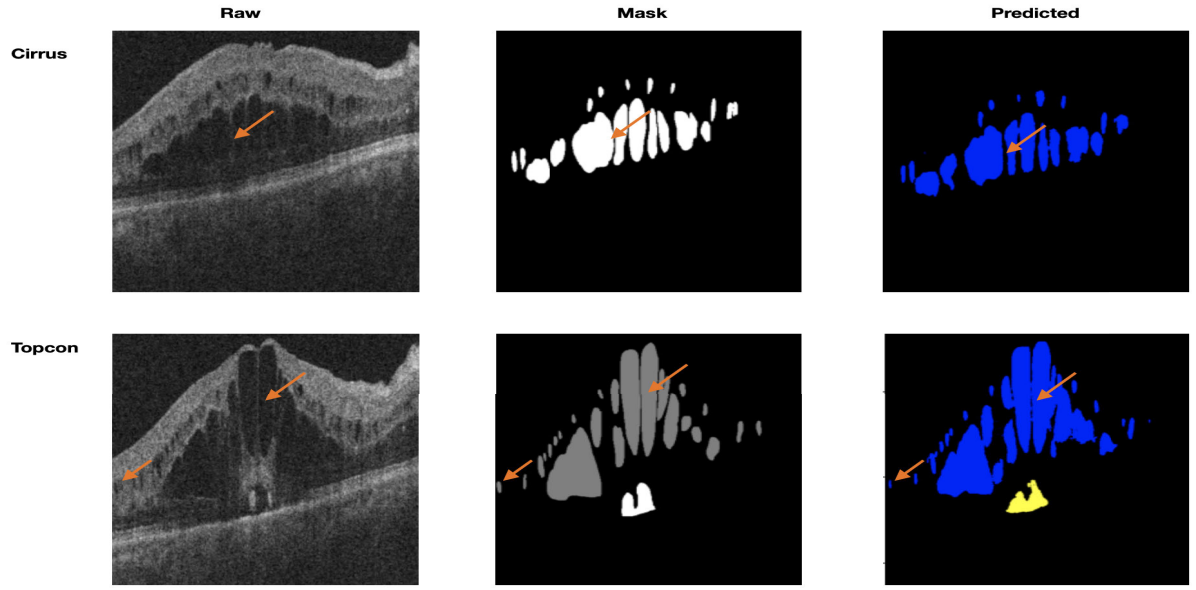


**FIGURE 8.** Examples of B-scans to illustrate the visualization output/predicted of nnUNet_RASPP, in order of the raw/inputs, mask/annotations and predicted/outputs in columns when trained on the training set of two vendor devices and tested on the training set of the third vendor device (Cirrus and Topcon in row 1 and row 2 respectively). Fine details captured by the model are indicated with orange arrows.

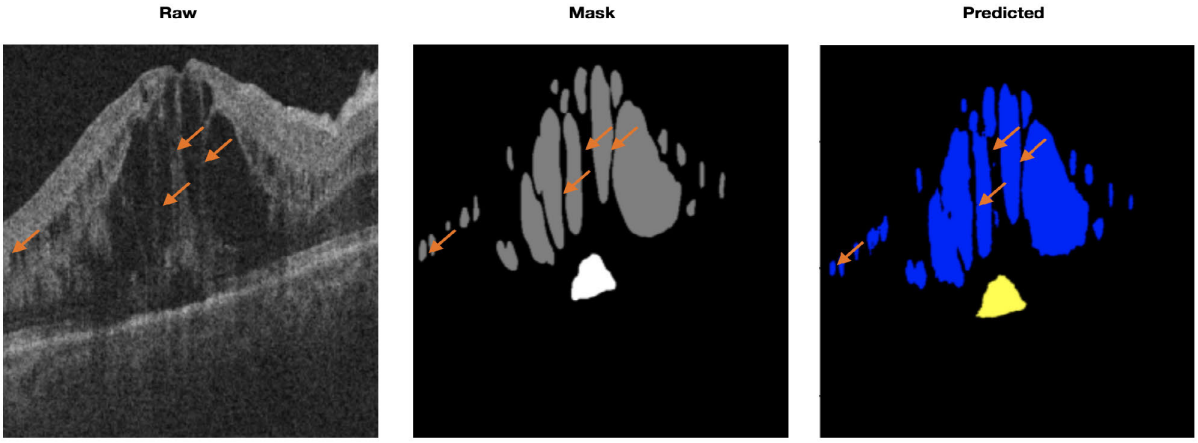


**FIGURE 9.** An example of a B-scan to illustrate the visualization output/predicted of nnUNet_RASPP, in order of the raw/inputs, mask/annotations and predicted/outputs when zoom out to highlights the fine details captured by the model using orange arrows. This is captured when trained on the training set of the Spectralis and Topcon devices and tested on the training set of the Cirrus device.

two vendor devices from the training set and tested on the third from the training set are illustrated in FIGURE 8 and FIGURE 9.

## V. CONCLUSION

In this work, we have investigated the problems of detection and segmentation of multiple fluids in retinal OCT volumes acquired from multiple device vendors using SOTA deep learning methods. We have included the most representative methods as appeared in the leading positions of the RETOUCH competition, and evaluated their performance based on the hidden test results where the testing datasets are not available to the competition participants. The RETOUCH datasets are among the largest for the underlying problems. It demonstrates a high level of variability due to data collected from devices of various vendors.

The key findings of the work include the following:

1) Firstly, it provides a comprehensive review of the representative models to address the problem of retina fluid segmentation, detection from OCT images and generalisation performance over variations of data source.
2) Secondly, through a blinded evaluation where the ground truth on the testing datasets is withheld by the competition organisers, it is demonstrated that large foundation models, including SAM and its variants SAMed and SAMedOCT, exhibit promising

performance in the segmentation task through a routine fine-tuning process.

3) Thirdly, at the current stage, the specifically designed and trained deep networks such as nnUNet and nnUNet_RASPP still offer a slightly advantage on the performance of all tasks of segmentation, detection and generalisation.

We believe nnUNet_RASPP's slight outperformance in this particular problem is achieved by incorporating residual blocks to learn complex features and reduce the training error rate and an ASPP to capture global information. Moreover, it has an inherent simple architecture with fewer model parameters than the foundation models and therefore offers better real-time performance.

The methods included in this study provide useful information for further diagnosis and monitoring the progress of retinal diseases such as AMD, DME and Glaucoma. In the future, we aim to leverage the availability of various small public medical image datasets to assess the performance of these methods on a more heterogeneous dataset (combination of datasets originating from different modalities and anatomical regions).

## ACKNOWLEDGMENT

The authors would like to thank Hrvoje Bogunovic for his invaluable support and advice during their participation in the RETOUCH competition, and also would like to thank the Department of Computer Science, Brunel University London for providing the computational resources to conduct the experiments.

## REFERENCES

[1] *Multi-Atlas Labeling Beyond the Cranial Vault—Workshop and Challenge*. [Online]. Available: https://paperswithcode.com/dataset/miccai-2015-multi-atlas-abdomen-labeling

[2] *Multi-Atlas Labeling Beyond the Cranial Vault—Workshop and Challenge*. [Online]. Available: https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html

[3] *Retouch: The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge*. [Online]. Available: https://retouch.grand-challenge.org/Background/

[4] S. Akbar, M. Hayat, M. Tahir, S. Khan, and F. K. Alarfaj, "CACP-DeepGram: Classification of anticancer peptides via deep neural network and skip-gram-based word embedding model," *Artif. Intell. Med.*, vol. 131, Sep. 2022, Art. no. 102349.

[5] S. Akbar, H. G. Mohamed, H. Ali, A. Saeed, A. A. Khan, S. Gul, A. Ahmad, F. Ali, Y. Y. Ghadi, and M. Assam, "Identifying neuropeptides via evolutionary and sequential based multi-perspective descriptors by incorporation with ensemble classification strategy," *IEEE Access*, vol. 11, pp. 49024–49034, 2023.

[6] S. Akbar, A. Raza, T. A. Shloul, A. Ahmad, A. Saeed, Y. Y. Ghadi, O. Mamyrbayev, and E. Tag-Eldin, "PAtbP-EnC: Identifying anti-tubercular peptides using multi-feature representation and genetic algorithm-based deep ensemble model," *IEEE Access*, vol. 11, pp. 137099–137114, 2023.

[7] K. Alsaih, M. Z. Yusoff, T. B. Tang, I. Faye, and F. Mériaudeau, "Retinal fluids segmentation using volumetric deep neural networks on optical coherence tomography scans," in *Proc. 10th IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Aug. 2020, pp. 68–72.

[8] B. N. Anoop, R. Pavan, G. N. Girish, A. R. Kothari, and J. Rajan, "Stack generalized deep ensemble learning for retinal layer segmentation in optical coherence tomography images," *Biocybernetics Biomed. Eng.*, vol. 40, no. 4, pp. 1343–1358, Oct. 2020.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[10] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[11] H. Bogunovic, F. Venhuizen, S. Klimscha, S. Apostolopoulos, A. Bab-Hadiashar, U. Bagci, M. F. Beg, L. Bekalo, Q. Chen, C. Ciller, and K. Gopinath, "RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1858–1874, Aug. 2019.

[12] F. Botond, J. Morano, D. Lachinov, G. Aresta, and H. Bogunovic, "SAMedOCT: Adapting segment anything model (SAM) for retinal OCT," 2023, *arXiv:2308.09331*.

[13] A. Carass, S. Roy, A. Gherman, J. C. Reinhold, A. Jesson, T. Arbel, O. Maier, H. Handels, M. Ghafoorian, B. Platel, A. Birenbaum, H. Greenspan, D. L. Pham, C. M. Crainiceanu, P. A. Calabresi, J. L. Prince, W. R. G. Roncal, R. T. Shinohara, and I. Oguz, "Evaluating white matter lesion segmentations with refined Sørensen-dice analysis," *Sci. Rep.*, vol. 10, no. 1, p. 8242, May 2020.

[14] U. K. Challa, P. Yellamraju, and J. S. Bhatt, "A multi-class deep all-CNN for detection of diabetic retinopathy using retinal fundus images," in *Proc. Int. Conf. Pattern Recognit. Mach. Intell.* New York, NY, USA: Springer, 2019, pp. 191–199.

[15] C. Chen, J. Miao, D. Wu, Z. Yan, S. Kim, J. Hu, A. Zhong, Z. Liu, L. Sun, X. Li, T. Liu, P.-A. Heng, and Q. Li, "MA-SAM: Modality-agnostic SAM adaptation for 3D medical image segmentation," 2023, *arXiv:2309.08842*.

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[17] 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, pp. 801–818.

[18] 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. Exp.*, vol. 6, no. 4, pp. 1172–1194, 2015.

[19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[20] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, Athens, Greece. New York, NY, USA: Springer, 2016.

[21] B. I. Dodo, Y. Li, D. Kaba, and X. Liu, "Retinal layer segmentation in optical coherence tomography images," *IEEE Access*, vol. 7, pp. 152388–152398, 2019.

[22] B. I. Dodo, Y. Li, X. Liu, and M. I. Dodo, "Level set segmentation of retinal OCT images," in *Proc. BIOIMAGING*, 2019, pp. 49–56.

[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[24] 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. Exp.*, vol. 8, no. 5, pp. 2732–2744, 2017.

[25] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, and G. Van Den Driessche, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Med.*, vol. 24, no. 9, pp. 1342–1350, Sep. 2018.

[26] P. Gholami, P. Roy, M. K. Parthasarathy, and V. Lakshminarayanan, "OCTID: Optical coherence tomography image database," *Comput. Electr. Eng.*, vol. 81, Jan. 2020, Art. no. 106532.

[27] B. Hassan, S. Qin, R. Ahmed, T. Hassan, A. H. Taguri, S. Hashmi, and N. Werghi, "Deep learning based joint segmentation and characterization of multi-class retinal fluid lesions on OCT scans for clinical use in anti-VEGF therapy," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104727.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision—ECCV*, Amsterdam, The Netherlands. New York, NY, USA: Springer, 2016, pp. 630–645.
[30] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," 2021, *arXiv:2106.09685*.
[31] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NnU-net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.
[32] S. Jetley, N. A. Lord, N. Lee, and P. H. S. Torr, "Learn to pay attention," 2018, *arXiv:1804.02391*.
[33] D. Kaba, Y. Wang, C. Wang, X. Liu, H. Zhu, A. G. Salazar-Gonzalez, and Y. Li, "Retina layer segmentation using kernel graph cuts and continuous max-flow," *Opt. Exp.*, vol. 23, no. 6, pp. 7366–7384, 2015.
[34] D. Kaba, A. G. Salazar-Gonzalez, Y. Li, X. Liu, and A. Serag, "Segmentation of retinal blood vessels using Gaussian mixture models and expectation maximisation," in *Proc. Int. Conf. Health Inf. Sci.*, 2016, pp. 105–112.
[35] D. Kaba, C. Wang, Y. Li, A. Salazar-Gonzalez, X. Liu, and A. Serag, "Retinal blood vessels extraction using probabilistic modelling," *Health Inf. Sci. Syst.*, vol. 2, no. 1, pp. 1–10, Dec. 2014.
[36] 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," Amazonaws, USA, Tech. Rep., 2017.
[37] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, and J. Dong, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, Feb. 2018.
[38] J. Kim and L. Tran, "Retinal disease classification from OCT images using deep learning algorithms," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2021, pp. 1–6.
[39] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, and R. Girshick, "Segment anything," 2023, *arXiv:2304.02643*.
[40] V. Koch, O. Holmberg, H. Spitzer, J. Schiefelbein, B. Asani, M. Hafner, and F. J. Theis, "Noise transfer for unsupervised domain adaptation of retinal OCT images," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2022.
[41] J. Kugelman, D. Alonso-Caneiro, S. A. Read, J. Hamwood, S. J. Vincent, F. K. Chen, and M. J. Collins, "Automatic choroidal segmentation in OCT images using supervised deep learning methods," *Sci. Rep.*, vol. 9, no. 1, p. 13298, Sep. 2019.
[42] J. Q. Li, T. Welchowski, M. Schmid, J. Letow, A. C. Wolpers, F. G. Holz, and R. P. Finger, "Retinal diseases in Europe," *Eur. Soc. Retina Specialists (EURETINA)*, vol. 24, pp. 20–30, Aug. 2017.
[43] Q. Li, S. Li, Z. He, H. Guan, R. Chen, Y. Xu, T. Wang, S. Qi, J. Mei, and W. Wang, "DeepRetina: Layer segmentation of retina in OCT images using deep learning," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 61, Dec. 2020.
[44] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5168–5177.
[45] X. Liu, S. Wang, Y. Zhang, D. Liu, and W. Hu, "Automatic fluid segmentation in retinal optical coherence tomography images using attention based deep learning," *Neurocomputing*, vol. 452, pp. 576–591, Sep. 2021.
[46] D. Lu, M. Heisler, S. Lee, G. Ding, M. V. Sarunic, and M. Faisal Beg, "Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network," 2017, *arXiv:1710.04778*.
[47] D. Ma, D. Lu, M. Heisler, S. Dabiri, S. Lee, G. W. Ding, M. V. Sarunic, and M. F. Beg, "Cascade dual-branch deep neural networks for retinal layer and fluid segmentation of optical coherence tomography incorporating relative positional map," in *Proc. Med. Imag. Deep Learn.*, 2020, pp. 493–502.
[48] I. Mantel, A. Mosinska, C. Bergin, M. S. Polito, J. Guidotti, S. Apostolopoulos, C. Ciller, and S. De Zanet, "Automated quantification of pathological fluids in neovascular age-related macular degeneration, and its repeatability using deep learning," *Transl. Vis. Sci. Technol.*, vol. 10, no. 4, p. 17, Apr. 2021.
[49] N. McConnell, A. Miron, Z. Wang, and Y. Li, "Integrating residual, dense, and inception blocks into the nnUNet," in *Proc. IEEE 35th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2022, pp. 217–222.
[50] N. McConnell, N. Ndipenoch, Y. Cao, A. Miron, and Y. Li, "Exploring advanced architectural variations of nnUNet," *Neurocomputing*, vol. 560, Dec. 2023, Art. no. 126837.
[51] M. Melinscak, M. Radmilovic, Z. Vatavuk, and S. Loncaric, "Annotated retinal optical coherence tomography images (AROI) database for joint retinal layer and fluid segmentation," *Automatika*, vol. 62, nos. 3–4, pp. 375–385, Oct. 2021.
[52] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, and R. Wiest, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.
[53] Y.-H. Nai, B. W. Teo, N. L. Tan, S. O'Doherty, M. C. Stephenson, Y. L. Thian, E. Chiong, and A. Reilhac, "Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset," *Comput. Biol. Med.*, vol. 134, Jul. 2021, Art. no. 104497.
[54] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "EdgeConnect: Generative image inpainting with adversarial edge learning," 2019, *arXiv:1901.00212*.
[55] N. Ndipenoch, A. Miron, Z. Wang, and Y. Li, "Simultaneous segmentation of layers and fluids in retinal OCT images," in *Proc. 15th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Nov. 2022, pp. 1–6.
[56] N. Ndipenoch, A. Miron, Z. Wang, and Y. Li, "Retinal image segmentation with small datasets," 2023, *arXiv:2303.05110*.
[57] M. Pekala, N. Joshi, T. Y. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," *Comput. Biol. Med.*, vol. 114, Nov. 2019, Art. no. 103445.
[58] 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*, 2017 pp. 54–62.
[59] A. Raza, J. Uddin, A. Almuhaimeed, S. Akbar, Q. Zou, and A. Ahmad, "AIPs-SnTCN: Predicting anti-inflammatory peptides using fastText and transformer encoder-based hybrid word embedding with self-normalized temporal convolutional networks," *J. Chem. Inf. Model.*, vol. 63, no. 21, pp. 6537–6554, Nov. 2023.
[60] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, Munich, Germany. New York, NY, USA: Springer, 2015.
[61] 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. Exp.*, vol. 8, no. 8, pp. 3627–3642, 2017.
[62] R. Sznitman S. Apostolopoulos, C. Ciller, and S. De Zanet, "Simultaneous classification and segmentation of cysts in retinal OCT," in *Proc. MICCAI* , 2017, pp. 22–29.
[63] A. Salazar-Gonzalez, D. Kaba, Y. Li, and X. Liu, "Segmentation of the blood vessels and optic disk in retinal images," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 6, pp. 1874–1886, Nov. 2014.
[64] A. Salazar-Gonzalez, Y. Li, and D. Kaba, "MRF reconstruction of retinal images for the optic disc segmentation," in *Health Information Science*, Beijing, China. Berlin, Germany: Springer, 2012.
[65] A. G. Salazar-Gonzalez, Y. Li, and X. Liu, "Retinal blood vessel segmentation via graph cut," in *Proc. 11th Int. Conf. Control Autom. Robot. Vis.*, Dec. 2010, pp. 225–230.
[66] A. G. Salazar-Gonzalez, Y. Li, and X. Liu, "Optic disc segmentation by incorporating blood vessel compensation," in *Proc. IEEE 3rd Int. Workshop Comput. Intell. Med. Imag.*, Apr. 2011, pp. 1–8.
[67] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[68] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, and S. Farsiu, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 5, no. 10, pp. 3568–3577, 2014.
[69] M. Subramanian, K. Shanmugavadivel, O. S. Naren, K. Premkumar, and K. Rankish, "Classification of retinal OCT images using deep learning," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2022, pp. 1–7.
[70] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[71] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, pp. 1–28, Dec. 2015.
[72] R. Tennakoon, A. K. Gostar, R. Hoseinnezhad, and A. Bab-Hadiashar, "Retinal fluid segmentation in OCT images using adversarial loss based convolutional neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1436–1440.
[73] J. Tian, B. Varga, E. Tatrai, P. Fanni, G. M. Somfai, W. E. Smiddy, and D. C. Debuc, "Performance evaluation of automated segmentation software on optical coherence tomography volume data," *J. Biophotonics*, vol. 9, no. 5, pp. 478–489, May 2016.
[74] I. A. Viedma, D. Alonso-Caneiro, S. A. Read, and M. J. Collins, "Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey," *Neurocomputing*, vol. 507, pp. 247–264, Oct. 2022.
[75] C. Wang, Y. X. Wang, and Y. Li, "Automatic choroidal layer segmentation using Markov random field and level set method," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 6, pp. 1694–1702, Nov. 2017.
[76] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, Jan. 1992.
[77] J. Wu, J. Chen, Z. Xiao, and L. Geng, "Automatic layering of retinal OCT images with dual attention mechanism," in *Proc. 3rd Int. Conf. Intell. Med. Image Process.*, Apr. 2021, pp. 63–69.
[78] J. Xie and Y. Peng, "The head and neck tumor segmentation using nnU-Net with spatial and channel 'squeeze & excitation' blocks," in *Head and Neck Tumor Segmentation*, Lima, Peru. Springer, 2021.
[79] G. Xing, L. Chen, H. Wang, J. Zhang, D. Sun, F. Xu, J. Lei, and X. Xu, "Multi-scale pathological fluid segmentation in OCT with a novel curvature loss in convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 41, no. 6, pp. 1547–1559, Jun. 2022.
[80] Z. Yang and S. Farsiu, "Directional connectivity-based segmentation of medical images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023.
[81] H. Zhang, J. Yang, K. Zhou, F. Li, Y. Hu, Y. Zhao, C. Zheng, X. Zhang, and J. Liu, "Automatic segmentation and visualization of choroid in OCT with knowledge infused deep learning," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3408–3420, Dec. 2020.
[82] K. Zhang and D. Liu, "Customized segment anything model for medical image segmentation," 2023, *arXiv:2304.13785*.
[83] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu, "nnFormer: Interleaved transformer for volumetric segmentation," 2021, *arXiv:2109.03201*.

**NCHONGMAJE NDIPENOCH** (Student Member, IEEE) received the B.Sc. degree in software development and the M.Sc. degree in computer science. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Brunel University London. His research interests include artificial intelligence and data science for medical imaging, and software development.

**ALINA MIRON** (Member, IEEE) received the B.Eng. and M.Eng. degrees from Babes-Bolyai University, Romania, and the Ph.D. degree in machine learning in the field of autonomous vehicles from INSA de Rouen. She is currently a Lecturer with Brunel University London. She is also an Artificial Intelligence Researcher, a Developer, and an Educator. Her current research interests include computer vision and machine learning applied to medical imagining, and the analysis of human behavior from videos.

**YONGMIN LI** (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from Tsinghua University, China, and the Ph.D. degree from the Queen Mary University of London. Before joining Brunel University London, he was a Research Scientist with the British Telecom Laboratories. His research interests include the areas of data science, machine learning, artificial intelligence, image processing, computer vision, video analysis, medical imaging, bio-imaging, biomedical engineering, healthcare technologies, automatic control, and nonlinear filtering. Together with his colleagues, he has won the Most Influential Paper over the Decade Award at MVA 2019 and Best Paper Awards at Bioimaging 2018, HIS 2012, BMVC 2007, BMVC 2001, and RATFG 2001. He is a Senior Fellow of the Higher Education Academy.