Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

KHALED ALSAIH[1], **MOHD ZUKI YUSOFF**[1], **(Member, IEEE)**,
IBRAHIMA FAYE[1], **(Senior Member, IEEE)**, **TONG BOON TANG**[1], **(Senior Member, IEEE)**,
AND FABRICE MERIAUDEAU[2], **(Member, IEEE)**

[1]Centre for Intelligent Signal and Imaging Research, Electrical and Electronic Engineering Department,
Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
[2]ImViA/IFTIM, Universite Bourgogne Franche-Comté, 21078 Dijon, France

Corresponding author: Khaled Alsaih (khaledalsaih@gmail.com)

ABSTRACT Morphological changes related to different diseases that occur in the retina are currently extensively researched. Manual segmentation of retinal fluids is time-consuming and subject to variability, giving prominence to the demand for robust automatic segmentation methods. The standard in assessing the existence and mass of retinal fluids at present is through the optical coherence tomography (OCT) modality. In this study, semantic segmentation deep learning networks were examined in 2.5D and ensembled with 2D networks. This analysis aims to show how these networks can perform in-depth than using only a single B-scan and the effects of 2.5 patches when fitted to the deep networks. All experiments were evaluated using public data from the RETOUCH challenge as well as the OPTIMA challenge dataset and Duke dataset. The networks trained in 2.5D performed slightly better than 2D networks in all datasets. The average performance of the best network was 0.867, using the dice similarity coefficient score (DSC) metric on the RETOUCH dataset. On the DUKE dataset, $Deeplabv3+^{Pa}$ outperformed other networks in this study with a dice score of 0.80. Experiments showed a more robust performance when networks were ensembled. Intraretinal fluid (IRF) was recognized better than other fluids with a DSC of 0.924. $Deeplabv3+^{Pa}$ model outperformed all other networks with a p-value average of 0.03 on the RETOUCH challenge dataset. Methods used in this study to distinguish retinal disorders outperform human performance as well as showed competitive results to the teams who joined both challenges. Three consecutive B-scans, including partial depth information in training neural networks, were stacked as a single image built for more robust networks compared to providing only 2D information.

INDEX TERMS 2.5D networks, deep learning, ensembled networks, retinal diseases, SD-OCT segmentation.

I. INTRODUCTION

Human eyes are one of the main organs, which makes it crucial to keep them healthy. However, due to increasing age among humans, some diseases such as age-related macular degeneration (AMD) can affect the eyes' condition. Additionally, patients diagnosed with diabetes are prone to develop retinal diseases known as diabetic retinopathy (DR) that eventually can lead to diabetic macular edema (DME)

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

if left untreated. The exudation from retinal capillaries and gradual congregation of the drained fluid within the spaces of the center of the retina causes swelling, known as macular edema [1], which can lead to unexpected or acute loss of vision. Age-related macular degeneration (AMD) costs $4.6 billion per annum in the United States (US) alone for direct healthcare [2]. Individuals diagnosed with AMD has reached 11 million people in the US, and the figure is anticipated to double by 2050. Anti-vascular endothelial growth factor therapy (anti-VEGF) is a prevalent treatment for AMD, whereby the growth of abnormal blood vessels are dissolved.

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

IEEE*Access*

Patients who undergo anti-VEGF and monitoring sessions would be tracked on the effectiveness of the suggested medical treatment. However, anti-VEGF drugs and the monitoring sessions after each treatment are expensive, which would result in a substantial financial burden for the patients and the healthcare system.

Moreover, DME and AMD are only a small portion of the diseases /conditions affecting the human eye (presbyopia, glaucoma, cataract, DR,....), and the number of ophthalmologists is rather low compared to the ever-increasing number of patients. As such, the number of ophthalmologists is comparatively lower than the ever-increasing number of patients with eye conditions. Automated procedures are introduced to ease the stress of ophthalmologists by managing their time as well as reducing the financial burden of the patients and the healthcare systems. Among the procedures that could be automated are image segmentation and classification, which remove inter-graders variability and efficiently track the disease's progress during the screening process [3].
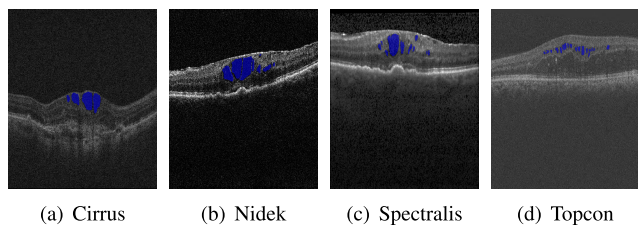

(a) Cirrus   (b) Nidek   (c) Spectralis   (d) Topcon

FIGURE 1. SD-OCT B-scan examples in OPTIMA challenge dataset with intraretinal cystoid fluid (IRC) annotation from different scanners.

Visualizing fluid that piles up for various retinal diseases is best monitored and assessed using spectral-domain optical coherence tomography (SD-OCT). This technique uses a three dimensional and non-invasive imaging modality that enables a clear and accurate view of the fluid and the retinal layers through a stack of B-scans (2D slices of the volume). The imaging principle would, however, have some natural noise, and the obtained images would be slightly altered in terms of resolution, numbers, and signal-to-noise ratio from one supplier to another. Fig. 1 shows the example of scans produced by each supplier of different settings for imaging the B-scans, different resolution range, and different scanning time, using a dataset from the OPTIMA challenge. The number of slices that have been provided by each supplier for each eye ranges from 49 slices to 128 slices. Due to a large amount of data produced from these scans, a clinician would likely take a long time to process and screen through the slices. Hence, automatic algorithms are needed to alleviate the time spent by the clinician and to reduce variability among experts.

Many studies had used private datasets to distinguish various retinal fluids, resulting in several standards on the detection and segmentation of retinal lesions being established. The aim of the OPTIMA challenge [4] was to segment the intraretinal cystoid fluid (IRC), while the RETOUCH challenge [3] aimed to differentiate intraretinal fluid (IRF), sub-retinal fluid (SRF), and pigment epithelial detachment (PED).

For the DUKE dataset [5], the target is to segment DME fluids. In the RETOUCH challenge, the volumes were defined with AMD cases and retinal vein occlusion (RVO) cases. However, the main objective of this study was the retinal fluid segmentation.

This study analyzes mainly four different deep networks for retinal disease segmentation aims. Following the introduction section, this work is constructed as follows: Sect. II concentrates on the related works on the AMD fluids segmentation. Sect. III illustrates our segmentation frameworks while Sect. IV is dedicated to show experiment results. Sect. V is devoted to discussion. Finally, this study is summed up in Sect. VI.

II. RELATED WORK

In recent years, efforts had been made by the medical imaging community in providing better images and analysis of the retina structure. Initial studies found in the retinal field was carried out by Hee [6] using OCT modality (A-scans). Optical coherence tomography began in 1991 [7] and gained attention among researchers by 2006 with the implementation of spectral-domain SD-OCT. Numerous studies since 2005 began to propose the diagnosis of the healthy retina from AMD [8] or DME [9]. Studies using B-scans that segmented fluids, such as IRF and SRF, were introduced based on the active contours [10]. Similar to other studies of other areas within the medical field, the initial objective of improving the image quality of retinal scans was to remove the noise and segment meaningful features (retinal layers) [11]. In contrast, current researches were mainly focused on deep learning architectures [12]–[14]. Hence, there were very few datasets available on retinal fluids segmentation, and in turn, may contain only a restricted number of fluids. A large dataset that had all retinal fluids was released [15], but the dataset remained private, with only two suppliers provided with the data in the study.

This study used the OPTIMA, DUKE, and the RETOUCH challenge datasets. Most of the papers submitted to these challenges analyzed the OCT volumes in a 2D manner. Machine learning techniques were used by participants of the OPTIMA challenge but [16]. Venhuizen *et al.* trained three independent convolution neural networks at various scales to differentiate between IRC fluid and background. The study used different scales for each image, and each patch was fitted to different CNNs. Then, the AND operation was applied to merge the three scales. Retinal layers were detected using the Iowa reference algorithm [17] to restrict the search of fluids between the inner limiting membrane (ILM) layer and the retinal pigment epithelium (RPE) layer. The method achieved a mean dice value of 0.64. In the DUKE dataset [5], a kernel regression method was developed to segment the retinal fluid and layers. The overall dice score of 0.53 and 0.45 for expert 1 and expert 2 were achieved, while [18] has developed a deep learning model to segment the retinal fluids in the DUKE dataset. The dice score of 0.62 and 0.53 for expert 1 and expert 2 were obtained from [18]. The ReLayNet model [19]

IEEE *Access*

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

scored the dice value of 0.81 using the DUKE dataset. Girish *et al.* [20] used a modified *U-net* model with the use of depthwise separable convolution operations. Additionally, Gopinath and Sivaswamy [21] validated the proposed CNN model using three datasets, including the OPTIMA dataset.

For the RETOUCH challenge dataset, eight groups with various methods using automated and semi-automated deep learning networks contributed in an end-to-end manner. First, the Helios group [22] trained a cascaded FCN network after the B-scans were denoised using a spectral total variation method. A generalized motion pattern (GMP) was then applied to generate motion in the resized image, with a size of (256 x 256) to suppress the background. After augmenting the data using the GMP algorithm, all images were fitted into the encoder-decoder architecture. In the study carried by the RetinaAI group [23], retinal layers were first segmented before the images were augmented by applying shear, Gaussian noise, and rotation. A residual branch was added to the *U-net* architecture with a 2D input to segment the fluids. The UMN group [18] also segmented the ILM layer and RPE layer to limit the search of fluids. PED was segmented by flattening the retinal layers, and the difference between the flattened RPE and elevated RPE was calculated. IRF and SRF were segmented in a supervised manner using four layers of CNN architecture, fed with the region of interest (ROI) images lying between the ILM layer and the RPE layer. MABIC group [24], on the other hand, added dropout and max out activations at each layer of the *U-net* architecture. Data was resized to (512 x 512) before being normalized and fitted into two *U-net* networks. The first network had one fully connected layer, and the output was fitted to the second network but without the fully connected layer. Data augmentation was also performed, and binary cross-entropy loss was used. Last, SFU group [25] had attached an extra channel to the B-scan that represented the intensity of the distance map for each B-scan. Groups mentioned above have used only 2D information, and no in-depth information was used during the network training. Only one group that participated in the RETOUCH challenge used 3D information. RMIT group [26] normalized the data before histogram matching. Images were then denoised using a median filter and resized to (256 x 128 x 3). An adversarial network was added to the *U-net* architecture and trained with cross-entropy loss, dice loss, and adversarial loss.

Few studies had used 2.5D information among them, with three groups using the extracting hand-crafted features of machine learning techniques [27], [28], and [29]. The two groups that contributed to the RETOUCH challenge used 2.5D information on deep learning models. First, a group named UCF [30] flattened the data using a 3D Gaussian kernel, and B-scans were resized to (512 x 256). CNN ResNet based architecture was proposed, and a novel myopic warping method was applied on B-scans to make the retina curvier. Another work performed by NJUST [31] used a combination of Faster R-CNN, region growing, and RPE layer segmentation to segment IRF, SRF, and PED, respectively.

The bilateral filter was used to reduce the speckle, and 3D smoothing was applied as a postprocessing step.

Segmentation of retinal fluids was still an issue in this area of research. Table 4 presents the summary of the segmentation methods. As shown, only a few algorithms utilized the 2.5D information to capture the fluid's extended depth. Hence, this study had analyzed four deep networks with 2.5D input, namely Seg-net [32], FCN [33], Deeplabv3+ [34], and *U-net* [35] to segment retinal fluids. Besides, 2.5D patches had been extracted and fitted to the networks, in which no medical field has examined the effects of the 2.5D patches. Also, this work had ensembled 2D trained networks along with 2.5D trained networks. Lastly, this analysis compared different dimensionality input to the networks to examine the effects of the input size on the deep models.

III. METHODOLOGY

Datasets involved in this study, intermediate steps to process the data, and networks used for the analysis were presented. Fig. 2 depicts the methodological process taken in this work.

A. DATASET DESCRIPTION

Three public datasets were used in this study; the RETOUCH challenge dataset [3], the DUKE dataset [5], and the OPTIMA challenge dataset [4]. In this analysis, the RETOUCH challenge dataset was used to train and validate the networks. The RETOUCH testing data was not used in this study as there were no ground truth (GT) images provided by the challenge organizers. Furthermore, the networks trained using the RETOUCH datasets were fine-tuned over the OPTIMA challenge dataset and DUKE dataset. The OPTIMA and DUKE dataset were released with labels for the training and testing dataset. Table 1 summarizes the details of each dataset.

B. PRE-PROCESSING

This section described the steps used in the pre-processing of the data on the RETOUCH, DUKE, and OPTIMA, before fitting the images into deep learning architectures. SD-OCT volumes were processed as follows:

1) SLICES EXTRACTION

All volumes were sliced in 2.5D. Fig. 2 shows the methodology of slicing the B-scans to integrate auxiliary information from the 3D volume. Table 1 presents the description of each dataset. 2.5D B-scans, as well as 2.5D patches, were extracted to establish equal comparison for each network. Three adjacent B-scans were stacked together to feed the RGB channels of each network, and the ground truth of the middle slice was used as the label mask. This process was repeated throughout the entire volume. Following the method mentioned, volume size for each vendor was changed as the Cirrus scanner volume size changed from (512 x 1024 x 1 x 128) to (512 x 1024 x 3 x 126). Each B-scan was then resized to (384 x 384) using the nearest-neighbor interpolation to feed networks with inputs of dimension (384 x 384 x 3). Ground truth images were then resampled
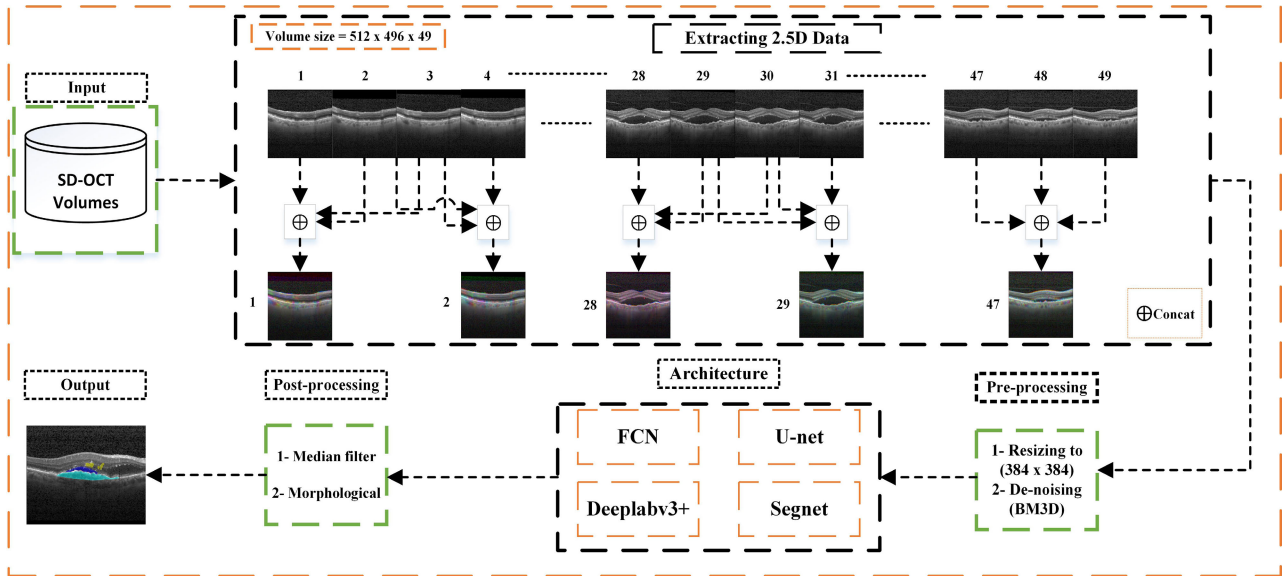
K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

IEEE *Access*

FIGURE 2. Segmentation Pipeline.

TABLE 1. Dataset description.

| Vendor | Size | RETOUCH | | OPTIMA | | DUKE |
		Training	Testing	Training	Testing	
Cirrus	512 x 1024 x 128	24	14	4	4	-
Spectralis	512 x 496 x 49	24	14	4	4	10
Nidek	512 x 512 x 128	-	-	3	3	-
Topcon	512 x 885 x 128	22	14	4	4	-
GT availability		✓	-	✓	✓	✓

(a) Spectralis Image (b) Grader 1 (c) Grader 2 (d) STAPLE
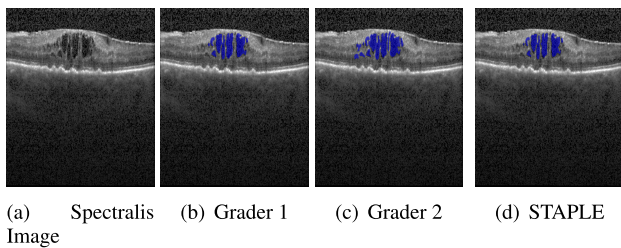
FIGURE 3. Simultaneous truth and performance level estimation process on OPTIMA challenge dataset.

(a) Original B-scan (b) De-noised B-scan

FIGURE 4. BM3D de-noising.

the same way. Also, for the case of fitting patches, the resized images were used to extract patches with the size of (128 x 128 x 3) with an overlapping percentage at 85%. RETOUCH dataset provided single grader annotation, while the OPTIMA and DUKE dataset were released with two graders annotation. Simultaneous truth and performance level estimation (STAPLE) [36] algorithm was applied to the OPTIMA and DUKE dataset references to produce one ground-truth (GT), as presented in Fig. 3. In the DUKE dataset, the size of the released volume is 496 x 768 x 61, and the number of annotated images are 11 per volume. However, the images were not annotated consecutively.

2) SLICE DE-NOISING
Like other medical data, SD-OCT images were prone to noise. All resized images were denoised using
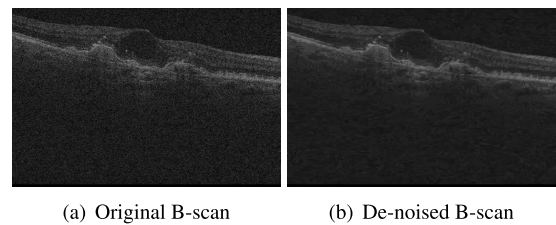
block-matching 3D filtering (BM3D) [37] algorithm. The sigma value applied to all images and patches was set to 25. Fig. 4 shows an example of the denoising method applied on SD-OCT images.

C. DEEP CNN ARCHITECTURES
CNNs were efficiently used for computer vision tasks. There were many proposed architectures for each task, such as Vanilla CNN, which had better performance for classification tasks. For segmentation tasks, two common architecture types were proposed, which are single-path and multi-path. Multi-path networks fused the information from the encoder part, which led to better segmentation outcomes at the cost of computation speed and memory usage. In contrast, single-path networks were faster in the training process. This study utilized four proposals, which were fully convolutional networks (FCN) [33], *U-net* [35], Seg-net [32], and

IEEE Access

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

Deeplabv3+ [34]. The implementation details of all architectures were depicted as follows:

1) FULLY CONVOLUTIONAL NETWORKS (FCN) [33]

Pioneer work in segmentation tasks used deep neural networks. Inspired by popular classification networks, FCN utilized the pre-trained network; VGG16 [38], Alexnet network [39], and GoogLenet network [40]. In this study, the 8s model was used, which is upsampled from the third layer with VGG16 as the backbone architecture. In the last layer of the VGG16 network, skip connections were used to fuse the information with the lower layers. Three different layers were triggered to upsample the information. The last layer was upsampled 32 times, and that model was named 32s model. The model trained with patches was recognized with the superscript Pa as in FCN^{Pa}.

2) U-net [35]

A deep neural network was designed to overcome the data shortage in the biomedical imaging field. The network employed two paths to predict the retinal fluid location, which were the contracting part of obtaining global information, and the expanded path to capture the local information. The combination of the two paths was performed either by addition or concatenation. Patches were extracted in a sliding window pattern, which supplied the network with various shapes. The *U-net* contained four successive convolution layers and four deconvolution layers. The feature map was transferred from the encoder path to the decoder path. After each successive max-pooling, the input was halved to train the network at different scales, and padding was used to retain the exact size of the input. This study used the vanilla *U-net*, and full images and patches were fitted in the network. The patched trained network was defined with the superscript pa as in $U\text{-}net^{Pa}$. Small size fluids were mostly missed. Hence, *U-net* was extended with two more convolution layers to force the network to recognize smaller shapes. The extended *U-net* network in this study was referred to with the subscript Ex as $U\text{-}net_{Ex}$, and the patched network as $U\text{-}net_{Ex}^{Pa}$.

3) SEG-NET [32]

Another architecture was proposed for segmentation purposes that followed the encoder-decoder style and resembled the *U-net* architecture. There were two main differences of this network with the *U-net* network, which was, firstly, the encoder path was replaced with the VGG16 network. Hence, a convolution layer and deconvolution layer were added in comparison with *U-net*. Secondly, the skip connections only transferred pooling indices instead of transferring the full feature map resulting in saving memory usage. Seg-net was also trained with full image resolution and patches, and the patched network was observed with the superscript Pa as $Seg\text{-}net^{Pa}$.

4) Deeplabv3+ [34]

The final network examined in this study was to have a decoder added to the Deeplabv3 model [41], known as Deeplabv3+. This network trained different backbones, for example, ResNet-101 [42] and Xception [43]. The model used depthwise separable convolution operations instead of normal convolution operations. The core idea of this replacement was to cover a wider distance of the same number of parameters and to speed up the training phase. The network reported that depthwise separable convolution operations retained the same or performed better than standard convolution operations, with reduced computation complexity. The model was also trained with patches, whereby the patched network recognized the superscript Pa as $Deeplabv3+^{Pa}$.

D. EXPERIMENTAL DETAILS

This work studied mainly four deep models to analyze the performance of each network on RETOUCH dataset. All models are trained three times, which is equal to 9 fold validation. Each model is trained twice with Adam optimizer options, then the network is trained once with stochastic gradient descent momentum (SGDM). The score of each network is fused to output one network using the majority voting rule. In each run, the model (Seg-net, FCN, Deeplabv3+, or *U-net*) is trained using 3 fold validation, and the summary of each network specifications during the training phase as follows:

- Cross entropy loss function with median frequency class balancing.
- Data augmentation with random rotation between -10 and $+10$ degrees, random mirroring, random translation between -10 and $+10$ degrees, and random magnitude of noise between 0 to 0.35.
- ADAM optimizer with an initial learning rate of 0.001 and 0.0001, and the decay rate of the squared gradient is 0.95 and 0.99.
- SGDM optimizer with an initial learning rate of 0.0001, the learning rate is reduced by 4 after 8 epochs, and momentum is set to 0.9.
- Minibatch size of 8 images and 200 epochs is applied.
- Training is killed using the early stopping policy with patience equal to 15.
- Images are shuffled after each epoch.
- 3 fold validation.
- Each model is trained three times with different parameters (two networks with ADAM optimizer and one network with SGDM optimizer), and the majority vote is performed.
- The average training phase duration for each network is 49 hours.
- The average testing phase duration for each network to segment a volume is almost 12.4 seconds.

The average training phase time in hours for Deeplabv3+, *U-net*, $U\text{-}net_{Ex}$, FCN, and Seg-net is 41, 43, 49, 53, and 61, respectively. The average testing phase time in seconds for Deeplabv3+, *U-net*, $U\text{-}net_{Ex}$, FCN, and Seg-net is 9.6, 11.9, 13.6, 13.8, and 13, respectively. Fig. 5 shows the model fusing criteria used in this work.

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

IEEE *Access*

TABLE 2. Network performance evaluation On 2D input over RETOUCH datasets using dice score metric [44].

Scanner	Disease	Network									
		FCN	U-net	Seg-net	Deeplabv3+	$U\text{-net}_{Ex}$	FCN^{Pa}	$U\text{-net}^{Pa}$	$Seg\text{-net}^{Pa}$	$Deeplabv3+^{Pa}$	$U\text{-net}^{Pa}_{Ex}$
Cirrus	IRF	0.58	0.43	0.74	0.73	0.65	0.62	0.78	0.69	**0.75**	**0.81**
	SRF	0.42	0.5	0.71	0.64	0.59	0.63	0.59	0.71	**0.78**	**0.74**
	PED	0.51	0.29	0.68	0.69	0.71	0.56	0.66	0.73	**0.7**	**0.62**
Spectralis	IRF	0.34	0.24	0.65	0.55	0.59	0.45	0.59	0.68	**0.89**	**0.88**
	SRF	0.55	0.11	0.71	0.46	0.58	0.38	0.42	0.51	**0.76**	**0.73**
	PED	0.46	0.38	0.59	0.63	0.67	0.4	0.66	0.62	**0.73**	**0.81**
Topcon	IRF	0.43	0.34	0.49	0.49	0.62	0.56	0.53	0.51	**0.78**	**0.71**
	SRF	0.62	0.46	0.53	0.67	**0.92**	**0.85**	0.62	0.72	**0.84**	**0.85**
	PED	0.48	0.25	0.43	0.63	0.67	0.51	0.54	0.78	**0.83**	**0.79**
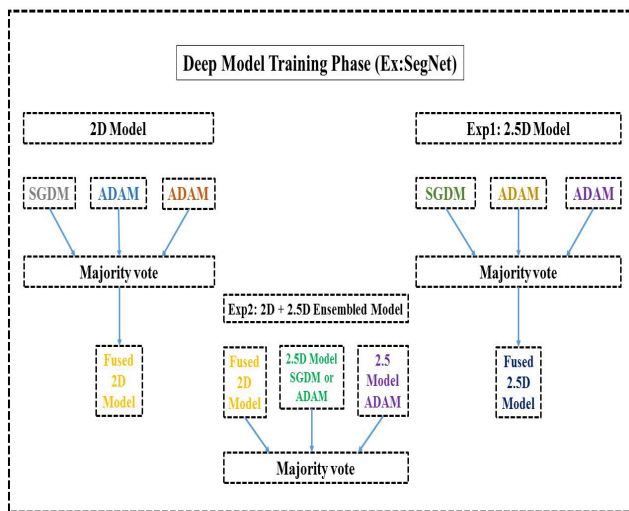
FIGURE 5. Training procedure.

E. POST-PROCESSING

This work employs two pre-processing operations. A median filter with the size of (3 x 3) is applied to refine the output. A closing operation is applied with a radius value of 4 pixels to fill up the holes.

F. DATA EVALUATION

Deep architectures are evaluated using the dice similarity coefficient score (DSC).

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (1)$$

TP indicates the true-positive pixels, FP symbolizes the false-positive pixels, and FN represents the false-negative pixels. Furthermore, the Wilcoxon signed-rank test is considered to evaluate the significance among deep neural models statistically.

IV. RESULTS

The evaluation of this study used the RETOUCH challenge dataset. In this study, the effect of using 2.5D deep networks over 2D deep networks was examined. Additionally, 2D networks with 2.5D networks were ensembled. The best performing networks were used on the OPTIMA challenge dataset as well as the DUKE dataset. No extra data was added

to the networks during the training phase in all performed experiments.

A. NETWORKS TRAINED WITH 2.5D INPUT

This section described the effects of feeding 2.5D data or 2D data to deep models. Table 2 shows the performance of ten networks trained using 2D input [44]. Each network was trained in 3 fold validation, and the score of the network was preserved. This process was repeated thrice with different hyper-parameters, as described in the experimental details section. The resulting scores from the three networks (of the same model, Ex: Seg-net) were fused using the majority voting rule. Hence, in most of the networks, IRF was identified better than other fluids on Cirrus and Spectralis data. SRF was localized well in the $U\text{-}net_{Ex}$ network with DSC = 0.92. Most networks performed better on Topcon data than Cirrus data but significantly better than Spectralis data (p = 0.04). The overlapping patches with the rate of 85% led to better performance in DSC values compared to lower overlapping ratios. The networks trained in full image input performed lower than networks trained with patches. Under other conditions, Table 3 presents the performance of networks trained using 2.5D input, including some depth information to the network that led to better DCS values.

Similar to the 2D network training method, each network was trained thrice using various parameters, and the majority voting technique was applied. As observed in experiment one, most networks enhanced the overall performance with a slight increment in DSC values compared to the networks trained with only 2D information. Also, many networks reduced the gap in DSC values for fluid detection. For example, the output of $Deeplabv3+^{Pa}$ network on 2D input using Spectralis data was 0.89, 0.76, and 0.73 for IRF, SRF, and PED segmentation, respectively. However, the output of $Deeplabv3+^{Pa}$ network on 2.5D input using Spectralis data was 0.86, 0.82, and 0.84 for IRF, SRF, and PED segmentation, respectively. Additionally, in the Seg-net network, the output on 2D input over Spectralis data was 0.65, 0.71, and 0.59 for IRF, SRF, and PED segmentation, respectively. On the other hand, the output of Seg-net network on 2.5D input over Spectralis data was 0.68, 0.70, and 0.66 for IRF, SRF, and PED segmentation, respectively. In 2.5D networks, most networks performed better on Topcon data compared to Cirrus data.
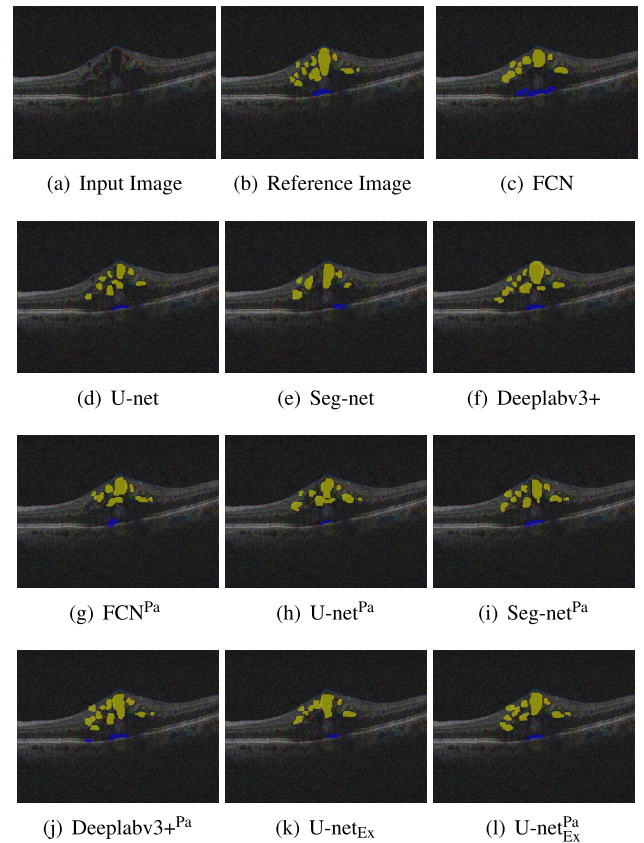
IEEE Access·

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

TABLE 3. *Exp1* - Network performance evaluation on 2.5D input over RETOUCH datasets using dice score metric.

Scanner	Disease	Network									
		FCN	U-net	Segnet	Deeplabv3+	$U-net_{Ex}$	FCN^{Pa}	$U-net^{Pa}$	$Segnet^{Pa}$	$Deeplabv3+^{Pa}$	$U-net_{Ex}^{Pa}$
Cirrus	IRF	0.52	0.38	0.76	0.75	0.72	0.58	0.75	0.78	**0.78**	**0.71**
	SRF	0.49	0.47	0.73	0.7	0.69	0.55	0.68	0.77	**0.81**	**0.75**
	PED	0.55	0.34	0.7	0.72	0.75	0.58	0.74	0.68	**0.76**	**0.73**
Spectralis	IRF	0.41	0.29	0.68	0.72	0.66	0.52	0.62	0.74	**0.86**	**0.75**
	SRF	0.48	0.26	0.7	0.77	0.62	0.46	0.56	0.76	**0.82**	**0.69**
	PED	0.49	0.35	0.66	0.69	0.78	0.57	0.69	0.7	**0.84**	**0.82**
Topcon	IRF	0.55	0.41	0.55	0.65	0.68	0.56	0.58	0.66	**0.8**	**0.76**
	SRF	0.57	0.49	0.62	0.71	0.82	0.67	0.66	0.75	**0.88**	**0.88**
	PED	0.52	0.37	0.5	0.59	0.74	0.62	0.62	**0.81**	0.79	**0.81**

Networks performed significantly better in Topcon data than Spectralis data with ($p = 0.0487$). $Deeplabv3+^{Pa}$ network and U-net_{Ex}^{Pa} network outperformed other networks significantly in this study. $Deeplabv3+^{Pa}$ network outperformed U-net_{Ex}^{Pa} network with ($p = 0.164$). $Deeplabv3+^{Pa}$ recognized fluids better in Spectralis data over other scanners, while U-net_{Ex}^{Pa} recognized fluids in Topcon data over other networks. In 2.5D networks using Cirrus data, there was no favorite fluid to be recognized. In Spectralis data, PED was recognized more than other fluids. Lastly, SRF was more recognized by networks using Topcon data.

B. NETWORK ENSEMBLING OF 2.5D AND 2D DATA

Experiment two is shown in Table 5, whereby 2D networks were ensembled along with 2.5D networks from the same model. The network ensembling in this study was only for the fused 2D and 2.5D networks. In experiment two, three networks were ensembled, with one network trained with 2D input, and two networks trained with 2.5D input. Fig. 5 describes the training procedure for network's ensembling. The highest two performing networks score from 2.5D input were selected. The three networks were ensembled using the majority vote. To be noted, there was no fusion between, such as U-net and Seg-net. The fusion occurs only between the networks trained from the same model. The overall performance was enhanced in the ensembled networks, which outperformed networks trained with 2D or 2.5D input separately but not significantly ($p = 0.16$). The highest average DSC score was 0.86 for $Deeplabv3+^{Pa}$ network trained on Topcon data. IRF was most recognized with 0.924 DSC value in the ensembled networks. In comparison to 2.5D networks, the ensembled $Deeplabv3+^{Pa}$ network performed significantly better than U-net_{Ex}^{Pa} model ($p = 0.048$) and other networks with an average p-value of 0.028. The ensembled network outputs for each model over RETOUCH dataset are represented in Fig. 6 for Cirrus data, Fig. 7 for data, and Fig. 8 for Topcon data. Since the challenge organizers provided no reference images to the testing set, the outcome of the best two ensembled networks on some selected images were selected from the three scanners. The ensembled $Deeplabv3+^{Pa}$ network and U-net_{Ex}^{Pa} network are evaluated on the RETOUCH testing dataset in Fig. 9. In all these figures, the colors represented different fluids, with yellow showing IRF, dark blue depicting SRF fluids, and light blue illustrating PED.

(a) Input Image	(b) Reference Image	(c) FCN
(d) U-net	(e) Seg-net	(f) Deeplabv3+
(g) FCN^{Pa}	(h) U-net^{Pa}	(i) Seg-net^{Pa}
(j) $Deeplabv3+^{Pa}$	(k) U-net_{Ex}	(l) U-net_{Ex}^{Pa}

FIGURE 6. Network output over Cirrus RETOUCH training dataset.

C. ENSEMBLED NETWORKS ON OPTIMA DATASET

Table 6 illustrates the performance of ensembled networks on the OPTIMA dataset, which were fine-tuned using the training dataset of the OPTIMA challenge on Cirrus, Spectralis, and Topcon scanners. The network that achieved the highest DSC value from any scanner (Cirrus, Spectralis, and Topcon) in experiment two was fine-tuned again using the Nidek training dataset. The OPTIMA challenge released two testing datasets, which was combined in this study, and reported as one set. Although the ensembled networks were trained to distinguish three fluids, OPTIMA datasets were aimed to only segment IRF fluids. Thus, the output of any network established the pixels identified as SRF or PED to zero. Also, in experiment three, $Deeplabv3+^{Pa}$ network and U-net_{Ex}^{Pa} network outperformed all other networks, as well as

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

IEEE *Access*

TABLE 4. Summary of the state-of-the-art deep learning models for retinal diseases segmentation.

Reference	OCT scanner	Data size	Pre-processing				Fluid type	Model	Evaluation	2D/3D	Results	
			De-noise	Data Aug	Lay Seg	Cropping					Training	Testing
Gopinath et al. [21]	Cirrus, Nidek, Optovue, Spectralis, Topcon	50	✓	✓	✓		IRF	U-net	DSC	3D		0.69
Girish et al. [20]	Cirrus, Nidek, Spectralis, Topcon	30		✓	✓		IRF	U-net	DSC	2D		0.74
Chiu et al. [5]	Spectralis	10	✓	✓	✓		IRF	Kernel Regression	DSC	2D		0.53, 0.45
Roy et al. [19]	Spectralis Spectralis, Topcon	10			✓		IRF	U-net	DSC	2D		0.81
Helios [22]	Cirrus, Topcon, Spectralis	112	✓			✓	IRF, SRF, PED	U-net	DSC	2D	0.61, 0.70, 0.73	0.63, 0.68, 0.66
RetinaAI [23]	Cirrus, Topcon, Spectralis	112		✓	✓		IRF, SRF, PED	U-net	DSC	2D	0.59, 0.55, 0.77	0.74, 0.67, 0.71
MABIC [24]	Cirrus, Topcon, Spectralis	112		✓			IRF, SRF, PED	U-net	DSC	2D	0.86, 0.93, 0.91	0.78, 0.66, 0.70
UMN [18]	Cirrus, Topcon, Spectralis	112			✓		IRF, SRF, PED	CNN	DSC	2D	0.85, 0.82, 0.84	0.70, 0.71, 0.77
SFU [25]	Cirrus, Topcon, Spectralis	112		✓	✓		IRF, SRF, PED	U-net	DSC	2D	0.77, 0.83, 0.76	0.82, 0.75, 0.74
Venhuizen et al. [16]	Cirrus, Topcon, Spectralis, Nidek	30			✓		IRF	CNN	DSC	2D		0.64
RMIT [26]	Cirrus, Topcon, Spectralis	112		✓			IRF, SRF, PED	U-net adversarial	DSC	3D	0.69, 0.67, 0.85	0.73, 0.70, 0.69
Xu et al. [28]	Topcon	30	✓		✓		IRF	stratified sampling	recall	3D		0.97, 0.19
Chen et al. [27]	not reported	15			✓		all	Constrained Graph-Search	DSC	2.5D		0.81
Wang et al. [29] 0.81	Optovue	10			✓		IRF	Fuzzy level-set	Jacard metric	2.5D		
Sun et al. [45]	Topcon	25	✓		✓		PED	shape-constrained	DSC	2.5D		0.91
UCF [30]	Cirrus, Topcon, Spectralis	112		✓		✓	IRF, SRF, PED	ED-ResNet	DSC	2.5D	0.52, 0.68, 0.61	0.49, 0.54, 0.63
NJUST [31]	Cirrus, Topcon, Spectralis	112	✓		✓		IRF, SRF, PED	Faster R-CNN	DSC	2.5D	0.99, 0.99, 0.99	0.56, 0.53, 0.64

IEEE *Access*

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

(a) Input Image   (b) Reference Image   (c) FCN

(d) U-net   (e) Seg-net   (f) Deeplabv3+

(g) FCN^{Pa} (h) U-net^{Pa} (i) Seg-net^{Pa}

(j) Deeplabv3+^{Pa} (k) U-net$_{Ex}$ (l) U-net$_{Ex}^{Pa}$

FIGURE 7. Network output over Spectralis RETOUCH training dataset.

(a) Input Image   (b) Reference Image   (c) FCN

(d) U-net   (e) Seg-net   (f) Deeplabv3+

(g) FCN^{Pa} (h) U-net^{Pa} (i) Seg-net^{Pa}

(j) Deeplabv3+^{Pa} (k) U-net$_{Ex}$ (l) U-net$_{Ex}^{Pa}$

FIGURE 8. Network output over Topcon RETOUCH training dataset.

(a) Cirrus Image   (b) U-net$_{Ex}^{Pa}$ (c) Deeplabv3+^{Pa}

(d) Spectralis Image   (e) U-net$_{Ex}^{Pa}$ (f) Deeplabv3+^{Pa}

(g) Topcon Image   (h) U-net$_{Ex}^{Pa}$ (i) Deeplabv3+^{Pa}
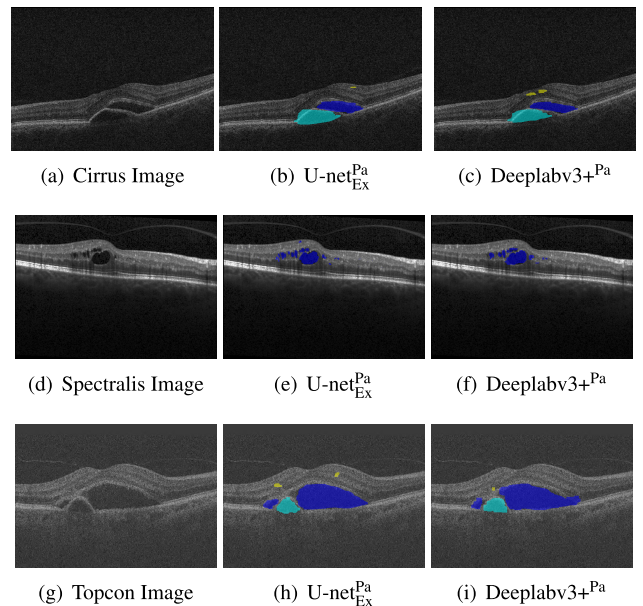
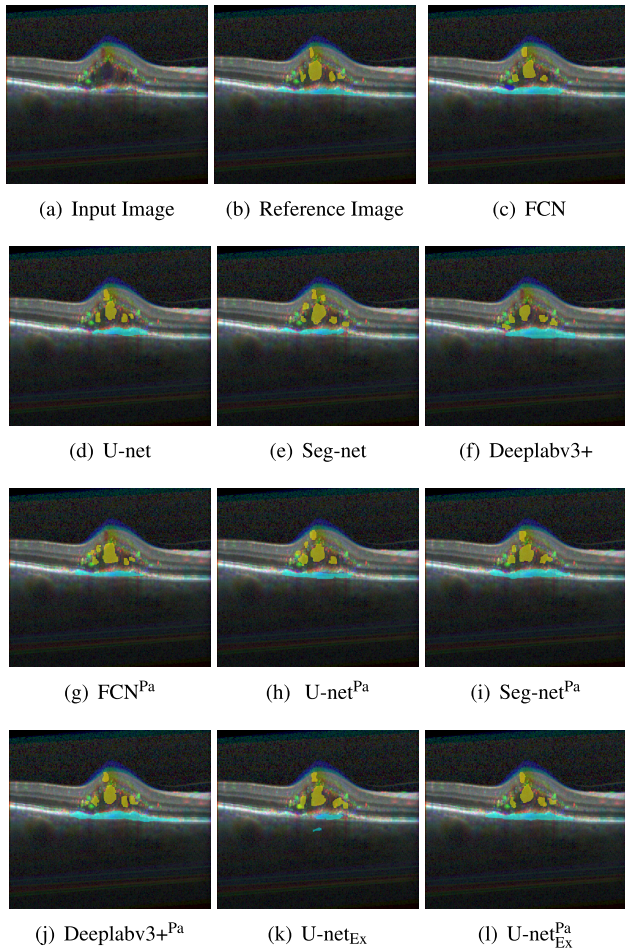**FIGURE 9.** Network output over the RETOUCH testing dataset.

teams, contributing to the challenge. IRF was recognized well in Topcon data using a *Deeplabv*3+^{Pa} network. In the same dataset, the *U-net*$_{Ex}^{Pa}$ network recognized IRF higher than other models. Extracting patches led to better recognition of fluids, and this could be explained as providing more data with various shapes helped in improving the efficiency of the learning phase. Fig. 11 shows an example of the two highest-performing network output over OPTIMA datasets.

D. DUKE DATASET EXPERIMENTS

In Table 7, three different experiments are applied to the DUKE dataset. First, the highest performing networks from *Exp2* on the Spectralis dataset are evaluated directly on the 110 images from the DUKE dataset. No fine-tuning (NFT) method was applied and tested to the data for the first row in *Exp4*, in which the Duke images are fitted directly to the resulted networks from *Exp2*. After that, a fine-tuning (FT) method was performed to push the performance of the networks. In the fine-tuning stage, the data were split into 50% for training and 50% for testing. At the fine-tuning stage, the networks were trained using the same method on the RETOUCH dataset, which were trained thrice. The deep models were fine-tuned twice with ADAM optimizer param-
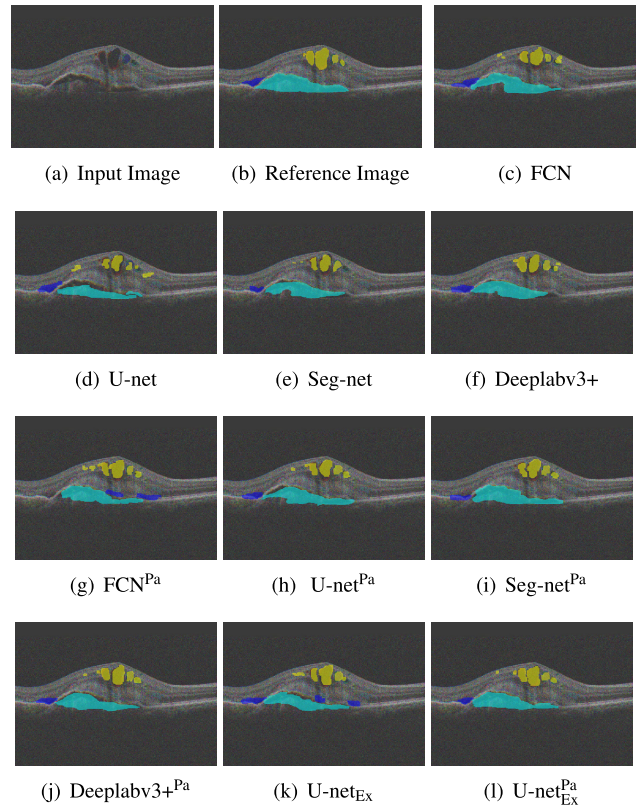
eters and once with SGDM optimizer. Last, on the third row of the Table 7, the models were trained from scratch (SC). The same steps were repeated for the data and training procedures as the FT stage using 3 fold validation.

In *Exp4*, 2.5D input to networks was used. *Deeplabv*3+^{Pa} network has outperformed other networks significantly

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

IEEE *Access*

TABLE 5. *Exp2* - Network performance evaluation on Ensembled 2D input and 2.5D input over RETOUCH datasets using dice score metric.

Scanner	Disease	Network									
		FCN	U-net	Seg-net	Deeplabv3+	U-net_{Ex}	FCN^{Pa}	U-net^{Pa}	Seg-net^{Pa}	Deeplabv3+^{Pa}	U-net_{Ex}^{Pa}
Cirrus	IRF	0.56	0.48	0.77	0.73	0.67	0.63	0.72	0.72	**0.77**	**0.79**
	SRF	0.53	0.52	0.78	0.69	0.69	0.59	0.71	0.7	**0.83**	**0.73**
	PED	0.6	0.45	0.75	0.76	0.62	0.61	0.68	0.75	**0.81**	**0.78**
Spectralis	IRF	0.46	0.4	0.72	0.75	0.64	0.59	0.66	0.67	**0.84**	**0.77**
	SRF	0.55	0.35	0.76	0.76	0.61	0.53	0.62	**0.81**	**0.83**	**0.81**
	PED	0.55	0.46	0.71	0.73	0.75	0.62	0.64	0.71	**0.79**	**0.72**
Topcon	IRF	0.53	0.48	0.61	0.68	0.72	0.6	0.61	0.62	**0.92**	**0.75**
	SRF	0.59	0.52	0.6	0.73	**0.8**	0.65	0.64	0.73	**0.81**	**0.81**
	PED	0.49	0.44	0.62	0.66	0.76	0.64	0.68	0.75	**0.84**	**0.76**

TABLE 6. *Exp3* - Network performance evaluation on 2.5D input over OPTIMA datasets using dice score metric.

Scanner	Network									
	FCN	U-net	Seg-net	Deeplabv3+	U-net_{Ex}	FCN^{Pa}	U-net^{Pa}	Seg-net^{Pa}	Deeplabv3+^{Pa}	U-net_{Ex}^{Pa}
Cirrus	0.53	0.39	0.48	**0.62**	**0.61**	0.58	0.49	0.55	**0.72**	0.67
Spectralis	0.58	0.41	0.54	**0.62**	0.59	**0.62**	0.53	0.58	**0.7**	0.68
Topcon	0.59	0.48	**0.62**	**0.67**	0.5	**0.62**	0.55	**0.65**	**0.78**	0.77
Nidek	0.5	0.44	0.53	0.55	0.49	0.53	0.49	0.55	**0.62**	0.57

TABLE 7. *Exp4* - Network performance evaluation on DUKE dataset using dice score metric.

Scanner	Network									
	FCN	U-net	Seg-net	Deeplabv3+	U-net_{Ex}	FCN^{Pa}	U-net^{Pa}	Seg-net^{Pa}	Deeplabv3+^{Pa}	U-net_{Ex}^{Pa}
Spectralis (NFT)	0.45	0.41	0.51	**0.62**	0.51	0.49	0.48	0.50	**0.68**	0.59
Spectralis (FT)	0.53	0.48	**0.60**	**0.71**	**0.63**	0.51	0.57	**0.66**	**0.80**	**0.72**
Spectralis (SC)	0.48	0.41	0.52	0.62	0.47	0.48	0.51	0.58	**0.72**	**0.67**

in all experiments using the DUKE dataset. Fine-tuned (FT) $Deeplabv3+^{Pa}$ network outperformed the trained from scratch (SC) $Deeplabv3+^{Pa}$ network significantly with (p = 0.024). The ensembled models are performing better when fine-tuned over OPTIMA and DUKE datasets. This led to the fact that incorporating some depth information could enhance the performance of deep networks. $U\text{-}net_{Ex}^{Pa}$ network has shown a competitive performance to the segmentation of the retinal fluids when compared to the $Deeplabv3+^{Pa}$ network.

V. DISCUSSION

In this study, ten networks have been trained to distinguish various retinal fluids on three available benchmarks. Three main experiments are assessed in this study to show the importance of (i) 2.5D data input over the 2D data to neural models, (ii) 2.5D data ensembled with 2D data to enhance the accuracy, and (iii) networks previously trained on RETOUCH dataset being fine-tuned. The first experiment trained ten networks with 2.5D input, which are shown in Table 3, and the comparison of the results are presented in Table 2. The overlapping patches performed better than the input of full images in the deep models. The $U\text{-}net_{Ex}$ network has performed higher compared to networks trained with patches, which may be caused by the various scales each image has been trained in a network, resulting in each image being down-sampled six times.

The $Deeplabv3+^{Pa}$ network specifically outperformed all other networks because it used the Xception model as a backbone architecture. The network also adopts atrous spatial pyramid pooling (ASPP) to train an image in several parallel scales to capture the contextual information. Hence, the input of patches assists a network in feeding various shapes of the original image, as well as the network that trained the patches at different scales. $U\text{-}net_{Ex}^{Pa}$ network has performed relatively well compared to $Deeplabv3+^{Pa}$. The extracted patches have filled some gaps in the output images, especially in the areas predicted to be homogeneous. Still, some gaps have not been filled, which further emphasizes the need to apply morphological operations. Finally, random noisy pixels are found to be diagnosed as fluids outside the retinal layers once removed by the median filter. The second experiment ensembled the 2D networks and 2.5D networks. Three networks are ensembled to segment the retinal fluids, with one network from 2D input and two networks from 2.5D input. Ensembled networks enhanced fluid recognition, though not statistically significant. $Deeplabv3+^{Pa}$ network has outperformed all other networks, and IRF fluid is recognized more than other fluids. The last experiment presents the performance of ensembled networks over the testing set released to the OPTIMA challenge. The ensembled networks are fine-tuned using the training set from each network. For any model, the best performing network on Cirrus, Spectralis, and Topcon data is further fine-tuned with Nidek data. $Deeplabv3+^{Pa}$ network outperformed other networks participating in the challenge as in [16]. Venhuizen *et al.* scored a 0.64 to segment the IRF fluids. Besides, Girish *et al.* [20] used a modified $U\text{-}net$ model with the use of depthwise separable convolution

IEEE *Access*

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

(a) Spectralis Image   (b) Grader 1   (c) Grader 2

(d) STAPLE   (e) Deeplabv3+^{Pa} (f) U-net$_{Ex}^{Pa}$
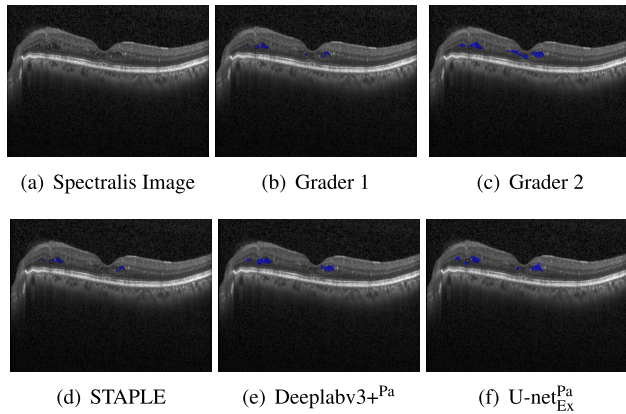
FIGURE 10. Network output over the Duke dataset.

operations in the OPTIMA dataset, and the average DSC score is 0.74. The best DSC score of all networks belongs to the *Deeplabv3+^{Pa}* model, which scores 0.78 on the Topcon dataset.

All results are compared with the eight teams that are associated with the training set of RETOUCH challenge, as shown in Table 4. The models are reported in Table 4 evaluated their models using the DSC metric. The winner of the RETOUCH challenge is SFU group [46], which attached an extra channel to the B-scan. The attached channel represents the distance map of each B-scan intensity. The SFU model performed well in the training and testing set compared to the NJUST [31] model. The NJUST group has segmented the IRF fluid using the Faster R-CNN model, region growing to segment SRF fluid, and RPE layer to segment the PED. The NJUST model, which utilized 2.5D methods, performed very well in the training set but poorly in the testing set. Another group that utilized the 2.5D information is the UCF group [30]. In [30], a decoder-encoder ResNet model is employed, and their model performed low. Moreover, the UMN group [47] proposed a CNN model, which performed better in the training set (average DSC for all fluid segmentation = 0.83) in comparison to the testing datasets (average DSC for all fluid segmentation = 0.73). Additionally, MABIC group [24] trained two consecutive U-net models, and performed better in the training datasets (average DSC for all fluid segmentation = 0.90) compared to the testing set (average DSC for all fluid segmentation = 0.71). Thus, releasing the reference for the testing set could be more helpful to draw a better conclusion. Moreover, most of the methods show stable performance in both datasets, yet their performance is low as in [30].

RMIT group [26] utilized 3D information, and the U-net model is linked with an adversarial model. The RMIT model has performed stably in both training and testing datasets. Most of the models utilized the U-net model, which some methods performed well in the training set and performed low in the testing set (maybe due to overfitting), or performed low in both datasets. In this work, the best performing model is *Deeplabv3+^{Pa}*, and the average DSC for all

(a) Cirrus Image   (b) STAPLE   (c) U-net$_{Ex}^{Pa}$ (d) Deeplabv3+^{Pa}

(e) Nidek Image   (f) STAPLE   (g) U-net$_{Ex}^{Pa}$ (h) Deeplabv3+^{Pa}

(i) Spectralis Image   (j) STAPLE   (k) U-net$_{Ex}^{Pa}$ (l) Deeplabv3+^{Pa}

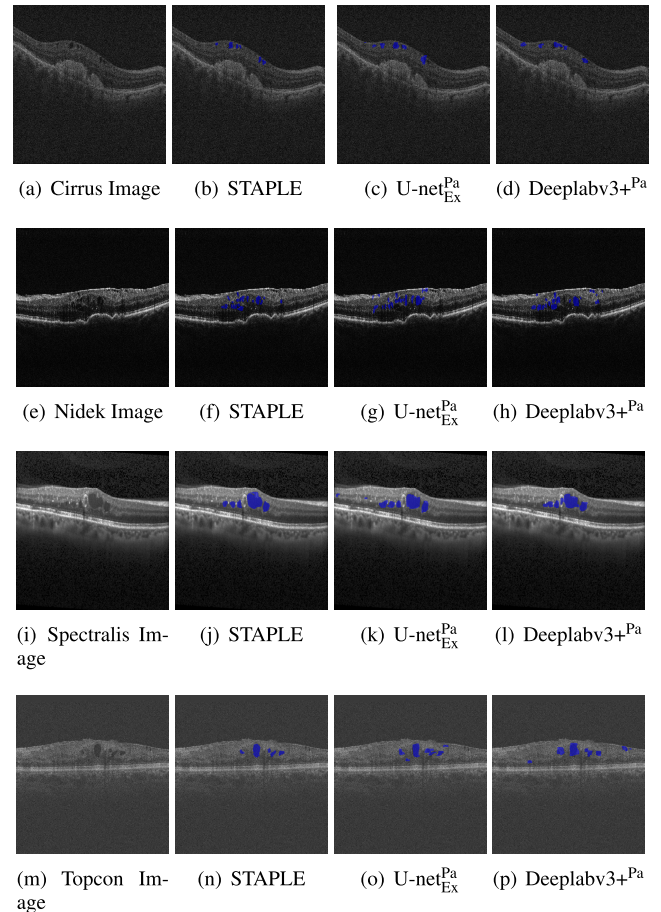(m) Topcon Image   (n) STAPLE   (o) U-net$_{Ex}^{Pa}$ (p) Deeplabv3+^{Pa}

FIGURE 11. Network output over OPTIMA testing dataset.

fluid segmentation is 0.82. We believe that our implemented models will show a slight decrement or no decrement in the DSC values if they are applied on the RETOUCH challenge testing datasets, due to the network ensembling. This study has shown the network's performance over the testing data in Fig. 9, whereby some IRF fluids are recognized wrongly.

In experiment four, the networks have shown a competitive performance over the DUKE dataset. The *Deeplabv3+^{Pa}* network has outperformed [5] and [18] but not [19]. In [19], the network proposed has scored 0.81 when trained using 80% of the data and tested using the remaining of the data. The authors in [19] did not follow the norm of [5] in terms of training and testing criteria. RelayNet [19] outperformed *Deeplabv3+^{Pa}* network but not significantly (p = 0.067). Fine-tuning the ensembled networks outperformed the networks trained on (NFT) and (SC). The data were not annotated consecutively, hence including some depth information was not utilized properly. Each 2.5D input was not linked to the next image, as only 11 images are annotated out of 61 images provided in a volume. Thus, the FT ensembled networks might perform better if all images containing diseased pixels were annotated. Fig. 10 shows an example of the two highest network performance segmentation results along with the expert 1, expert 2 and STAPLE reference images.

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

IEEE *Access*

VI. CONCLUSION

In this study, ten deep learning networks that are ensembled in different dimensionality were examined. Each model is trained to detect IRF, SRF, and PED fluids on RETOUCH datasets and to detect IRF fluid only on OPTIMA and DUKE datasets. Networks trained in a 2.5D manner outperformed networks trained in a 2D manner. The combination of information from different dimensionality led to a better output. Even though feeding the networks with full images accompanied by the augmented images, patched images performed significantly better. The networks that are examined in this study show a competitive performance of the teams that joined both challenges. *Deeplabv*3+^{Pa} network is found to have outperformed other networks on most of the datasets. Data normalization and layers segmentation, however, have not been considered in this study. The use of full 3D information is also not utilized, which could be an aspect to be explored in future studies.

REFERENCES

[1] C. A. Curcio, N. E. Medeiros, and C. L. Millican, "Photoreceptor loss in age-related macular degeneration," *Investigative Ophthalmol. Vis. Sci.*, vol. 37, no. 7, pp. 1236–1249, 1996.

[2] W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C.-Y. Cheng, and T. Y. Wong, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis," *Lancet Global Health*, vol. 2, no. 2, pp. e106–e116, Feb. 2014.

[3] H. Bogunović *et al.*, "RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1858–1874, Aug. 2019.

[4] J. Wu, A.-M. Philip, D. Podkowinski, B. S. Gerendas, G. Langs, C. Simader, S. M. Waldstein, and U. M. Schmidt-Erfurth, "Multivendor spectral-domain optical coherence tomography dataset, observer annotation performance evaluation, and standardized evaluation framework for intraretinal cystoid fluid segmentation," *J. Ophthalmol.*, vol. 2016, pp. 1–8, Jun. 2016.

[5] 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, 2015.

[6] M. R. Hee, "Optical coherence tomography of the eye," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep., 1997. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=M.+R.+Hee%2C+%E2%80%98%E2%80%98Optical+coherence+tomography+of+the+eye%2C%E2%80%99%E2%80%99+Ph.D.+dissertation%2C&btnG=

[7] 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. 1178–1181, 1991.

[8] F. G. Venhuizen, B. van Ginneken, B. Liefers, M. J. van Grinsven, S. Fauser, C. Hoyng, T. Theelen, and C. I. Sánchez, "Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks," *Biomed. Opt. Express*, vol. 8, no. 7, pp. 3292–3316, 2017.

[9] K. Alsaih, G. Lemaitre, M. Rastgoo, J. Massich, D. Sidibé, and F. Meriaudeau, "Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images," *Biomed. Eng. OnLine*, vol. 16, no. 1, p. 68, Dec. 2017.

[10] D. C. Fernandez, "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.

[11] X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," *IEEE Access*, vol. 7, pp. 3046–3061, 2019.

[12] L. Bekalo, S. Niu, X. He, P. Li, I. P. Okuwobi, C. Yu, W. Fan, S. Yuan, and Q. Chen, "Automated 3-D retinal layer segmentation from SD-OCT images with neurosensory retinal detachment," *IEEE Access*, vol. 7, pp. 14894–14907, 2019.

[13] X. Li, L. Shen, M. Shen, and C. S. Qiu, "Integrating handcrafted and deep features for optical coherence tomography based retinal disease classification," *IEEE Access*, vol. 7, pp. 33771–33777, 2019.

[14] 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.

[15] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Med.*, vol. 24, no. 9, p. 1342, 2018.

[16] F. G. Venhuizen, M. Grinsven, C. B. Hoyng, T. Theelen, B. Ginneken, and C. I. Sanchez, "Vendor independent cyst segmentation in retinal SD-OCT volumes using a combination of multiple scale convolutional neural network," in *Proc. Med. Image Comput. Comput. Assist. Intervent.-Challenge Retinal Cryst. Segmentation*, 2015, pp. 1–8.

[17] K. Li, X. Wu, D. Z. Chen, and M. Sonka, "Optimal surface segmentation in volumetric Images–A graph-theoretic approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 119–134, Jan. 2006.

[18] 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.

[19] 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, 2017.

[20] G. N. Girish, B. Saikumar, S. Roychowdhury, A. R. Kothari, and J. Rajan, "Depthwise separable convolutional neural network model for intra-retinal cyst segmentation," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2027–2031.

[21] K. Gopinath and J. Sivaswamy, "Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 273–282, Jan. 2019.

[22] 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, pp. 1–8.

[23] S. Apostolopoulos, C. Ciller, R. Sznitman, and S. De zanet, "Simultaneous classification and segmentation of cysts in retinal OCT," in *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*, Sep. 2017, pp. 22–29.

[24] 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, pp. 9–14.

[25] D. Lu, M. Heisler, S. Lee, G. W. Ding, E. Navajas, M. V. Sarunic, and M. F. Beg, "Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network," *Med. Image Anal.*, vol. 54, pp. 100–110, May 2019.

[26] 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, pp. 30–37.

[27] X. Chen, M. Niemeijer, L. Zhang, K. Lee, M. D. Abràmoff, 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. 1521–1531, Aug. 2012.

[28] X. Xu, K. Lee, L. Zhang, M. Sonka, and M. D. Abràmoff, "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.

[29] 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. 1577–1589, 2016.

[30] 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, pp. 1–9.

[31] 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)*, Sep. 2017, pp. 15–21.

IEEE *Access*

K. Alsaih *et al.*: Retinal Fluid Segmentation Using Ensembled 2-Dimensionally and 2.5-Dimensionally Deep Learning Networks

[32] 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.

[33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[34] 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.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28

[36] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, Jul. 2004.

[37] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[40] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[41] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[42] 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.

[43] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1251–1258.

[44] K. Alsaih, M. Yusoff, T. Tang, I. Faye, and F. Mériaudeau, "Deep learning architectures analysis for age-related macular degeneration segmentation on optical coherence tomography scans," *Comput. Methods Programs Biomed.*, p. 105566, 2020.

[45] 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, no. 1, p. 21739, Feb. 2016.

[46] D. Lu, M. Heisler, S. Lee, G. Ding, M. V. Sarunic, and M. F. Beg, "Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network," in *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*, 2017, pp. 1–9.

[47] A. Rashno, D. Koozekanani, and 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, pp. 54–62.
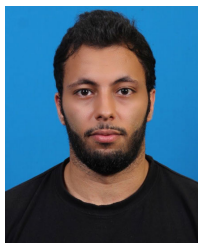
KHALED ALSAIH received the joint master's degree in computer vision, robotics, and mechatronics from the University of Technology Malaysia and Universite Bourgogne Franche-Comté, France, in 2016, and the Ph.D. degree in electrical and electronics engineering from Universiti Teknologi PETRONAS (UTP), Malaysia, in 2020. His research interests include artificial intelligence, retinal disease segmentation and classification, image and signal processing, medical imaging, and deep learning.

MOHD ZUKI YUSOFF (Member, IEEE) received the B.Sc. degree in electrical engineering from Syracuse University, in 1988, the M.Sc. degree in communications, networks, and software from Surrey University, in 2001, and the Ph.D. degree in electrical and electronic engineering from Universiti Teknologi PETRONAS (UTP), Malaysia, in 2010. He is currently an Associate Professor with UTP. He has international publications and holds patents. His research interests include transport safety and telecommunication. He is a member of the Tau Beta Pi and the Eta Kappa Nu.

IBRAHIMA FAYE (Senior Member, IEEE) received the Ph.D. degree in mathematics from the University of Toulouse 3, France, in 2001. He is currently an Associate Professor with Universiti Teknologi PETRONAS, Malaysia. His research interests include engineering mathematics, signal and image processing, multiresolution analysis, and pattern recognition.

TONG BOON TANG (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees (Hons.) from the University of Edinburgh. He is currently the Director of the Institute of Health and Analytics, Universiti Teknologi PETRONAS. His research interest includes biomedical instrumentation, from device and measurement to data fusion. He serves as the Secretary of the HICoE Council, a Board Member of the INCF, and the Vice Chair of the IEEE Circuits and Systems Society Malaysia Chapter.

FABRICE MERIAUDEAU (Member, IEEE) was born in March 1971. He received the Engineering degree in material sciences and the master's degree in physics, and the Ph.D. degree in image processing from Dijon University, France, in 1994 and 1997, respectively. He was a Postdoctoral Researcher for a year with The Oak Ridge National Laboratory. He was the Director of the Institute Health and Analytics (2017/2018) at the Universiti Teknologi PETRONAS, Malaysia, and the Le2i (UMR CNRS) France, which had more than 200 staff members, from 2011 to 2016. He is currently a Professeur des Universites with the ImVia Research Group, University of Burgundy, where he leads the Artificial Intelligence theme. He coordinated an Erasmus Mundus Master in the field of computer vision and robotics from 2006 to 2010 and was the Vice President for International Affairs for the University of Burgundy, from 2010 to 2012. He has authored or coauthored more than 150 international publications and holds three patents. His research interests include image processing for nonconventional imaging systems (UV, IR, and polarization) and more recently on medical/biomedical imaging.

• • •