Segmentation of Intra-Retinal Cysts From Optical Coherence Tomography Images Using a Fully Convolutional Neural Network Model

G. N. Girish, Bibhash Thakur, Sohini Roy Chowdhury, Abhishek R. Kothari, and Jeny Rajan

***Abstract*—Optical coherence tomography (OCT) is an imaging modality that is used extensively for ophthalmic diagnosis, near-histological visualization, and quantification of retinal abnormalities such as cysts, exudates, retinal layer disorganization, etc. Intra-retinal cysts (IRCs) occur in several macular disorders such as, diabetic macular edema, retinal vascular disorders, age-related macular degeneration, and inflammatory disorders. Automated segmentation of IRCs poses challenges owing to variations in the acquisition system scan intensities, speckle noise, and imaging artifacts. Several segmentation methods have been proposed in the literature for IRC segmentation on vendor-specific OCT images that lack generalizability across imaging systems. In this paper, we propose a fully convolutional network (FCN) model for vendor-independent IRC segmentation. The proposed method counteracts image noise variabilities and trains FCN models on OCT sub-images from the OPTIMA cyst segmentation challenge dataset (with four different vendor-specific images, namely, Cirrus, Nidek, Spectralis, and Topcon). Further, optimal data augmentation and model hyperparametrization are shown to prevent over-fitting for IRC area segmentation. The proposed method is evaluated on the test dataset with a recall/precision rate of 0.66/0.79 across imaging vendors. The Dice correlation coefficient of the proposed method outperforms that of the published algorithms in the OPTIMA cyst segmentation challenge with a Dice rate of 0.71 across the vendors.**

***Index Terms*—Optical coherence tomography, segmentation, retinal cyst, cystoid macular edema, convolutional neural networks, deep learning.**

Manuscript received August 31, 2017; revised January 22, 2018; accepted February 21, 2018. Date of publication February 27, 2018; date of current version January 2, 2019. This work was supported by the Science and Engineering Research Board (Department of Science and Technology, India) under Project EMR/2016/002677. *(G. N. Girish and Bibhash Thakur contributed equally to this work.) (Corresponding author: G. N. Girish.)*

G. N. Girish, B. Thakur, and J. Rajan are with the Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal 575025, India (e-mail: girishanit@gmail.com; bibhashthakur92@gmail.com; jenyrajan@gmail.com).

S. R. Chowdhury is with the Department of Electrical and Computer Engineering, University of Washington, Bothell, WA 98011 USA (e-mail: roych@uw.edu).

A. R. Kothari is with the Pink City Eye and Retina Center, Jaipur 302015, India (e-mail: dr.a.kothari@gmail.com).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2018.2810379

I. Introduction

RETINAL disorders pose a significant health challenge in the current era. Macular edema is a leading cause of visual impairment in the working population. It affects central vision, thereby impacting reading, writing, fine motor activities and driving. Macular edema results from vascular incompetence, leading to fluid leakage into the retina [1]. This extravasated fluid accumulates in retina tissue into multiple interconnected labyrinthine cavities. In cross-sectional views, these fluid spaces can be observed as round or oval empty spaces and are referred to as cysts, even though this is a misnomer. These '*intra-retinal cysts*' (IRC) are seen in macular edema associated with several eye disorders such as age-related macular degeneration, diabetic retinopathy, retinal vein occlusion, etc. Also, studies have demonstrated that cystoid macular edema is strongly associated with visual loss, and the volume of cystic fluid space, and their relative location in retinal micro-structure may correlate with the degree of visual impairment [2].

Optical Coherence Tomography (OCT) is a non-invasive technique that uses low coherence light for resolving internal structures of biological tissues [3]. OCT is widely used for the diagnosis of several ocular diseases and visualization of retinal structures such as retinal layers, cysts, exudates, and sub-retinal fluid. OCT produces a large amount of volumetric data; analyzing these images completely for locating cysts and measuring specific volumes is time-consuming as well as it requires expert annotations. Automated segmentation methods can assist the ophthalmologists for better interpretation and quantification of retinal features to enhance diagnosis and decision making in the treatment of retinal pathologies.

Over the past few years, several automated IRC segmentation methods have been proposed in the literature. Retinal cyst segmentation has been performed using 3D volumetric approaches [4]–[8] as well as 2D B-scan approaches [2], [9]–[19]. Recently, a benchmark study [20] was conducted to analyze the performance of different automated IRC segmentation methods. To the best of our knowledge, the existing methods are dependent on vendor-specific acquisition systems and require human intervention to determine acquisition and visualization-specific parameters. Additionally, segmentation of fluid-filled spaces in retinal volumes across different OCT vendors is a challenging task due to the variations in the noise levels

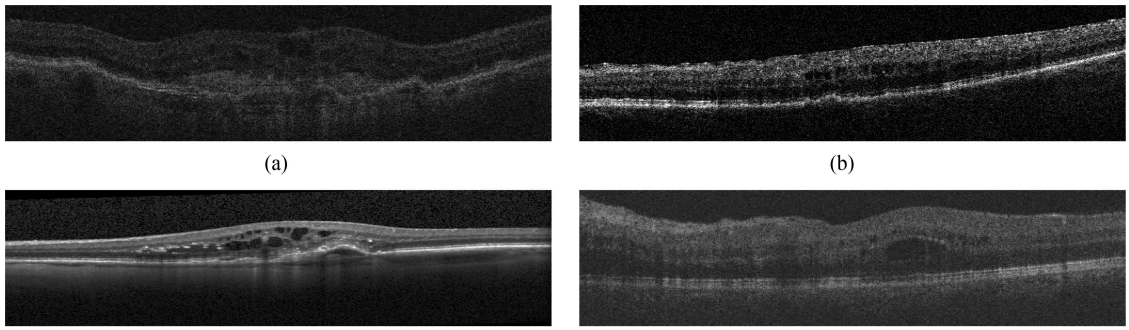

Fig. 1. Sample OCT B-scans of retina from four different vendors (Image courtesy: OPTIMA cyst challenge dataset [21]). (a) Cirrus (b) Nidek (c) Spectralis (d) Topcon.

and pixel intensities across scans [see Fig. 1]. The proposed method addresses these challenges posed to the development of automated vendor independent cyst segmentation methods.

Recent forays into the usage of Convolutional Neural Networks (CNNs) for biomedical image segmentation have shown significant improvements over other approaches [22], [23]. CNNs are a specific variant of standard neural network models, which take an image as an input and perform repetitive convolution operations to abstract features specific to the segmentation task. The learnable weights and biases of the neurons serve as convolution filters. These filters may be arranged in different ways to design architectural variants for various segmentation problems. One such significant CNN model used in a variety of different medical image segmentation tasks, is the U-Net architecture proposed by Ronneberger *et al.* [24]. U-Net is a fully convolutional network (FCN), such as the one popularized by Long and Shelhammer [25], where, if an image is provided as the input, the output obtained is its corresponding segmentation mask. Milletari *et al.* [26] further expanded this concept and converted their model to accept 3D image stacks as input. They also introduced the concept of residual connections described in the widely cited ResNet architecture [27]. The self-learning and abstract feature learning capabilities of CNNs for distinguishing subtle spatial variations motivate our choice of FCN in vendor-independent cyst segmentation tasks. So far, FCNs have not been utilized for IRC segmentation from vendor-independent OCT scans.

This paper makes two key contributions. First, a customizable state-of-the-art FCN model is presented that is capable of automated IRC area segmentation from OCT images across vendor-specific imaging systems. We analyze the sensitivity of model parameters, such as number of layers and kernel dimensions, towards the IRC area segmentation goals. We observe that an optimally parametrized model can achieve higher recall rate of 0.66 while preserving the precision rate of 0.79 across multiple vendor data, when compared to state-of-the-art methods. Second, the importance of OCT image pre-processing by image noise suppression using Gamma noise models, sub-retinal region of interest (ROI) segmentation and optimal data augmentation methods are presented. We observe that image pre-processing and domain-specific data augmentation methods significantly prevent model over-fitting while ensuring generalizability across vendor-specific imaging systems.

The rest of the paper is organized as follows. In Section II the methods and data are discussed. In Section III the experimental setups are explained. Experimental results are presented and discussed in Section IV. Finally, conclusions and discussions are drawn in Section V.

II. Methods and Data

The proposed method comprises of two stages, namely: pre-processing stage that includes image noise removal and retinal layer segmentation, followed by IRC segmentation stage. OCT scans are inherently affected by speckle noise, that influences the performance of automated IRC segmentation methods. The initial pre-processing module suppresses speckle noise and improves the IRC segmentation performance. The two methodological stages are described below.

A. Pre-processing: Denoising and Sub-retinal Layer Segmentation

OCT scans contain varying degrees of speckle noise as in other medical imaging modalities like ultrasound. The work in [28] experimentally showed that OCT images with presence of noise can be approximated with a Gamma distribution. Motivated by this work, we implement the approach given in [29] as Unbiased Fast Non Local Means (UFNLM) to denoise the OCT image data set. First, each input image is denoised with a Fast NLM (FNLM) method proposed in [30] followed by the subtraction of bias due to speckle. Here, the bias term is estimated from the noisy image using the maximum likelihood method described in [28], [29],

$$\hat{B}_e = \hat{\rho}_{\mathrm{ML}} \cdot \hat{\beta}_{\mathrm{ML}} \tag{1}$$

where $\hat{B}_e$ is the estimated bias and $\hat{\rho}_{\mathrm{ML}}$, $\hat{\beta}_{\mathrm{ML}}$ are the estimated shape and scale parameters (of the Gamma distribution) from the noisy image. Since the shape and scale parameters can be assumed to be consistent throughout all regions in an input OCT image, they can be estimated from any homogeneous image region (where the underlying intensity is constant). If we assume that many such piece-wise homogeneous regions may exist, then

TABLE I
AUTOMATICALLY ESTIMATED BIAS VALUE '$\hat{B}_e$' AND SMOOTHING PARAMETER 'h' ON DIFFERENT VENDOR OCT SCANS CONSIDERED IN THIS STUDY FOR SPECKLE NOISE REDUCTION

Vendor	Smoothing Parameter (h)	Bias Value ($\hat{B}_e$)
Cirrus	10	15
Nidek	15	26
Spectralis	10	32
Topcon	10	9

ρ and β can also be estimated as [28], [29],

$$\hat{\rho}_{\mathrm{ML}} = \mathrm{mode}\left\{\rho_{\mathrm{ML}(i,j)}\right\}, \qquad \hat{\beta}_{\mathrm{ML}} = \mathrm{mode}\left\{\beta_{\mathrm{ML}(i,j)}\right\} \quad (2)$$

where, $\hat{\rho}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{ML}}$ represent the estimated values of ρ and β, respectively, for each pixel at location (i, j) using a neighborhood of size $[m \times n]$. In this work, the neighborhood window size is empirically determined as $[5 \times 5]$. Finally, the FNLM algorithm is executed with the following parameters: search window and similarity window sizes of the filter are $[11 \times 11]$ and $[7 \times 7]$, respectively. The smoothing parameter h is empirically estimated across vendors. The estimated bias parameter $\hat{B}_e$ and smoothing parameter h for images acquired with different vendors are reported in Table I.

Following the image denoising process, sub-retinal layer segmentation provides information regarding morphology variations and thickness of the layers. This information is utilized to refine the IRC segmentation results. In this work, the Iowa Reference Algorithm [31]–[33] is used to segment 11 different retinal layers. As intra-retinal cysts are prone to occur in the retinal area enclosed by Internal Limiting Membrane (ILM) to Boundary of Myoid and Ellipsoid of Inner Segments (BMEIS) layers, the OCT B-scans are cropped by delimiting this region as the ROI.

After delimiting the ROI of the OCT scans, contrast enhancement is performed using CLAHE [34] to increase the intensity difference between the cystic and non-cystic regions. Kernel size of $(\frac{1}{8})$th of the image width and height and clip limit of 0.01 is used for CLAHE.

B. Cyst Segmentation: FCN Model

The FCN model [25] accepts denoised sub-retinal layer regions of OCT images as input and produces a prediction score matrix that is then used to construct a binary output mask corresponding to segmented cyst regions. Our FCN model is inspired by Ronneberger's U-Net architecture [24] that captures both local and global features from an input image to construct an accurate segmentation map. While global features indicate the exact location and relative size of the cystic region, the local features determine the exact cyst boundaries.

The proposed architecture follows a multi-stage two-phase approach. The first phase performs convolutions in five stages. Every stage has two convolutional layers followed by a down-sampling layer with max-pooling operation to garner a larger receptive field. The second phase aims to revert the activations of the first phase to the original resolution. This up-sampling is performed with trainable deconvolution layers in 4 stages, thus enabling end-to-end predictions regarding the location of IRCs. For accurate segmentation both local and global features are taken into consideration. Thus, the features extracted from the previous stages, in phase one, are forwarded and concatenated with the features in the corresponding output of up-sampled stages in phase two. The advantage of this architecture is that every pixel is considered as an individual training sample with per-pixel back-propagation error. This process significantly increases the size of the training data set, thereby avoiding model over-fitting.

The proposed network architecture is shown in Fig. 2. The model is designed to process images of size $[256 \times 512]$. The convolution filter width and height ($[\omega \times \eta]$) is chosen as $[3 \times 3]$ across the entire network. This filter size is motivated by prior work [35] to ensure discriminative feature learning from pixel neighborhood while ensuring low parametrization when compared to larger filter sizes.

In the analysis/down-sampling phase, there are two $[3 \times 3]$ convolutions before a $[2 \times 2]$ max-pooling layer which reduces the resolution of the image exactly by half. All convolutions are followed by a Rectified Linear Unit (ReLU) activation function and batch normalization operation. The addition of batch normalization [36] speeds up the network training and convergence process. The number of filters (k) in each convolutional layer also doubles after every stage. The second phase of the network up-samples the activations using deconvolution or fractionally-strided convolution. It follows a path symmetric to the down-sampling path but the pooling operations are replaced with deconvolutions. The final layer in the network is a $[1 \times 1]$ convolution layer. The loss function of the network is computed by pixel-wise sigmoid activation function over the output of the final $[1 \times 1]$ layer with the binary cross entropy loss, which maps the input to a probability map having the same dimensions as the input image. The loss function of the network for a single input image with n_{out} pixels is given in (3),

$$L = -\sum_{i=1}^{n_{\mathrm{out}}} \left(t_i \log\left(s_i\right) + \left(1 - t_i\right) \log\left(1 - s_i\right)\right) \quad (3)$$

where t_i is the actual binary output (target) and s_i is the predicted binary output for the pixel i. The sigmoid activation function s applied on weighted sum of the inputs and the final output y are given in (4).

$$s_i = \frac{1}{1 + e^{-y_i}}, \qquad y_i = \sum_{j=1}^{n} x_j w_{ji} \quad (4)$$

Here, n signifies the number of neurons present in the layer just before the sigmoid activation. Hence, the model output represents a pixel-wise likelihood of being cystic or non-cystic. The resulting binary image mask is then compared with the actual target t cyst mask, and with the aid of the loss function L, model weights are updated such that the loss is minimized for subsequent epochs. Thus, the FCN model is trained.

The proposed model [see Fig. 2] is designed with 18 total convolutional layers (denoted by red arrows in Fig. 2) across the two phases. The number of filters (k) for the 2 starting

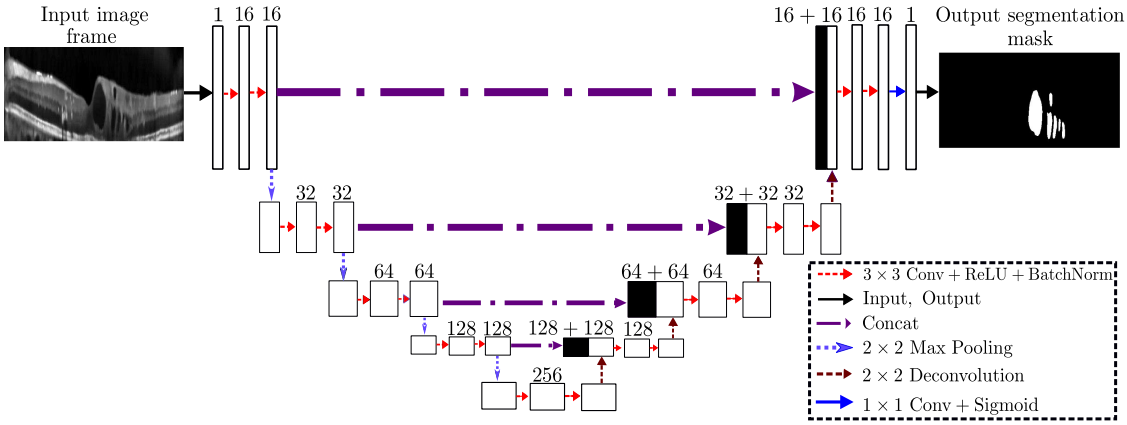

Fig. 2. The proposed network architecture. The FCN model comprises of 18 convolutional layers with the number of kernels varying from 16 to 256.

TABLE II
DATASET DESCRIPTION

Vendor	Train set	Test set	Total number of B-scans per volume
Cirrus	4	4	128–200
Nidek	3	3	7–128
Spectralis	4	4	7–49
Topcon	4	4	128

TABLE III
DIFFERENT FCN ARCHITECTURES FORMED BY VARYING NUMBER OF LAYERS

# Layers	Starting Filter Size ($\omega \times \eta \times k$)	*Depth*	# Parameters
4	$3 \times 3 \times 64$	0	112,513
6	$3 \times 3 \times 64$	1	367,809
10	$3 \times 3 \times 64$	2	1,866,817
14	$3 \times 3 \times 64$	3	7,707,457
18	$3 \times 3 \times 64$	4	31,054,145

layers are assigned as 16. After each pooling layer, the number of filters are doubled for the next 2 convolutional layers. There are 4 pooling layers, making the maximum number of filters in a convolutional layer to be 256. In the up-sampling phase, the number of convolutional filters is reduced by half after every up-sampling operation. Similar to the first phase, this is also performed 4 times, leaving the last 2 convolutional layers with $k = 16$ filters each. This is followed by a $[1 \times 1]$ convolutional layer with a filter to produce a $[256 \times 512]$ sized output image mask, to match the input dimensions.

C. Data

To evaluate the performance of the proposed method, experiments are conducted on the benchmarked OPTIMA cyst segmentation challenge dataset [21]. The OPTIMA cyst challenge dataset contains OCT scans with cystoid macular edema obtained using four different vendors, namely Zeiss Cirrus, Nidek, Spectralis Heidelberg and Topcon [21]. This dataset consists of separate training and test subsets. Each OCT volume is acquired over 6 × 6 mm of the macula and foveal center from subjects with CME. Table II shows the dataset description. A sample OCT B-scan frame from each vendor is shown in Fig. 1.

III. EXPERIMENTAL SETUP

The proposed FCN architecture is implemented in Keras 1.0 [37] with Tensorflow back-end on a workstation with a 64-bit Ubuntu operating system, Intel Xeon Processor E5-2600 (Intel, Mountain View, CA), solid state hard drive, 128 GB of RAM and the NVIDIA Quadro K2200 GPU with 4 GB dedicated memory.

A. FCN Model Training and Hyper-parametrization

Model hyper-parameters need to be optimally tuned for the final objectives. In the context of FCNs, the most important parameters/hyper-parameters that need to be optimized include: the number of weights and biases in the network, the number of layers, the number of filters/kernels in each layer and the learning rate of the model. The best combination of these parameters is identified by performing grid-search using the hold-out method [38]. Finally, the test data set is used for trained model performance analysis. Several experiments have been designed to validate the FCN model architecture as described in Section II-B. As described in [24], two convolutional layers are always followed by a max-pooling layer or a deconvolution layer. Keeping this feature constant and the number of filters as described in [24], the number of layers can be varied for different architectures. An overview of these architectures are presented in Table III. Here, *Depth* signifies the number of max-pooling layers, *parameters* signify the number of weights and biases in the network, and *layers* specify the convolutional layers.

For experimental evaluation of the FCN parameters on the various architectures, one volume from each vendor is extracted from the training set to act as the validation set, i.e., the training and validation sets contain 11 and 4 volumes, respectively. First, these 3D OCT volumes are separated into 2D B-scans,

which constitute training data of 1243 B-scans and validation data of 433 B-scans across all the vendors. As discussed in the Section II-A, only the ROI between the ILM and BMEIS sub-retinal layers are considered for intra-retinal cyst segmentation. All the images are then resized using bilinear interpolation to a standard resolution of $[256 \times 512]$. After resizing, the OCT scans are normalized to zero mean and unit variance.

The FCN network is trained with the binary cross-entropy loss function (refer (3)). The Adam optimizer [22], which is a variant of stochastic gradient descent, is used to update the weights. The default hyper-parameters used for this optimizer are: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$. The learning rate is empirically chosen as 3×10^{-4} based on a short random search. Initial network weights are assigned according to the *He initializer* [39]. No other pre-trained weights are used in the process. The FCN network is trained from scratch completely on the train data set. Nevertheless, the train data comprises of relatively less number of samples to train a full-scale FCN. Hence, data augmentation is applied to generalize the model and reduce over-fitting. Several data augmentation schemes are evaluated for their impact on training loss and model convergence. These include image gradient enhancement, edge enhancement, brightness and contrast variations. However, we observe that augmented data generated by horizontal flipping, random shear, height, width and zoom shifts are best for IRC segmentation tasks. This observation is intuitive since domain knowledge suggests cysts to have variable shapes, structures and orientations but similar appearance with respective to neighboring tissues. Thus, we observe that domain-specific data augmentation process is key to a generalizable FCN model. The data transformations for augmentation are performed on-the-fly during training, thereby alleviating storage concerns. The proposed model is trained up to 120 epochs, beyond which, no further change in loss function is observed.

Fig. 3 shows the training process for various FCN model architectures. *Depth* 0, 1, 2 models [see Fig. 3(a)–(c)] exhibit limited learning capabilities and the training loss does not reduce significantly. *Depth* 3, 4 models [see Fig. 3(d) and (e)] can be further investigated for their receptive fields. The receptive field is defined as the effective area of the original input image covered by a convolution filter [40]. If the convolution filter size is kept the same throughout the network, the receptive field of a filter increases after every pooling operation. This is because the resolution of the image is reduced, but the size of the filter remains the same. Even though both *Depth* 3 and *Depth* 4 models converge to a lower training loss, we choose the latter one for its higher receptive field, which can account for more features. However, in *Depth* 4 model the rate of change in loss function demonstrates some over-fitting trends. Keeping this *Depth* 4 as constant, the number of parameters can be varied by changing the number of filters in each layer, thereby reducing over-fitting. Table IV shows the number of parameters in the different architectures formed at *Depth* 4 by varying the starting filter size, and Fig. 4 shows their training processes, respectively. We observe that both $k = 32$ and $k = 16$ starting filter sizes converge to low training losses with significantly less over-fitting trends. Thus, the model with a fewer number of parameters is chosen as the final architecture, because of higher generalizability and low over-fitting trends. The proposed *Depth* 4 FCN model requires 338 seconds to train per epoch (total 1234 samples in the training data), and 0.06 seconds per test image (total 909 samples in the test data) for prediction.

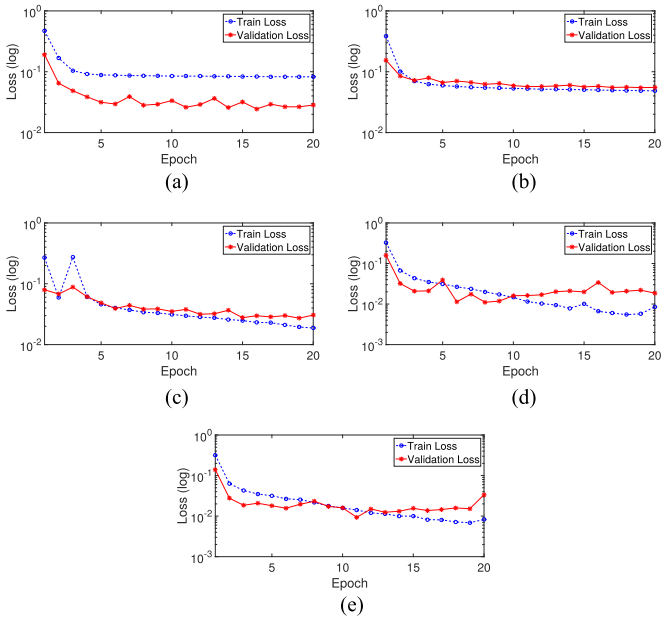

Fig. 3. Loss Value vs. Epoch graphs for FCN model *Depth*: [0, 1, 2, 3, 4] architectures after model training. Blue: Training Loss. Red: Validation Loss. (Plotted in Log Scale). (a) Layers $= 4$, Depth $= 0$. (b) Layers $= 6$, Depth $= 1$. (c) Layers $= 10$, Depth $= 2$. (d) Layers $= 14$, Depth $= 3$. (e) Layers $= 18$, Depth $= 4$.

TABLE IV
ARCHITECTURES DESIGNED BY VARYING NUMBER OF FILTERS IN *Depth* 4

Starting Filter Size ($\omega \times \eta \times k$)	# Parameters
$3 \times 3 \times 64$	31,054,145
$3 \times 3 \times 32$	7,771,297
$3 \times 3 \times 16$	1,946,705

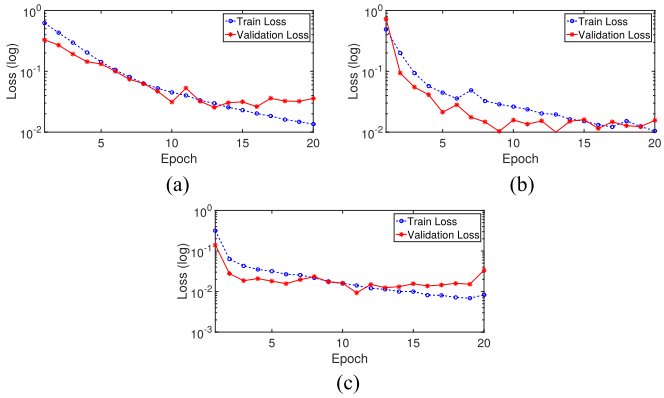

Fig. 4. Loss Value vs. Epochs graphs for *Depth 4* architecture with different starting filter sizes. Blue: Training Loss. Red: Validation Loss. (Plotted in Log Scale). (a) Layers $= 18$, Depth $= 4$, Starting Filter Size $k = 16$. (b) Layers $= 18$, Depth $= 4$, Starting Filter Size $k = 32$. (c) Layers $= 18$, Depth $= 4$, Starting Filter Size $k = 64$.

TABLE V
MEAN (STANDARD DEVIATION) OF PRECISION AND RECALL OF THE PROPOSED METHOD ON DIFFERENT VENDOR OCT SCANS

Vendor	G_1			G_2			$G_1 \cap G_2$		
	Precision	Recall	Dice	Precision	Recall	Dice	Precision	Recall	Dice
Cirrus	0.71 (0.21)	0.61 (0.38)	0.62 (0.35)	0.71 (0.21)	0.61 (0.38)	0.62 (0.35)	0.67 (0.18)	0.66 (0.40)	0.63 (0.35)
Nidek	0.83 (0.02)	0.75 (0.23)	0.77 (0.15)	0.81 (0.01)	0.75 (0.24)	0.76 (0.14)	0.74 (0.03)	0.80 (0.25)	0.75 (0.11)
Spectralis	0.82 (0.09)	0.62 (0.09)	0.71 (0.08)	0.84 (0.09)	0.64 (0.11)	0.72(0.10)	0.79 (0.09)	0.69 (0.10)	0.74 (0.09)
Topcon	0.82 (0.09)	0.70 (0.17)	0.75 (0.13)	0.82 (0.08)	0.73 (0.14)	0.77 (0.11)	0.76 (0.11)	0.78 (0.13)	0.77 (0.12)
Overall	**0.79 (0.12)**	**0.66 (0.22)**	**0.71 (0.20)**	**0.80 (0.12)**	**0.67 (0.22)**	**0.72 (0.19)**	**0.74 (0.12)**	**0.73 (0.23)**	**0.72 (0.19)**

IV. RESULTS AND ANALYSIS

Segmentation results of the proposed method is compared with manually graded ground-truth (GT) provided with dataset images (ascertained by two trained ophthalmologists). We used pixel-wise analysis for calculating True Positives, False Positives and False Negatives. True Positives (TP) are defined as true cystic pixels detected by the algorithm, False Positives (FP) are defined as pixels identified as cysts by algorithm that were not actually cysts, and False Negatives (FN) are defined as true cystic pixels that are undetected by the algorithm. Precision and recall metrics are computed for each of the OCT volumes in the test dataset against GT from two independent graders (G_1, G_2) and their intersection ($G_1 \cap G_2$). Precision and recall metrics are given as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \qquad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

To assess the correlation accuracy of the segmentation, Dice coefficient is computed between segmented results of the algorithm ($Detected$) and ground truths (GTs) for each test OCT volume. Dice coefficient is computed as [41]:

$$\text{Dice coefficient} = 2 \cdot \frac{|Detected \cap \text{GT}|}{|Detected| + |\text{GT}|}. \tag{6}$$

Table V shows the mean precision, recall and Dice coefficient results of the proposed method against G_1, G_2 and $G_1 \cap G_2$. It can be noticed that the proposed method gave highest mean recall rate on scans obtained from Nidek vendor compared to other vendors (0.75 on G_1, 0.75 on G_2 and 0.80 on $G_1 \cap G_2$). The precision of the proposed method is stabilized across the three vendors (Nidek, Spectralis and Topcon). In case of Dice coefficient, scans obtained from Nidek and Topcon result in higher correlation on G_1 (0.77 and 0.75, respectively) and G_2 (0.76 and 0.77, respectively) compared to other two vendors (Cirrus and Spectralis). On Cirrus scans, precision, recall and Dice correlation rates are lower compared to other vendors due to high-intensity variations and noise.

Also, we observe variations in the results obtained against G_1, G_2 and $G_1 \cap G_2$, due to inter-observer variability between manually annotated GTs. The scan intensity variations and retinal vessel shadows often increase the difficulty in cyst identification and delineation for human graders, thus accounting for such inter-observer variability. We observe a variable degree of inter-observer variability across vendors for the IRC segmentation task.

Qualitative results using the proposed segmentation method are shown in Figs. 5 and 6. Images in the first column of these figures show original scans from different vendors, the second column show GTs provided by $G_1 \cap G_2$, and the third column show the result of the proposed FCN model, respectively. We observe that due to intensity variations on the sample Cirrus scan, the boundary between individual cysts are ill-defined [see Fig. 5(a)]. Hence, the proposed segmentation method merges individual cysts into larger cystic region due to small intensity variations in the boundaries of cysts [see Fig. 5(c)]. In the sample Nidek scan, the proposed method over-segments the cystic structures by merging two cysts into a single large cyst, thereby resulting in false positive [see Fig. 5(d) and (f)]. In case of sample Topcon and Spectralis scans [see Fig. 5(g) and (j)], the proposed method finely segments the cystic structures when compared to the other two vendors owing to the clear boundary between the cystic structures [see Fig. 5(i) and (l)].

Fig. 6 shows the limited segmentation performance results of the proposed method on different vendor scans. For the sample Cirrus scan [see Fig. 6(a)–(c)], the proposed method unable to detect cysts due to low intensity variations between the cysts and neighboring non-cystic regions. In the sample Nidek scan [see Fig. 6(d)] the boundary between cysts beside the larger cysts is relatively unclear and the cyst merges with the outer nuclear layer (ONL). Here, the model ignores the cyst by misclassifying it as ONL layer [see Fig. 6(e) and (f)]. In the sample Spectralis scan, the grader had marked a non-cystic region as cystic region on right to the foveal center. Here, the proposed method shows its robustness to such instances by not segmenting this region [see Fig. 6(g)–(i)]. The proposed method segments 6 FPs in the sample Topcon scan due to intensity variations [see Fig. 6(j)–(l)]. These limiting results demonstrate the vital role of scan intensity variations towards OCT cyst segmentation tasks.

The Dice coefficient using the proposed method when compared with the published OPTIMA Cyst Segmentation Challenge results [21] are shown in Table VI. Here, the method proposed by de Sisternes *et al.* [42], uses 34 handcrafted features to train a machine learning model. While designing appropriate features manually across imaging vendors may be manually tedious and insufficient task, we observe that the designed features unable to encompass all the OCT scan variations. In [11], a cascaded CNN working on multiple resolutions is used to segment the IRC. This method takes a lot of time to train 3 different networks independently. Moreover, the selection of patch sizes is found to impact the final segmentation results. Another

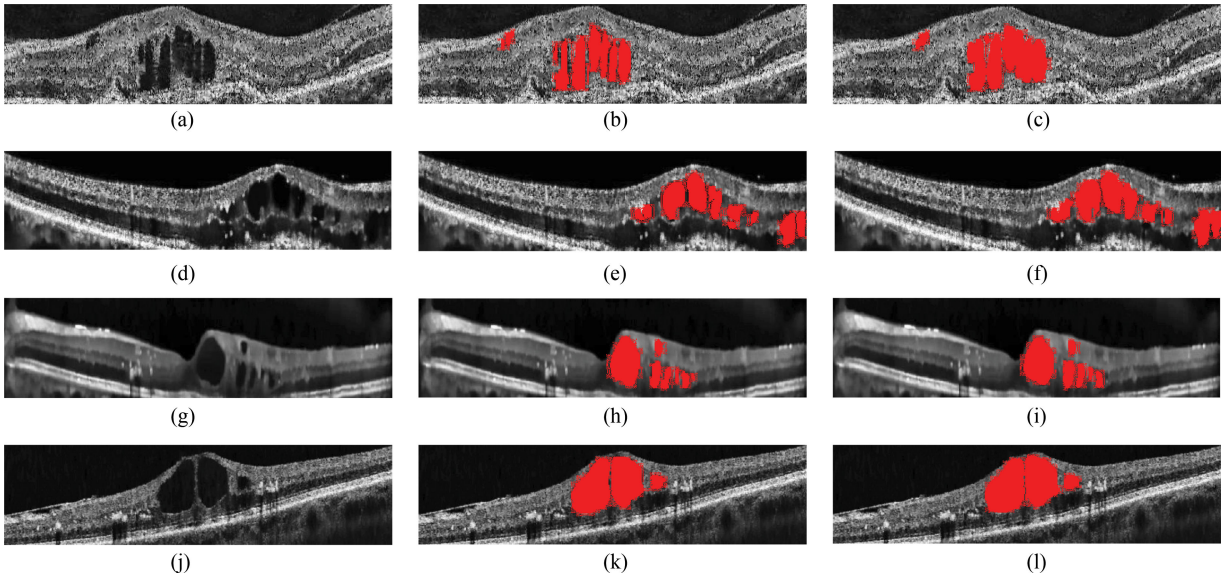

Fig. 5. Results of proposed method on different vendor scans against the GT from $G_1 \cap G_2$. (a) Sample B-scan of Cirrus Vendor. (b) Ground truth. (c) Segmented output. (d) Original B-scan of Nidek Vendor. (e) Ground truth. (f) Segmented output. (g) Sample B-scan of Spectralis Vendor. (h) Ground truth. (i) Segmented output. (j) Sample B-scan of Topcon Vendor. (k) Ground truth. (l) Segmented output.

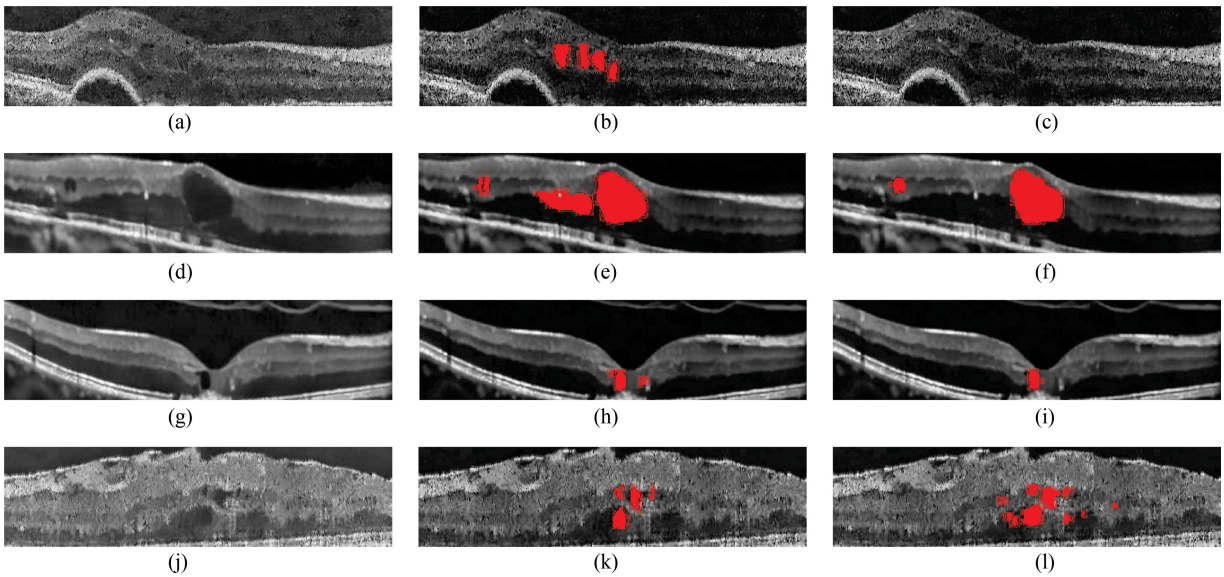

Fig. 6. Limiting segmentation results by the proposed method for each vendor, against the GT from $G_1 \cap G_2$. (a) Sample B-scan of Cirrus Vendor. (b) Ground truth. (c) Segmented output. (d) Sample B-scan of Nidek Vendor. (e) Ground truth. (f) Segmented output. (g) Sample B-scan of Spectralis Vendor. (h) Ground truth. (i) Segmented output. (j) Sample B-scan of Topcon Vendor. (k) Ground truth. (l) Segmented output.

TABLE VI
COMPARISON MEAN (STANDARD DEVIATION) OF DICE COEFFICIENT OF THE PROPOSED METHOD AGAINST THE RESULTS OF OPTIMA CYST SEGMENTATION CHALLENGE ON ALL VENDOR OCT VOLUMES (TEST SET)

Method	G_1	G_2	$G_1 \cap G_2$
de Sisternes *et al.* [42]	0.68 (0.15)	0.67 (0.14)	0.69 (0.15)
Venhuizen *et al.* [11]	0.61 (0.19)	0.60 (0.19)	0.59 (0.19)
Oguz *et al.* [4]	0.60 (0.15)	0.59 (0.15)	0.60 (0.14)
Esmaeili *et al.* [5]	0.55 (0.27)	0.55 (0.27)	0.55 (0.28)
Haritz *et al.* [10]	0.23 (0.15)	0.23 (0.15)	0.23 (0.15)
Proposed Method	**0.71 (0.20)**	**0.72 (0.19)**	**0.72 (0.19)**

method involving two-stage graph-cut segmentation approach is proposed in [4], where the cost function for the graph-cut algorithm is computed using initial intra-retinal layer segmentation results. Since segmentation of inner retinal layers is more prone to error on pathological images than on normal ones, the cost function is found to be impacted significantly. Haritz *et al.* [10] initially identified candidate cyst regions using a center-surround difference technique. In this method, local descriptors are extracted from cystic regions and classified using random forest classifiers. Nonetheless, this rule-based method is found to fail while segmenting small cyst regions. Esmaeili *et al.* [5] pro-

posed automated IRC segmentation method using 3D curvelet transform and K-SVD dictionary learning approaches. However, this method is proposed and validated only on Spectralis vendor OCT scans of the challenge dataset. Generalizing this method for vendor-independent IRC segmentation might affect the dictionary learning process and performance because of the higher level of noise present in other OCT vendor (Cirrus, Nidek, and Topcon) scans.

The method proposed in this paper uses a pre-processing module to equalize the OCT images across vendors. Next, the FCN model automatically captures both micro and macro-level features for better characterization of cystic structures. Thus, the proposed method outperforms reported challenge results with highest Dice rate (0.71 on G_1, 0.72 on G_2 and 0.72 on $G_1 \cap G_2$) with an improvement of 3% on ground truth provided by G_1 and 5% on G_2 compared to results reported in de Sisternes *et al.* [42]. Additionally, we observe slight variations in the segmentation results for the comparative methods on the ground-truth provided by both graders due to inter-observer variabilities.

We also conducted experiments to study the influence of noise and the results are depicted in Section I of the Supplementary material. It can be observed that there is 50% improvement in Dice coefficient by using denoised images instead of raw noisy images. Further, to demonstrate generalizability of the proposed method, the publicly available Duke DME dataset [14] images are used to test the proposed trained FCN model in Section II of the Supplementary material. We observe that without any additional training and fine-tuning, the proposed method achieves Dice coefficient of 0.53 on the Duke DME dataset. Due to the limited size of the Duke DME data set, the proposed FCN model was not retrained further to fit these images.

V. Conclusion and Discussion

This work presents an FCN model-based vendor independent IRC area segmentation technique. The FCN model is customized for IRC area segmentation by utilizing denoised retinal OCT images to train it from scratch. Sensitivity analysis of the model hyper-parameters demonstrates that deeper networks exhibit better feature learning capabilities than shallower networks while higher receptive fields induce higher training losses that may lead to model over-fitting. The proposed optimal FCN model with lower receptive field (starting filter size 16) and *Depth* 4 results in a computationally efficient yet robust vendor-independent model for IRCs. Additionally, domain-specific data augmentation methods are found to improve model training and convergence rates. The proposed model is evaluated qualitatively and quantitatively on OPTIMA cyst challenge dataset. The results demonstrate that the proposed method efficiently segments the IRCs by providing mean Dice rate of 0.71 on scans obtained across four different vendors.

Future works may be directed towards cystic boundary detection or cyst counting tasks by suitably modifying the FCN parameters, goals and loss functions. It is noteworthy that for area-based segmentation tasks, penalizing FP pixels more than FN pixels in the loss function will lead to smaller cysts getting missed while over detections and merged cysts will be observed for higher penalization of FNs over FPs. Future efforts can be directed towards the implementation of such modified proposed architectures for the detection of sub-retinal cysts, pigment epithelial detachments and other macular degenerative pathologies.

Acknowledgment

The authors would like to thank Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology, Medical University of Vienna for providing data.

References

[1] S. R. Irvine, "A newly defined vitreous syndrome following cataract surgery," *Amer. J. Ophthalmol.*, vol. 36, no. 5, pp. 559–619, 1953.
[2] G. R. Wilkins, O. M. Houghton, and A. L. Oldenburg, "Automated segmentation of intraretinal cystoid fluid in optical coherence tomography," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 1109–1114, Apr. 2012.
[3] D. Huang *et al.*, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.
[4] I. Oguz, L. Zhang, M. D. Abràmoff, and M. Sonka, "Optimal retinal cyst segmentation from OCT images," *Proc. SPIE*, vol. 9784, 2016, Art. no. 97 841E.
[5] M. Esmaeili, A. M. Dehnavi, H. Rabbani, and F. Hajizadeh, "Three-dimensional segmentation of retinal cysts from spectral-domain optical coherence tomography images by the use of three-dimensional curvelet based K-SVD," *J. Med. Signals Sens.*, vol. 6, no. 3, pp. 166–171, 2016.
[6] 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.
[7] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abràmoff, and M. Sonka, "Three-dimensional analysis of retinal layer texture: Identification of fluid-filled regions in SD-OCT of the macula," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1321–1330, Jun. 2010.
[8] L. Zhang, W. Zhu, F. Shi, H. Chen, and X. Chen, "Automated segmentation of intraretinal cystoid macular edema for retinal 3d OCT images with macular hole," in *Proc. 2015 IEEE 12th Int. Symp. Biomed. Imag.*, 2015, pp. 1494–1497.
[9] A. Gonzalez, B. Remeseiro, M. Ortega, M. Penedo, and P. Charlon, "Automatic cyst detection in OCT retinal images combining region flooding and texture analysis," in *Proc. 2013 IEEE 26th Int. Symp. Comput.-Based Med. Syst.*, 2013, pp. 397–400.
[10] K. Gopinath and J. Sivaswamy, "Domain knowledge assisted cyst segmentation in OCT retinal images," 2015. [Online]. Available: https://optima.meduniwien.ac.at/fileadmin/Challenge2015/Gopinath-CystChallenge15.pdf
[11] F. Venhuizen, M. J. van Grinsven, B. van Ginneken, C. C. Hoyng, T. Theelen, and C. I. Sanchez, "Fully automated segmentation of intraretinal cysts in 3d optical coherence tomography," *Investigative Ophthalmol. Vis. Sci.*, vol. 57, no. 12, pp. 5949–5949, 2016.
[12] G. N. Girish, A. R. Kothari, and J. Rajan, "Automated segmentation of intra-retinal cysts from optical coherence tomography scans using marker controlled watershed transform," in *Proc. 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 1292–1295.
[13] G. N. Girish, A. R. Kothari, and J. Rajan, "Marker controlled watershed transform for intra-retinal cysts segmentation from optical coherence tomography B-scans," *Pattern Recognit. Lett.*, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865517304658
[14] 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.
[15] A. Lang *et al.*, "Automatic segmentation of microcystic macular edema in OCT," *Biomed Opt. Express*, vol. 6, no. 1, pp. 155–169, 2015.
[16] W. Wieclawek, "Automatic cysts detection in optical coherence tomography images," in *Proc. 2015 IEEE 22nd Int. Conf. Mixed Des. Integr. Circuits Syst.*, 2015, pp. 79–82.

[17] E. K. Swingle *et al.*, "Segmentation of microcystic macular edema in cirrus OCT scans with an exploratory longitudinal study," *Proc. SPIE-Int. Soc. Opt. Photon.*, vol. 9417, 2015, Art. no. 94 170P.
[18] E. K. Swingle, A. Lang, A. Carass, H. S. Ying, P. A. Calabresi, and J. L. Prince, "Microcystic macular edema detection in retina OCT images," *Proc. SPIE-Int. Soc. Opt. Photon.*, vol. 9038, 2014, Art. no. 90 380G.
[19] A. Rashno *et al.*, "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.*, 2018, doi: 10.1109/TBME.2017.2734058.
[20] G. N. Girish, V. Anima, A. R. Kothari, P. Sudeep, S. Roychowdhury, and J. Rajan, "A benchmark study of automated intra-retinal cyst segmentation algorithms using optical coherence tomography b-scans," *Comput. Methods Programs Biomed.*, vol. 153, pp. 105–114, 2018.
[21] "Optima cyst segmentation challenge," 2015. [Online]. Available: https://optima.meduniwien.ac.at/research/challenges/
[22] K. Kamnitsas, L. Chen, C. Ledig, D. Rueckert, and B. Glocker, "Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI," *Ischemic Stroke Lesion Segmentation*, vol. 13, pp. 13–16, 2015.
[23] Y. Lequan, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Apr. 2016.
[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention.*, 2015, pp. 234–241.
[25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 3431–3440.
[26] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 2016 4th Int. Conf. 3-D Vision*, 2016, pp. 565–571.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.
[28] P. V. Sudeep *et al.*, "Enhancement and bias removal of optical coherence tomography images: An iterative approach with adaptive bilateral filtering," *Comput. Biol. Med.*, vol. 71, pp. 97–107, 2016.
[29] P. Sudeep *et al.*, "Speckle reduction in medical ultrasound images using an unbiased non-local means method," *Biomed. Signal Process. Control*, vol. 28, pp. 1–8, 2016.
[30] J. Froment, "Parameter-free fast pixelwise non-local means denoising," *Image Process. On Line*, vol. 4, pp. 300–326, 2014.
[31] 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.
[32] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, Dec. 2010.
[33] M. K. Garvin, M. D. Abramoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1436–1447, Sep. 2009.
[34] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. New York, NY, USA: Academic, 1994, pp. 474–485.
[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, to be published.
[36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[37] F. Chollet *et al.*, "Keras," GitHub, 20156. [Online]. Available: https://github.com/keras-team/keras
[38] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, Stanford, CA, USA, 1995, vol. 14, pp. 1137–1145.
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 1026–1034.
[40] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 4898–4906.
[41] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
[42] L. de Sisternes, J. Hong, T. Leng, and D. L. Rubin, "A machine learning approach for device-independent automated segmentation of retinal cysts in spectral domain optical coherence tomography images," 2015. [Online]. Available: https://optima.meduniwien.ac.at/fileadmin/Challenge2015/Sisternes-CystChallenge15.pdf