This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3185653

A Device-independent, Shape Preserving Retinal Optical Coherence Tomography Image Alignment Method Applying TV-Unet for RPE Layer Detection

Narges Saeedizadeh, Mahnoosh Tajmirriahi, Alireza Haghani, Zahra Amini, Elias Khalili Pour, Hamid Riazi-Esfahani, Kaveh Fadakar, Rahele Kafieh, Hossein Rabbani

***Abstract*— Retinal optical coherence tomography (OCT) images are widely used in diagnosis of ocular conditions. However, random shift and orientation changes of the retinal layers in OCT B-scans yield to appearance variations across the scans. These variations reduce the accuracy of the algorithms applied in the analysis of OCT images. In this study, we propose a preprocessing step to compensate these variations and align B-scans. At first, by incorporating total variation (TV) loss in the well-known Unet model, we propose a TV-Unet model to accurately detect the retinal pigment epithelium (RPE) layer in each B-scan. Then we use the detected RPE layer in the alignment method to form a curvature curve and a reference line. A novel window transferring-based alignment approach is applied to force the curve points to form a straight line, while preserving the shape and size of the pathological lesions. Since detection of RPE layer is a crucial step in the proposed alignment method, we utilized various datasets to train and test the TV-Unet and provided a multimodal, device-independent OCT image alignment method. The TV-Unet localizes the RPE layer in OCT images with low boundary error (maximum of 1.94pixels) and high Dice coefficient (minimum of 0.98). Quantitative and qualitative results indicated that the proposed method can efficiently detects the RPE layer and align OCT images while preserving the structure and size of the retinal lesions (biomarkers) in the OCT scans.**

***Index* Optical coherent tomography (OCT), RPE detection, alignment, TV-Unet, Device-independent**

I. Introduction

Optical Coherence Tomography (OCT) is an indispensable supplementary tool for ophthalmologists in detection and treatment of the retinal disorders [1]. Till now, automatic algorithms have been proposed for OCT image processing. However, one of the main issues in these algorithms is random shift and curvature changes of the retinal layers in OCT images (B-scans) caused by anatomical structures and acquisition distortions [2]. For example, in myopic eyes, the macula might have an aberrant shape due to the presence of strongly curved concave staphylomata. Patients with a higher myopic spherical equivalent, a longer axial length, or an elevated lesion such as a choroidal tumor, retinal detachment, or retinoschisis are more likely to have alterations in the curvature of the retina and retinal pigment epithelium (RPE) [3]. An alignment method for preprocessing is, therefore, necessary to reduce the curvature variation across scans to improve further clinical measurements. The importance of image alignment as a preprocessing step in OCT image processing has been declared in segmentation [4]–[7], registration algorithms [8] and classification [9]–[12]. Given the importance of this preprocessing step in the accuracy of analysis and measurements, before explanation of proposed method, we review some recent OCT curvature alignment techniques.

Several methods have already been proposed for OCT image alignment. Liu et al. [13] aligned retinal regions by fitting a second-order polynomial to the whole retinal OCT image. They applied this method on the normal, macular edema (ME), macular hole (MH), and age-related macular degeneration (AMD) B-scans. However, the effectiveness of their method was reduced in elevated lesions. Baghaie et al. [14] used sparse and low-rank decomposition of vectored image data for batch alignment and denoising of the OCT images. Sun et al. [11] proposed a method that used linear or second-order polynomial fitting method to flatten the retina region and to reduce the variations between OCT images in morphology. They employed cropped images in training a sparse dictionary and used feature descriptors to classify the AMD, normal, and diabetic macular edema (DME) data. Duan et al. [15] proposed a group-wise curve alignment-based method to segment retinal layers. This method was applied to normal and abnormal B-scans for segmentation of retinal layers. Pan et al. [8] proposed a novel design-detection-deformation mechanism and a feature-based registration method to align OCT B-scans. They evaluated their method in registration of longitudinal OCT images from healthy subjects and subjects diagnosed with serious Choroidal Neovascularization (CNV).

Given that the initial step in our proposed method is to detect

This paragraph of the first footnote will contain the date on which you submitted your paper for review.

Narges Saeedizadeh, Mahnoosh Tajmirriahi, Alireza Haghani, Zahra Amini, Rahele Kafieh, and Hossein Rabbani are with the Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan 81746734641, Iran.

Rahele Kafieh is also with Department of Engineering, Durham University, South Road, Durham, UK.

Elias Khalili Pour, Hamid Riazi-Esfahani, and Kaveh Fadakar are with the Retina Ward, Farabi Eye Hospital, Tehran University of Medical Sciences, Tehran, Iran

Corresponding author: Rahele Kafieh

Narges Saeedizadeh and Mahnoosh Tajmirriahi has the same contribution as first authors.

the retinal pigment epithelium (RPE) in each B-scan, we also reviewed several RPE segmentation techniques that were used for purposes other than alignment. Kafieh et al. [16] presented an OCT curvature correction algorithm using graph-based geometry detection to localize the hyper-reflective complex (HRC). They examined the efficiency of their method in alignment of normal and abnormal HRC layers. Chen et al. [17] used highly reflective and locally connected pixels located below the retinal nerve fiber layer (RNFL) to segment RPE layer. They applied their algorithm on AMD B-scans for drusen segmentation. Srinivasan et al. [9] aligned retinal regions by fitting a second-order polynomial to RPE layer and then flattened the retina. Aligned images were applied in the classification task to discriminate AMD and DME cases from normal ones. However, their approach was confirmed only for moderately deformed RPE layers in order to fix the curvature of the retina's RPE border. Rashno et al. [4] used 1D linear interpolation which led to aligning RPE layer in OCT images. They applied this method for segmentation of three types of fluids including intra-retinal fluid (IRF), sub-retinal fluid (SRF) and pigment epithelial detachment (PED) in OCT B-scans of subjects with AMD and retinal vein occlusion (RVO) or diabetic retinopathy by utilizing a convolutional neural network. Thomas et al. [12] extracted the RPE layer and the baseline from OCT B-scans to find the height of abnormalities in the RPE layer and classify AMD B-scans from normal ones. For this purpose, the randomly selected points of the RPE layer were used to estimate the baseline using iterative polynomial fitting over the chosen points. Then, they obtained the difference between the RPE layer and baseline and used it to estimate the drusen heights as discriminative features.

In this work, we proposed a deep learning-based OCT alignment method consisting of two main steps: At first, we begin by detecting the RPE layer in each B-scan with a deep learning model called TV-Unet, which is made up of Unet with the total variation (TV) added to its loss functions. TV is a measure of spatial continuity in the image and can improve the connectivity requirements to predict segmentation maps [18], [19]. The concept of employing TV-Unet in computerized tomography (CT) image segmentation tasks has also been investigated [20]–[22]. In the second step, we flatten the B-scans along the RPE layer by aligning a set of small windows. In a simple transfer-based alignment approach, each window is translated up or down to force the curve points to form a straight line. One of the challenges in such approach is that regarding differences in various equipment, OCT data from distinct devices exhibit appearance disagreement, resulting in performance fluctuations in the alignment process. Therefore, we propose a device-independent OCT alignment approach to address this issue. This method is successfully trained and tested with combination of different datasets acquired from several OCT devices. The results show acceptable performance of the method in the test phase. Our contribution can be summarized into three folds:

(I) in this work, we present a deep learning-based framework for alignment of the retinal OCT images regarding the benefits of adding TV regularization to the ordinary Unet network.

(II) We develop a robust, shape-preserving and effective alignment method, successfully tested in a large OCT dataset acquired from several OCT platforms.

(III) We will publicly share our dataset as a benchmark for other researchers.

In addition, our work brings additional advances attributable to independency of alignment method from OCT device and demonstrates the performance improvement compared to other approaches. The proposed method can also be utilized to locate drusen using the same mentioned method [12]. Overall, consequent OCT image analysis such as automatic measurements of biomarkers can be effectively performed by using the trained TV-Unet as the preprocessing step.

The rest of this paper is organized as follows: in Section II, we provide detailed description of the proposed method and implementation approach. In section III, we introduce the datasets and report experimental results. Section IV provides a brief conclusion.

II. Method and Materials

A. Dataset

In this study, we used a dataset of 3173 OCT images from healthy and abnormal subjects Images acquired using different OCT devices. We divided these images into six subsets as described in Table I. Subset I consisted of 1568 images from healthy subjects and patients with multiple sclerosis (MS) obtained by the Heidelberg OCT platform (Heidelberg Engineering, Heidelberg, Germany). This subset has been presented in [23]. Subset II is composed of 116 images from abnormal cases (AMD pathology).This dataset has been acquired from the spectral domain (SD-OCT) imaging system from Bioptigen, Inc (Research Triangle Park, NC) and has been utilized in [24]. Subset III is acquired from a Custom-made swept-source OCT (SS-OCT) imaging system designed and built in Department of Biomedical Engineering, University of Basel. This dataset consists of 45 subjects without eye pathologies collected in Didavaran eye clinic, Isfahan, Iran. This subset has been used in [25], [26]. Subset IV has also been used in [27] and contains thirteen 3D macular SD-OCT images obtained from eyes without pathologies using Topcon 3D OCT-1000 imaging system in Ophthalmology Department, Feiz Hospital, Isfahan, Iran. Subset V consists of 63 images with a resolution of 1024x960 pixels obtained at the Farabi Eye Hospital's Retinal Service Center in Tehran, Iran from 30 patients with a variety of eye pathologies. This subset has been used in [28]. Subset VI is studied in [29] and consists of 193 OCT images from 19 DME patients. This dataset has been acquired from the Heidelberg system version 5.1. The dataset will be available at https://github.com/narges-sa/Multi-center-multi-device-Retinal-OCT-dataset-with-RPE-deliniation

B. Unet architecture

The Unet CNN model was first presented by Ronneberger et al. [30] for biomedical image segmentation. It consists of two main parts: an encoder and a decoder. However, one difference between Unet and ordinary auto-encoders is skipped

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3185653

connections between layers of the encoder to the decoder. These connections provide better capturing of the details. In our proposed Unet network, the encoding path consists of 3×3 convolutions which are repeatedly applied. Each convolutional layer is followed by a rectified linear unit (ReLU), a batch normalization, a 2×2 max pooling operation for down sampling, and a drop-out equal to 0.2. At each down-sampling step, the number of feature channels is doubled, and the dimension of the input image is halved. Each step in the decoding path is composed of a) a transposed convolutional layer, b) a concatenation of cropped features from the corresponding contracting path, c) a drop-out equals 0.2, and d) two 3×3 convolutional layers, each followed by a ReLU activation function and batch normalization. The final layer consists of a 1×1 convolution to map feature vector to the desired number of classes with a sigmoid activation function. Figure 1 illustrates the architecture of the proposed TV-Unet.

A. *TV regularization*

The traditional loss function utilized for training U-Net architecture is a weighted cross-entropy loss defined as follows.

$$\mathcal{L}_{Unet} = \sum_{X\epsilon\Omega} w(X) log\left(p_{l(X)}(X)\right) \quad (1)$$

where $w(X)$ is a weight for pixel X. The weights are pre-computed for each ground truth segmentation by assigning higher values to challenging boundary pixels based on a distance map [30]. $l:\Omega \rightarrow \{1,\ldots,K\}$ is true label of each pixel, K denotes the total number of classes and $p_{k(X)}$ is a pixel-wise soft-max that is defined as:

$$p_k(X) = \frac{exp\left(a_k(X)\right)}{\sum_{i=1}^{k} exp\left(a_i(X)\right)} \quad (2)$$

where $a_k(X)$ is the activation in feature channel k at the pixel position $X \in \Omega$ with $\Omega \subset Z^2$. Various regularization terms can be added to the loss function for different purposes such as reduction of overfitting of the network. One of these regularization terms is total variation (TV) which has been widely used in image restoration studies such as [31], and segmentation studies such as [32], [33]. TV is defined as (3) where $\nabla(.)$ denotes the gradient operator.

$$TV\ (u) = \int_{\Omega} |\nabla u(x)| dx \quad (3)$$

In 2D discrete domain ($Y = [Y_{i,j}]$) the TV regularization can be depicted by anisotropic version of TV as (4). In applications such as segmentation where spatial continuity must be preserved, TV regularization can be added to the loss function to incorporate a meaningful relationship between neighboring pixels to the weights of the neural network.

$$TV(Y) = \sum_{i,j} |Y_{i+1,j} - Y_{i,j}| + |Y_{i,j+1} - Y_{i,j}| \quad (4)$$

In other words, using TV as a regularization term forces the neighboring pixels to have similar values except for small number of positions, where it might make sudden jumps. These sudden jumps occur in the edges of the image which are critical in image segmentation. In TV-Unet, to promote connectivity, the total variation is added to the loss function of Unet. Therefore, the loss function of TV-UNet is defined as:

$$\mathcal{L}_{TV-UNet} = \mathcal{L}_{Unet} + \lambda\, TV(M(X))$$
$$\lambda = \frac{1}{255 \times N} \quad (5)$$

where λ is a regularization coefficient, N is the number of pixels in each image, and $\mathcal{L}_{UNet}$ is the binary cross-entropy loss defined in (1), $TV(M(X))$ is the total variation of predicted binary mask for RPE pixels.

A. *Alignment of OCT images*

The OCT image is represented by a 2D matrix $I(x,y)$ where $x \in [1,M], y \in [1,N]$. Having the RPE layer coordinates, whole shape can be aligned to straighten this curve. Suppose that (x_{RPE}, y_{RPE}) indicates a point on the RPE layer and $y_{PRE} = f(x_{RPE})$ where f indicates the RPE curve (Fig. 2 (a)). In order to preserve RPE curvatures created by some pathological effects (such as drusen, fluid-filled regions or sub-retinal mass), which are informative curvatures, we proposed using the estimated coordinates of the RPE layer to form an auxiliary curve so called curvature curve (Fig. 2 (a)). The curvature curve follows the estimated RPE coordinates in the areas without informative curvatures, but, it follows the straight line in informative areas.

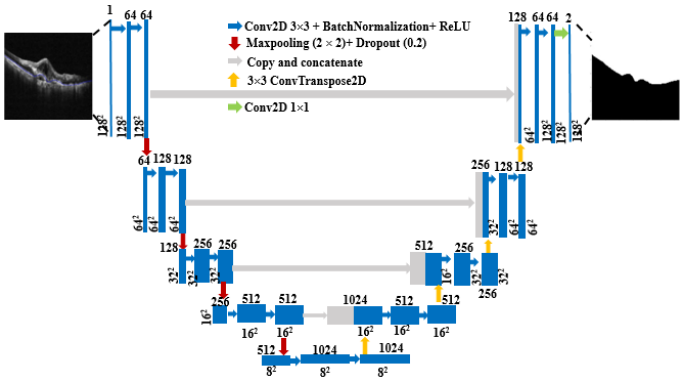

Fig. 1. An overview of the detailed structure of the proposed TV-Unet network.

TABLE I
DETAILS OF THE UTILIZED DATASET

Dataset	Data type	Number of images	Subjects	Imaging system	Image size	Resolution (μm^3)
Subset I	Normal	686	14	Heidelberg (SD-OCT)	1024×496	5.8×3.9×123.6
	Abnormal (MS)	882	18			
Subset II	Abnormal (AMD)	116	50	Bioptigen (SD-OCT)	1024×496	3.87×6×124
Subset III	Normal	45	45	Custom-made (SS-OCT)	1024×300	10×20×12
Subset IV	Normal	1188	11	Topcon (SD-OCT)	650×512	13.67×4.81×24.41
Subset V	Abnormal	63	30	Optovue (EDI SD-OCT)	1024×960	8 × 12 × 3.8
Subset VI	Abnormal (DME)	193	19	Heidelberg (SD-OCT)	875×656	6×6×3.8

This goal was attained by linear interpolating the RPE points through the running minimum filter. Given an input sequence

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3185653

of size W, $(x_i, f(x_i)), (x_{i+1}, f(x_{i+1})), \dots, (x_M, f(x_W)), \quad 1 \leq i \leq W$ we define the response of the minimum filter as follows.

$$\hat{x}_W = arglocalmin_x(f(x)) \;\; x_i \leq x \leq x_W \tag{6}$$

In fact, a running minimum filter is used to find the place, $\hat{x}_W$ (local solutions), and minimum value, $\hat{y}_W = f(\hat{x}_W)$, of the RPE curve in neighboring window of size W from each pixel. Hence, for each point (x_{RPE}, y_{RPE}) of the RPE layer curve, the location $(x, C(x))$ of a curvature curve is calculated by a linear interpolation as follows, where $x_{RPE} \leq x \leq x_W$.

$$C(x) = \hat{y}_W + \left(\frac{x - x_{RPE}}{\hat{x}_W + W - x_{RPE}}\right)(\hat{y}_W - y_{RPE}) \tag{7}$$

The width of running filter can be adjusted to cover all informative curvatures of the RPE layer. Here, we set $W = 15$ for our application. The curvature curve has crucial role in the correct estimation of the required transferring (Figure 2 (b)). Without curvature curve, the amount of required transfer is overestimated in informative area and the lesions are dislocated in the new aligned image as shown in Fig. 2 (c). After creating curvature curve, a straight line passing through its lowest height is considered as the reference alignment line (Fig. 2 (a)).

For simple transfer-based alignment, the image is divided into windows of width equal to one pixel and the height from top to the bottom of the image (Fig. 3 (b)). The distance between curvature curve and the reference line is considered as amount of required transfer for each pixel inside the corresponding window. Therefore, new image of the transferred pixels, (x, y_{new}) would be determined as follows (Fig. 3 (e)):

$$I(x, y_{new}) = I\left(x, y + C(x) - min\big(C(x)\big)\right) \tag{8}$$

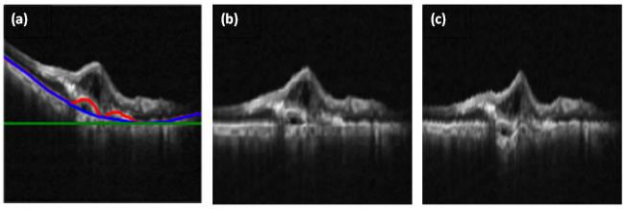

Fig. 2. (a) The estimated RPE layer (red), curvature curve (blue), and reference alignment line (green) for a sample B-scan of dataset V. Correction is based on the distance between curvature curve (blue) and the reference line (green), (b) image aligned based on the proposed curvature curve, (c) image incorrectly aligned based on tracking the estimated RPE layer (red).

B. Shape- preserving corrections

The main drawback of simple transfer-based alignment method is that it cannot preserve the shape and size of the pathological lesions as well as the curvatures of them. Distortions and flattening in morphology of drusen can be widely seen in these methods. This point is more highlighted when the lesion area is asymmetric and has stretched to the left or right. To address this issue, we updated the algorithm in such a way that it can resolve these deformations in the alignment process. To accomplish this, in the curvature curve we determined the points that are inclined relative to the vertical axis and applied the geometric transform to their coordinates before final alignment in (8). Figure 3 (a) illustrates the vertical axis (yellow) and the calculated perpendicular line (red) to the pixel of the curvature curve. Suppose a point of the curvature curve with coordinates $(x_1, C(x_1))$. First, we calculated the line perpendicular to the curve at this point and extended it throughout the image as a line with following equation.

$$y_p = \frac{-1}{C'(x_1)}x + b \tag{9}$$

where $C'(x_1)$ is the derivative of the curvature curve at point x_1 and b denotes the value of line at $x = 0$. Values of x and y_p change in range (0, 128). Then an intermediate window I_{int} (Fig. 3 (c)) is created as follows:

$$I_{int}(x, y) = \begin{cases} I(x, y) & \theta = 90 \\ I\left(x, \frac{y_p}{cos\theta}\right) & \theta \neq 90 \end{cases} \tag{10}$$

where θ is the angle between perpendicular line, depicted in (9), and the vertical line $x = x_1$ and is calculated as follows:

$$\theta = \arctan\left(\frac{-1}{C'(x_1)}\right) \tag{11}$$

The aligned image is then created using the following equation (Fig. 3 (f)).

$$I(x, y_{new}) = I_{int}\left(x, y + C(x) - min\big(C(x)\big)\right) \tag{12}$$

In this way, the real distance between oblique points and curvature curve are preserved during transferring the pixel to the objective image. Therefore, deformation of lesions is prevented in the proposed alignment method.

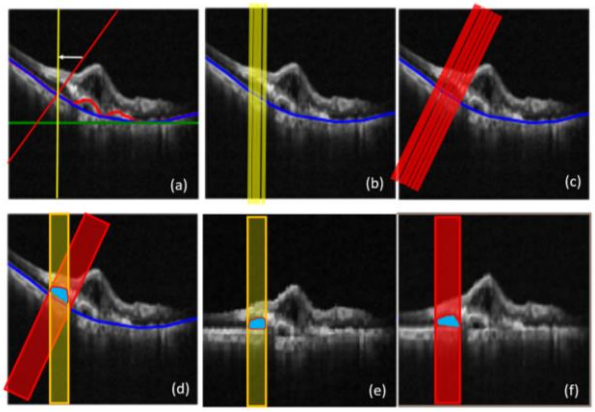

Fig. 3. (a) The estimated RPE layer (red), curvature curve (blue), and reference alignment line (green) for a sample B-scan of dataset V. yellow line and red line represent vertical axis, and prependicular line of a pixle respectively. (b) windows of width equal to one pixel and the height from top to the bottom of the image, designed for simple transfer-based alignment, (c) intermediate windows designed with the proposed alignement method, (d) blue-colored ground truth segmented fluid region and corresponding areas covered with windows in part (b) and part (c), (e) aligned b-scan with simple transfer-based alignment, (f) aligned b-scan with proposed shape preserving alignement method.

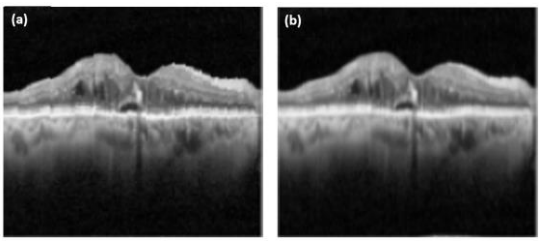

Fig. 4. Results of alignments of sample B-scan of dataseta VI. (a) before supperesion of saw tooth artifacts. (b) after supperesion of saw tooth artifacts.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3185653

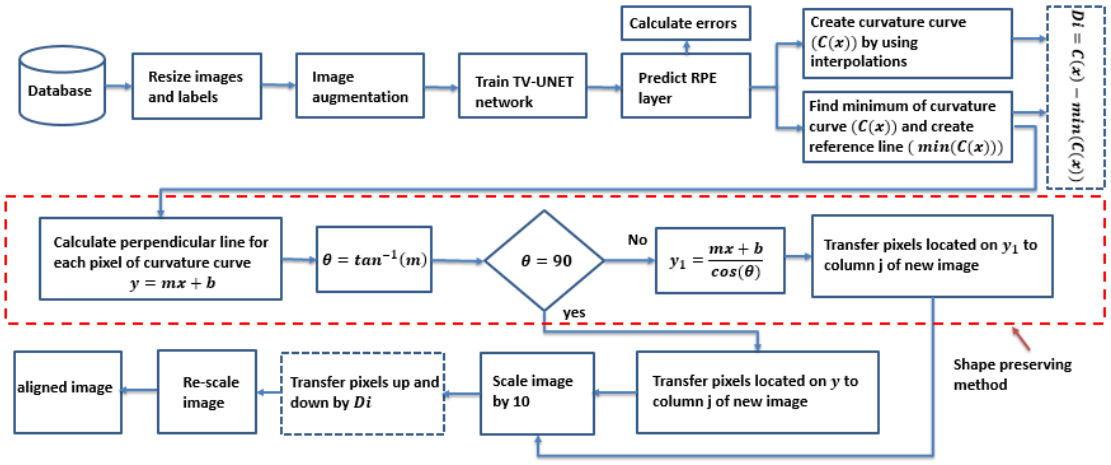

Fig. 5. Block diagram of proposed alignment method.

Figure 3 (d) illustrates the sample B-scan of dataset V and the ground truth segmented fluid region. This image is aligned without the shape-preserving correction (Fig .3 (e)) and with the proposed correction method (Fig .3 (f)). The shape, size and area of the fluid area is clearly preserved in proposed shape preserving method (Fig .3 (f)) in contrast to ordinary alignment where the overlaid area is deformed and smaller than original area (Fig .3 (e)).

C. *Suppression of saw-tooth artifacts*

The interpolations applied in creating the curvature curve can cause saw-tooth artifacts in the intra-layer edges of the aligned image. This issue arises as a result of rounding decimal numbers in order to get integer pixel coordinates during the interpolation. A suitable option is to scale the height of the test image and any relevant values to a larger size before alignment, apply the alignment procedure, and then rescale the image height to its original size after alignment. For instance, suppose the rescaling is done with the factor of 10. In this way, the decimal point will move one place to the right and rounding will be done more accurately. Therefore, the saw tooth steps, created by the decimal numbers during the rounding process, will be eliminated. Figure 4 illustrates the aligned sample B-scan before and after the suppression of artifacts.

III. Results and Discussion

The block diagram of the proposed alignment method is depicted in Fig .5 which will be described in this section. All the executable codes were implemented in Python, TensorFlow backend utilizing 12 GB NVIDIA K80 GPU, 324 MHz memory of the Google Colaboratory (Colab). The scratch TV-Unet network was trained here, and no pre-trained parts or transfer learning methods were used.

A. *Training details*

Whole dataset, in a subject-wise manner, with an 80/20 ratio was split for training and testing. In this way, 754 B-scans were parted to be used as test dataset. Since, in the 2419 remnant B-scans there were more normal images than abnormal ones, in order to have a balanced dataset, abnormal data was augmented and 4089 B-scans were obtained to be applied in training and validation process in a balanced way. Given the variety of datasets used in this study, it was important to identify all B-scans and their related masks (labels) prior to applying them to the network. To do this, we retrieved the RPE layer from each ground truth B- scan's layers and stored it as the appropriate mask. Eventually, B-scans and their matched labels were resized to 128 ×128 and applied in training and validation process. Since, images of various vendors had different range of intensity, we normalized the intensity range of training and test datasets to the [0, 1] values as well. The sample images of the dataset and their masks are illustrated in Fig. 6.

B. *Results of RPE detection*

We used the trained network to predict RPE layers in test B-scans. Results of RPE detection for sample B-scans from different datasets are depicted in Fig. 7 for visual inspection. In addition, we calculated signed and unsigned (absolute) errors between the target RPE and predicted layer of test B-scans using a variety of datasets.

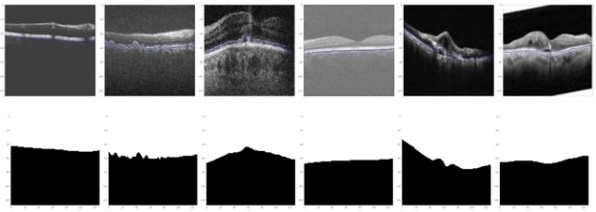

Fig. 6. The sample images (top panel) and corresponding ground-truth (down panel) from different Datasets.

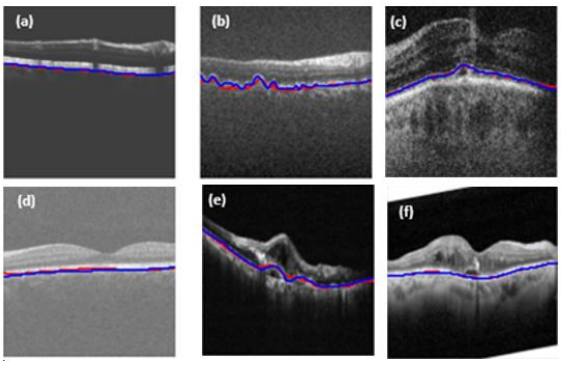

Fig. 7. Results of RPE detection on sample B-scans from datasets: (a) I, (b) II, (c) III, (d) IV, (e) V, (f) VI. The blue and red lines represent the ground-truth and predicted RPE layer respectively.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3185653

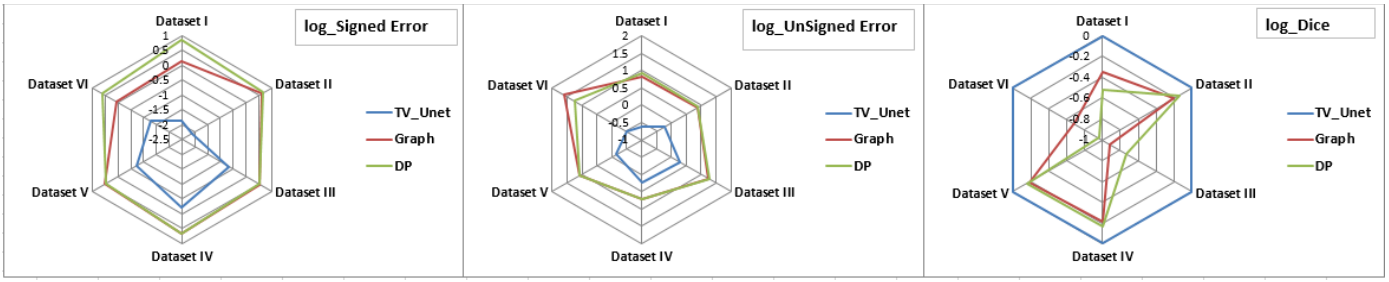

Fig. 8. The radar chart representing the logarithm of mean values of signed positioning error, unsigned positioning error, and Dice coefficients calculated for RPE layer versus different datasets.Results are acquired applying graph-based [27] , DP [34], and proposed TV_Unet method.

TABLE II
MEAN VALUES OF SIGNED AND UNSIGNED RPE DETECTION

	Signed error			Unsigned error		
	Graph [27]	DP [34]	Proposed TV_Unet	Graph [27]	DP [34]	Proposed TV_Unet
I	1.328	6.989	**0.014**	6.726	8.508	**0.243**
II	4.221	4.608	**-0.008**	7.3270	7.785	**0.59**
III	-3.441	-3.212	**-0.214**	18.191	18.470	**1.943**
IV	4.540	4.498	**0.613**	5.348	5.333	**1.754**
V	3.240	3.029	**-0.179**	12.158	12.049	**0.692**
VI	1.089	4.213	**0.054**	38.188	18.149	**0.306**

TABLE III
MEAN DICE COEFFICIENTS FOR RPE DETECTION

	I	II	III	IV	V	VI
Graph [27]	0.445	0.620	0.121	0.606	0.656	0.205
DP [34]	0.304	0.716	0.184	0.675	0.691	0.109
Proposed TV_Unet	**0.997**	**0.992**	**0.987**	**0.988**	**0.985**	**0.997**

Unsigned errors are computed to avoid underestimating the total number of mistakes produced by the cancellation of positive and negative errors. Mean values of these errors (in pixel) for six datasets are reported in Table II and compared with a graph-based method [27], and a dynamic programming (DP) [34]. According to Table II, the proposed TV-Unet is capable of accurately detecting the RPE layer in images. For more analysis, the RPE detection of each method was analyzed, by calculating the Dice coefficient corresponding to the overlap of each segmented RPE compared with its annotations (Table III). High Dice coefficients indicate the outstanding performance of the proposed method in RPE detection.

This satisfying performance has been provided by adding TV term to the loss function of the Unet network which helps accurate detection of the RPE layer and facilitate the following alignment step. Furthermore, we provided various training datasets from different capturing devices. This assisted the network in learning the requisite structural elements for detecting the RPE layer in a wide range of OCT images. Radar chart of the logarithmic values of boundary detection errors and Dice coefficients are represented in Fig. 8 for more comparison. According to Fig. 8, lower boundary positioning errors and higher Dice coefficients than comparing methods indicates the ability of the proposed TV-Unet to operate as a device-independent RPE-detection technique, and to be applicable for various datasets acquired from various vendors.

C. Results of alignment and its application

After indicating the coordinates of RPE, we aligned whole shape of each OCT image of the applied datasets to straighten this curve based on the method described in section II. Here, we devised and implemented a new quantitative evaluation approach for determining the efficiency of the alignment. To this purpose, we hypothesized that if the RPE layer's intensities are aligned horizontally, their first derivative must approach toward zero in locations devoid of pathologic curvature. In contrast, in the pathologic regions, we expect large variations in the values of derivatives which depend on the size and the shape of the curvature of lesions. Result of the derivative plots of a sample image before and after alignment is illustrated in Fig. 9. It can be seen that in the aligned image the derivatives have very small values (near zero) at non-pathologic areas and have large fluctuations in lesion areas. In addition, it can be seen from Fig. 9 (d) that the proposed evaluation method can be used in exact localization of the lesion area as well (red lines). This hopefully indicates the application of the proposed alignment method in the future macular pathology detection. Numerous diagnostic biomarkers, including the size and shape of lesions, are quantified using OCT images and are commonly used in ophthalmology to detect retinal abnormalities. As a result, these biomarkers must remain stable throughout alignment. As described in section II. F, by adding the geometric modification to the sloped zones of the RPE layer, we adjusted the alignment algorithm to preserve the size and shape of the lesions during the alignment.

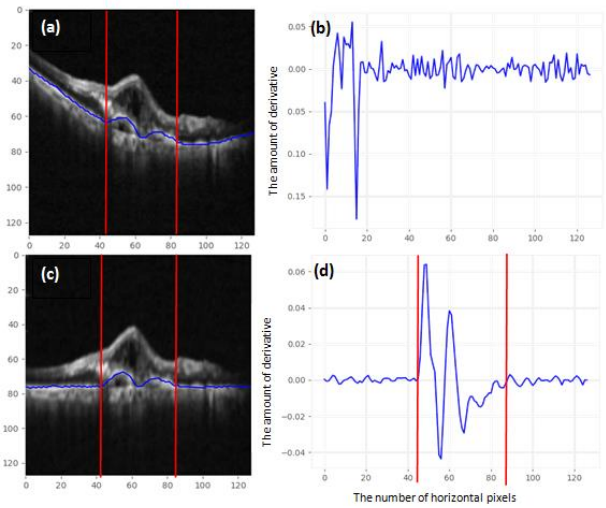

Fig. 9. (a,c) The sample image before and after alignment, (b,d) plot of derivatives of the intensities of the pixels located in the RPE layer. (red line indicates the fluid region which is also detrminable in the derivative plot after alignment.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3185653

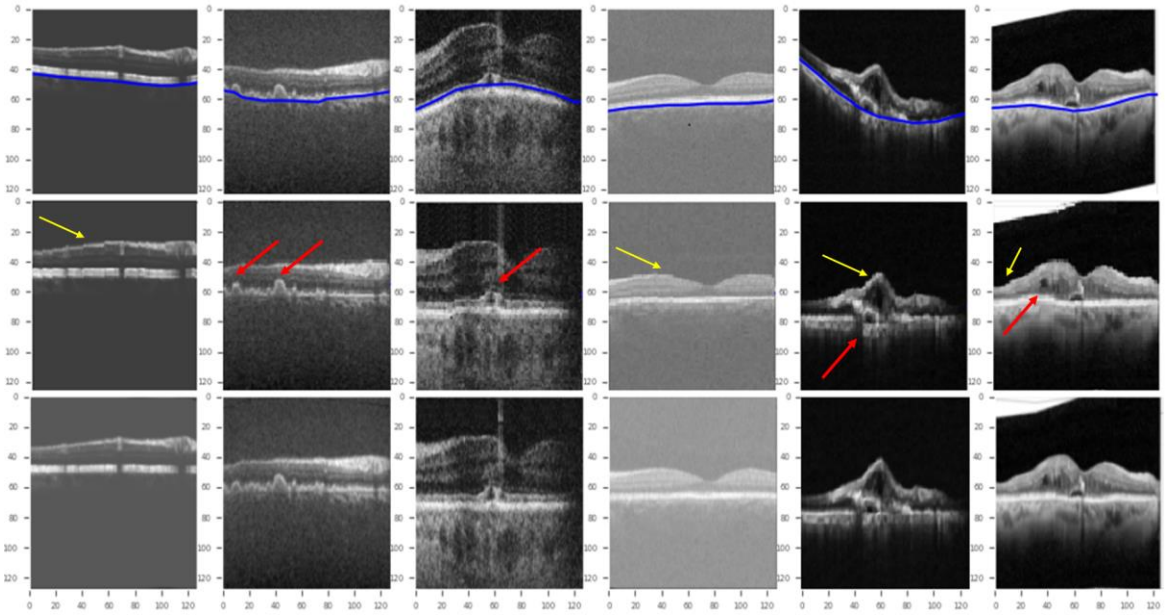

Fig. 10. Comparison of the proposed shape preserving alignments and simple transfer-based alignment on sample B-scan from datasets I, II, III, IV, V, VI from left to right, respectively. First row: Original image, Second row: corresponding aligned images with simple transfer-based alignment. Third raw: corresponding aligned version of image with proposed shape preserving and artifact supperession method. Yellow arrows indicate saw tooth artifacts and the red arrows show the change in lesions' shape using simple transfer-based alignment. Both mntioned problems are shown to be solved in proposed shape preserving method. In the first row, blue curves indicate the calculated curvature curve. $0 \leq X \leq 128$ and $0 \leq Y \leq 128$ axis represent the location (X, Y) of each correspondent pixel in the X-Y plane..

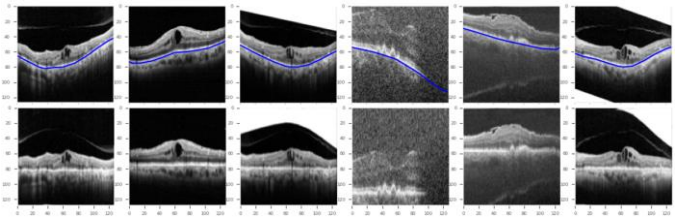

Fig. 11. Results of alignments on challenging B-scan from different datasets. Top panel: Original image, down panel: aligned version of image. In the first row, blue curves indicate the calculated curvature curve. $0 \leq X \leq 128$ and $0 \leq Y \leq 128$ axis represent the location (X, Y) of each correspondent pixel in the X-Y plane.

Furthermore, we suppressed the saw tooth artifacts occurred during interpolation-steps of the alignment procedure by adding a fake super-resolution step before alignment of the images as described in section II F. In Fig. 10, the improvement in alignment of images after proposed shape preserving modifications are depicted for visual inspection (2^{nd} and 3^{rd} raw show the aligned images without and with shape preserving method, respectively). According to this figure, the proposed modifications, particularly in presence of big lesions, helped preserving the lesions' shape and removing saw tooth artifacts, indicating the good performance of the proposed method in alignment of the OCT images. More results of alignment on sample challenging B-scans from different datasets are also provided in Fig. 11 to verify this feat.

IV. CONCLUSION

In this paper, a novel deep learning-based method is proposed to align retinal OCT B-scans by incorporating TV regularization in the Unet network. We reported visual and quantitative assessment of the TV-Unet network in detection of RPE layer to demonstrate the outstanding performance of the proposed method. When comparing with other methods, the RPE detection results showed noticeable performance for the TV-Unet method with Dice coefficients of 99.7%, and 99.2% in normal and abnormal datasets, respectively. The boundary errors of the proposed method has the lowest values among methods, with a difference of mean absolute error of 4 and 38 pixels for the RPE layer in various datasets. It is worth noting the retinal layers are small in area, so changes of a few pixels may have a large impact on clinical measurements. Afterward, an interpolation-based method was utilized to remove unwanted curvatures caused by distortions occurred during image acquisition. To preserve the shape and size of the lesions in the pathologic B-scans, the geometric modification was proposed and incorporated in the alignment method. Extra modification was proposed to suppress the saw tooth artifacts which is a common issue in the interpolation-based methods by adding a fake super-resolution step to the alignment method. Visual assessments of the results indicated the outperformance of the proposed alignment method. In addition, we proposed a new derivative-based approach for quantitative assessment of alignment method which can be used as future approach and be applied in localization of the lesions in the pathologic scans. The proposed method have been evaluated for various datasets and different vendors and the results indicated that the proposed method can be used as a device-independent, shape preserving alignment method in the preprocessing step for OCT image analysis.

However, the proposed alignment process may lead to inaccurate measurement of the retinal thickness due to creating

distortions in the actual morphology of retinal layers. Therefore, it is imperative to determine where the proposed alignment process could affect the measurements. It is expected that the thickness map is not affected for normal cases, however, the comparison of retinal thickness measurements of analyzed and original images, especially in images with macular pathologies, can be further studied in future works.

ACKNOWLEDGEMENT

This work was supported in part by the Leading House South Asia and Iran, Zurich University of Applied Sciences (Switzerland), under the Research Seed Money Grant; in part by the National Institute for Medical Research Development (NIMAD) under Grant 976795; and in part by the Vice-Chancellery for Research and Technology, Isfahan University of Medical Sciences, under Grant 298236.

REFERENCES

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

[2] J. Chhablani, T. Krishnan, V. Sethi, and I. Kozak, "Artifacts in optical coherence tomography," *Saudi J. Ophthalmol.*, vol. 28, no. 2, pp. 81–87, 2014.

[3] S. Lake, M. Bottema, K. Williams, and K. Reynolds, "The correlation between optical coherence tomography retinal shape irregularity and axial length," *PLoS One*, vol. 14, no. 12, p. e0227207, 2019.

[4] A. Rashno, D. D. Koozekanani, and K. K. Parhi, "Oct fluid segmentation using graph shortest path and convolutional neural network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 3426–3429.

[5] P. Seeböck *et al.*, "Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT," *IEEE Trans. Med. Imaging*, vol. 39, no. 1, pp. 87–98, 2019.

[6] B. Hassan, S. Qin, T. Hassan, R. Ahmed, and N. Werghi, "Joint Segmentation and Quantification of Chorioretinal Biomarkers in Optical Coherence Tomography Scans: A Deep Learning Approach," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–17, 2021.

[7] M. Monemian and H. Rabbani, "Analysis of A Novel Segmentation Algorithm for Optical Coherence Tomography Images Based on Pixels Intensity Correlations," *IEEE Trans. Instrum. Meas.*, 2020.

[8] L. Pan *et al.*, "OCTRexpert: a feature-based 3D registration method for retinal OCT images," *IEEE Trans. Image Process.*, vol. 29, pp. 3885–3897, 2020.

[9] P. P. Srinivasan *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express*, vol. 5, no. 10, pp. 3568–3577, 2014.

[10] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT classification using a multi-scale convolutional neural network ensemble," *IEEE Trans. Med. Imaging*, vol. 37, no. 4, pp. 1024–1034, 2017.

[11] Y. Sun, S. Li, and Z. Sun, "Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning," *J. Biomed. Opt.*, vol. 22, no. 1, p. 16012, 2017.

[12] A. Thomas *et al.*, "RPE layer detection and baseline estimation using statistical methods and randomization for classification of AMD from retinal OCT," *Comput. Methods Programs Biomed.*, p. 105822, 2020.

[13] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and J. M. Rehg, "Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding," *Med. Image Anal.*, vol. 15, no. 5, pp. 748–759, 2011.

[14] A. Baghaie, R. M. D'souza, and Z. Yu, "Sparse and low rank decomposition based batch image alignment for speckle reduction of retinal OCT images," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 226–230.

[15] W. Duan *et al.*, "A generative model for OCT retinal layer segmentation by groupwise curve alignment," *IEEE Access*, vol. 6, pp. 25130–25141, 2018.

[16] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Curvature correction of retinal OCTs using graph-based geometry detection," *Phys. Med. Biol.*, vol. 58, no. 9, p. 2925, 2013.

[17] Q. Chen *et al.*, "Automated drusen segmentation and quantification in SD-OCT images," *Med. Image Anal.*, vol. 17, no. 8, pp. 1058–1072, 2013.

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

[19] S. Minaee and Y. Wang, "An ADMM approach to masked signal decomposition using subspace representation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3192–3204, 2019.

[20] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[21] F. Jia, J. Liu, and X.-C. Tai, "A regularized convolutional neural network for semantic image segmentation," *Anal. Appl.*, vol. 19, no. 01, pp. 147–165, 2021.

[22] N. Saeedizadeh, S. Minaee, R. Kafieh, S. Yazdani, and M. Sonka, "COVID TV-Unet: Segmenting COVID-19 chest CT images using connectivity imposed Unet," *Comput. Methods Programs Biomed. Updat.*, vol. 1, p. 100007, 2021.

[23] Y. He, A. Carass, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Retinal layer parcellation of optical coherence tomography images: Data resource for multiple sclerosis and healthy controls," *Data Br.*, vol. 22, pp. 601–604, 2019.

[24] S. Farsiu *et al.*, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162–172, 2014.

[25] M. Tajmirriahi, R. Kafieh, Z. Amini, and H. Rabbani, "A Lightweight Mimic Convolutional Auto-encoder for Denoising Retinal Optical Coherence Tomography Images," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[26] M. Tajmirriahi, Z. Amini, A. Hamidi, A. Zam, and H. Rabbani, "Modeling of Retinal Optical Coherence Tomography Based on Stochastic Differential Equations: Application to Denoising," *IEEE Trans. Med. Imaging*, vol. 40, no. 8, pp. 2129–2141, 2021, doi: 10.1109/TMI.2021.3073174.

[27] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map," *Med. Image Anal.*, vol. 17, no. 8, pp. 907–928, 2013.

[28] Roya Arian, Tahereh Mahmoudi, Hamid Riazi-Esfahani, Raheleh Kafieh,Hooshang Faghihi,Ahmad Mirshahi, "Automatic Choroid Vascularity Index Calculation in Optical Coherence Tomography Images Low Contrast Sclerochoroidal Junction Using Deep Learning," 2021.

[29] M. Montazerin *et al.*, "Livelayer: A Semi-Automatic Software Program for Segmentation of Layers and Diabetic Macular Edema in Optical Coherence Tomography Images," *arXiv Prepr. arXiv2003.05916*, 2020.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[31] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, 1992.

[32] X. Tu *et al.*, "MR image segmentation and bias field estimation based on coherent local intensity clustering with total variation regularization," *Med. Biol. Eng. Comput.*, vol. 54, no. 12, pp. 1807–1818, 2016.

[33] C. Wang, B. Yang, and Y. Liao, "Unsupervised image segmentation using convolutional autoencoder with total variation regularization as preprocessing," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 1877–1881.

[34] A. Haghani, R. Kafieh, and M. Akhlaghi, "The Improvement of Extraction Retinal Pigment Epithelium (RPE) Layer in Optical Coherence Tomography (OCT) Images by Using Piecewise Dynamic Programming Method in Patients with Persistent Epithelial Defect (PED)," *J. Isfahan Med. Sch.*, vol. 39, no. 637, pp. 594–603, 2021.