A generalizable approach based on the U-Net model for automatic intraretinal cyst segmentation in SD-OCT images

Razieh Ganjee | Mohsen Ebrahimi Moghaddam | Ramin Nourinia

Abstract
In this article, we propose a new U-Net-based approach for intraretinal cyst segmentation across different vendors that improve some of the challenges faced by previous deep-based techniques. The proposed method has two main steps: (1) prior information embedding and input data adjustment, (2) the segmentation model. In the first step, we inject the information into the network in a way that overcomes some of the network limitations in receiving data and learning important contextual knowledge. And in the next step, we introduce a connection module between the encoder and decoder parts that transfers information more effectively from the encoder to the decoder. Two public datasets, namely, OPTIMA and KERMANY, are employed to evaluate the proposed method. The results show that the proposed method is an efficient vendor-independent approach for the segmentation of intraretinal cystoid fluid with mean Dice values of 0.78 and 0.81 on the OPTIMA and KERMANY datasets, respectively.

KEYWORDS
intraretinal cyst, optical coherence tomography (OCT), segmentation, U-Net

1 INTRODUCTION
Optical coherence tomography (OCT) is a non-invasive imaging technology used to diagnose ocular diseases. This imaging modality is a useful tool for visualizing morphological retinal tissue variations that occur due to macular disease. One of the macular pathologies that are efficiently visualized using this imaging modality is cystoid macular edema (CME). CME occurs due to vascular defects that lead to fluid leakage into the retina. If CME patients are not treated in time, they may lose their central vision forever. Therefore, automatic segmentation of intraretinal cystoid fluid (IRC) or cysts is valuable and can help ophthalmologists to assess the treatment progress efficiently. However, automatic segmentation of cysts obtained by various vendors is a challenging problem because OCT images captured by various vendors have different resolutions, intensity variations, and noise levels.

Over the past years, several automatic methods have been presented for the automatic segmentation of cysts but few of them such as are vendor-independent. For the first time, in 2015, OPTIMA cyst segmentation challenge published a database with a variety of retinal cysts provided by four different vendors consisting of Cirrus, Topcon, Spectralis, and Nidek to evaluate segmentation methods in terms of vendor independence. Different approaches participated in this challenge including a simple classifier trained on the 34 hand-crafted features, a patch-based CNN method, a rule-based method by employing center-surrounded filters and random forest classifier, and two unsupervised graph-based and curvelet-based methods. Among these methods, the rule-based method could not operate...
as well as others on the test set and failed to segment IRCs appropriately. The method proposed by Esmaeili et al.\textsuperscript{26} was vendor-specific and was only evaluated on Spectralis images. The only deep-based approach that participated in this challenge was proposed by Venhuizen et al.\textsuperscript{18} They used three different patch sizes to manage the size variety of cysts. Hence, training in this method is memory and time-consuming, and also the efficiency of that is strongly dependent on the patch size selection. De Sisternes et al.\textsuperscript{13} were the winner of this challenge. However, this method needs accurate segmentation of retinal layers, while middle retinal layers segmentation in the presence of cysts is an open challenging problem.

After the Optima challenge, the vendor-independent IRC segmentation development was also addressed in later works presented in Refs. [22–24]. A cascade of two U-Net architecture-based networks was employed in Ref. [22]. They employed the first U-Net for the segmentation of the retina and then integrated this output with the second U-Net for segmenting of IRCs. Another vendor-independent method was published in Ref. [23]. They attempted to customize U-Net parameters such as the optimal number of layers and the optimal kernel dimensions. In Ref. [24], a generalized motion pattern (GMP) and deep learning were employed for the segmentation of cysts. The recent IRC segmentation methods were proposed in Ref. [27] and our previous paper [28]. In Ref. [27], three types of fluids, including IRC, PED (Pigment Epithelial Detachment), and SRF (subretinal fluid), were segmented by employing a large training set captured by three different types of OCT devices. In this work, a separate network based on the U-Net was trained on images from each vendor. Hence, their method is not vendor-independent. However, they could obtain the first rank in the MICCAI RETOUCH\textsuperscript{29} challenge. And in our previous unsupervised approach,\textsuperscript{28} IRCs were segmented by limiting the searching space toward the targets in three levels of a hierarchical framework. Although our previous method could obtain valuable results and be competitive with the deep-based methods, deep learning-based methods are more of interests in recent years.

As it was reviewed, the previous deep learning-based methods followed two different approaches to dealing with the IRC segmentation in SD-OCT images captured by different types of devices; in one approach, input images are asymmetrically resized to an equal size,\textsuperscript{22–24} and in the other approach, a separate network is trained for each of the vendor.\textsuperscript{27} In the first approach, asymmetrically resizing causes to loss of small cysts, merging cysts placed near each other and preventing lesions to appear with their real shape, size, and location. And the second approach is not vendor-independent because of using a separate network for each vendor. In this study, we proposed a vendor-independent approach based on the U-Net architecture that is summarized in two parts: (1) prior information embedding and input data adjustment and (2) IRC segmentation model. In the first part, while the asymmetrically resizing problem was solved, we provided the possibility of expert knowledge injection to the network to manage some of the limitations of U-Net-based models in learning the location of target objects and receiving data with the same size. And in the second part, by adding a connection module to the core of the U-Net architecture, between encoder and decoder parts, we both increased field-of-view (FOV) and changed the focus of the model on salient regions. Therefore, this connection module helps U-Net to transfer more useful and meaningful features from encoder to decoder part resulting in more accurate segmentation. As a summary, the main contribution of this article can be highlighted as follows:

- A new U-Net-based method is presented for IRC segmentation.
- Adjustment of input data to make U-Net IRC segmentation model robust across the input image size.
- Embedding prior knowledge to avoid U-Net limitations in learning location features.
- Proposing a connection module in the U-Net model to extract more useful features.
- Generalizability across different vendors and sites.

In the following, first, the proposed method is explained in detail in part 2, then the evaluation results are presented in part 3, and finally, in parts 4 and 5, the discussion and conclusion are discussed, respectively.

2 | PROPOSED METHOD

The proposed method framework is shown in Figure 1. It contains two main parts: (1) prior information embedding and input data adjustment and (2) the IRC segmentation model. In the first part, by applying zero-padding, we provided the possibility of feeding OCT images to the neural network with their effective size. In addition, we employed expert knowledge to compensate for the side effects of zero-padding and also drive the network to learn the location of abnormalities more efficiently. In the second part, an extended structure of the U-Net model was implemented by applying a connection module between the encoder and decoder parts, indicated with black dash dots in Figure 1, comprise of attention gates (AGs)\textsuperscript{30} in the skip connection layers and atrous spatial pyramid pooling (ASPP) structure\textsuperscript{31} in the bottleneck layer. In this architecture, through the AGs, the
network can focus more on the area where the cysts occur, and through the ASPP structure, by multi-scale FOV enlargement, it can obtain more efficient multi-scale contextual features without more decreasing the resolution of the feature map. In the following, each part will be explained in detail.

2.1 | Prior information embedding and input data adjustment

In this section, the first purpose is to prepare the dataset such that OCT images are injected into the neural network with an effective size and keeping their aspect ratio instead of asymmetrically resizing to the equal dimensions; and the second purpose is to embed prior information as an extra channel to each input image. To achieve these goals, after enhancing the input images, first, the region of interest (ROI) that is a region prone to the cyst occurrence is extracted, then by considering the largest image dimension in different vendors as a reference dimension, preprocessed and ROI images are dilated to this dimension by zero-padding. Two following subsections describe this step in more detail.

2.1.1 | Denoising

The varying degree of speckle-noise degrades OCT images. we employ conventional bilateral filtering to
reduce speckle noise; because this simple denoising model can smooth noisy scans and keep edge data with relatively low time complexity. By applying this filter to the input image \( I(x) \), the denoised image \( I^d(x) \) is computed as follows:

\[
I^d(x) = k^{-1}(x) \sum_{x \in \Omega} I(x_i) G_{\sigma_d}(\|x_i - x\|) G_{\sigma_r}(I(x_i) - I(x)),
\]

where \( \sigma_r \) and \( \sigma_d \) are standard deviations of Gaussian functions named respectively as range and spatial parameters. \( \|\cdot\| \) shows Euclidean distance; \( \Omega \) indicates neighborhood pixels with center \( x \), and \( k = \sum_{x \in \Omega} (\|x_i - x\|) G_{\sigma_r}(I(x_i) - I(x)) \) is a normalization function.

As it is clear in Equation (1), using bilateral filtering, each noisy image pixel \( x \) enhanced by computing both Euclidean distance and intensity similarity of its nearby pixels and replacing its value with a weighted intensity average of close pixels. The result of applying this filter on an OCT sample is shown in Figure 2A,B. As can be seen, this filter could properly enhance the input image quality.

2.1.2 | ROI segmentation and zero-padding

One of the limitations of U-Net-based segmentation models is receiving the same size input data. This feature is challenging in cases such as vendor-independent IRC segmentation where images have different resolutions and dimensions. In this work, to preserve the originality of OCT images, zero-padding is used instead of asymmetric input image resizing. In this way, the largest possible dimensions of an OCT image in horizontal and vertical coordinates are considered reference dimensions, and then by applying zero-padding, each OCT image is embedded in the reference dimension. To compensate for the side effects of zero-padding and also injecting the prior information of the IRC location into the FCN model, each input data is enriched by adding a second channel with the information of ROI in addition to the original image. This strategy not only suppresses irrelevant information in the input image but also improves another limitation of the U-Net model in learning where cysts can be located.

Based on expert knowledge, IRCs happen in a limited area from the internal limiting membrane (ILM) layer to the retinal pigment epithelium (RPE) layer. Since in SD-OCT images these layers emerge with high contrast, a simple unsupervised graph-based method is applied for extraction of them. In this way, first, each OCT image is mapped to an undirected weighted graph and then desired layers are segmented by detecting two shortest graph paths using Dijkstra’s algorithm. The weighted graph is constructed by mapping each image pixel to a graph node and defining each graph edge as a linkage between two neighbor pixels. The vertical gradient is used to obtain the weight of each edge because, in retinal OCT visualization, pixel intensities from one layer to another layer vary in the vertical direction. Hence, each edge weight \( w_{ab} \) that connects two neighboring pixels \( a \) and \( b \) is obtained as follows:

\[
w_{ab} = 2 - (g_a + g_b) + w_{\text{min}},
\]

where \( w_{\text{min}} \) is a constant that is set to a small value, \( g_a \) and \( g_b \), respectively, indicate the vertical gradient values of pixels \( a \) and \( b \).

FIGURE 2 Preprocessing and region of interest (ROI) segmentation on an optical coherence tomography (OCT) sample. (A) Original image, (B) enhanced image, (C) internal limiting membrane (ILM) (blue) and retinal pigment epithelium (RPE) (green) segmentation, and (D) ROI image.
After constructing the weighted graph model, one of the desired layers is segmented by extracting the first shortest graph path using Dijkstra’s algorithm. Then, the constructed graph is cut into two subgraphs using this segmented layer. If the area above the cut consists of hyper-reflective data, the segmented layer is RPE, and hence, the top subgraph is considered for ILM segmentation. Otherwise, it is ILM, and thus, the bottom subgraph is investigated for RPE segmentation. ROI is considered as the area restricted between these two layers. Figure 2C,D shows ROI segmentation in an enhanced OCT sample shown in Figure 2B. The ROI image is also padded with zero and then integrated with the corresponding zero-padded image to make a two-channel input image for the next processing.

2.2 The IRC segmentation model

U-Net is a well-known FCN model that is widely used for medical image segmentation. Despite its success, due to the symmetric encoder-decoder architecture with skip connections that let to employ a combination of global and local features in image segmentation, it has limitations in learning the location of the target object and increasing FOV without decreasing the resolution of the feature map. Hence, we proposed a connection module in the core of U-Net architecture, where the encoder connects to the decoder, that uses AGs in the skip connections and ASPP after the first convolution in the bottleneck layer. By integrating AGs in the skip connections, the network is driven to learn where lesions can happen and by embedding an ASPP in the bottleneck layer, the receptive field at the high-level feature map is extended with various dilation rates that led to enriching the network with more multi-scale contextual features without more decreasing in the feature map resolution. As is shown in Figure 1, our FCN model is composed of three parts: encoder, decoder, and connection module. The encoder part is a contracting path to capture contextual features at different resolutions. It is composed of multiple down-sampling layers. Every down-sampling layer consists of two convolutional layers followed by a max-pooling operation. The decoder part is an expanding path to localize target objects, including boundaries and contours. This part is composed of multiple up-sampling layers so that every up-sampling layer consists of two convolutional layers followed by a de-convolutional layer. And the connection module with AG and ASPP concatenates symmetrical contextual and positional features obtained from convolution and de-convolution layers of encoder and decoder parts, respectively. Every AG in the connection module receives two inputs provided by the last convolution in each encoder layer \( X^i \) and the de-convolution in each decoder layer \( g \) and then computes the output \( X^i \) based on the attention coefficient \( a^i \) as follows:

\[
\hat{X}^i = a^i X^i,
\]

where \( a^i = \sigma_2 \left( \psi^T \left( \sigma_1 \left( W_{x} X^i + W_{g} g + b_{xg} \right) \right) + b_{\psi} \right),
\]

and the atrous convolution is defined as:

\[
y[i] = \sum_k x[i + rk] w[k],
\]

where \( r \) is a dilation rate that determines the number of zero values that should be added between two neighboring values of the convolution kernel, standard convolution occurs when \( r = 1 \). Performing the atrous convolution instead of the standard convolution allows us to enlarge the field-of-view and capture more efficient multi-scale contextual feature maps. These atrous convolution results are concatenated to each other, and then by passing through a standard 1*1 convolution, multi-scale contextual features are integrated and sent to the decoder part.

The proposed FCN model was trained end-to-end with the binary cross-entropy loss function. The specifics of training will describe in more detail in the experimental setup and measures subsection.

3 EXPERIMENTAL RESULTS

3.1 Dataset

Three publicly available databases, namely, OPTIMA, KERMANY, and UMN were employed.
in this work. The OPTIMA dataset contains 30 OCT volumes divided into training and testing sets, each set containing 15 volumes that have been captured from CME subjects by employing four vendors including Cirrus, Topcon, Spectralis, and Nidek. Both training and testing sets are included with three OCT volumes from Nidek and four volumes from each of the other three vendors. The total number of scans provided in this dataset is equal to 1560 and 909 in the training and testing set, respectively. In the UMN and KERMANY datasets, only Spectralis device was used for imaging, and they respectively contain 725 and 256 scans that were imaged from DME patients. Two trained experts manually delineated the ground truths in the OPTIMA and UMN datasets. In contrast, three raters were employed to perform this task in the KERMANY dataset. To assess the expert dependency of these datasets, we evaluated interobserver variability across the OPTIMA and UMN datasets using DC in Table 1. Providing this evaluation was not possible in the KERMANY dataset because three ground truths have not been shared separately (only the intersection of ground truths is provided along with this dataset). As can be seen, there is a good interobserver agreement in the OPTIMA datasets and UMN datasets. In contrast, three raters were employed to perform this task in the KERMANY dataset.

To have a fair comprising with state-of-the-art methods, the UMN dataset and the training set of the OPTIMA dataset were used to train the network, and the KERMANY dataset and the OPTIMA testing set were used for evaluations.

### 3.2 Experimental setup and measures

To denoise the input images, \( \sigma_r \) is estimated adaptively by automatic extraction of a patch from the region above the ILM layer. To control the degree of image smoothing and edges blurring, the \( \sigma_d \) parameter, which specifies the filter size, was set near the default value equal to 2\(^3\); for ROI segmentation, a weighted undirected graph was used to segment ILM and RPE layers. The weights of the graph were obtained using Equation (2) with the minimum weight \( w_{\text{min}} \) set to 10\(^{-5}\), and to unify the size of the images obtained from different vendors, all preprocessed images and their corresponding ROI images are passed through a padding module to construct two-channel images with the equal size of \( 640 \times 1024 \times 2 \). Preprocessed and ROI images are included in the center of this reference dimension.

In the segmentation model, the proposed extended U-Net model has two main components, the base U-Net, and the connection module. Although the higher number of convolutional layers in the base U-Net results in more accurate segmentation maps, it causes to increase in model parameters that can lead to overfitting and more computational cost. Hence, to have a tradeoff between the efficiency of the segmentation and computational cost, four convolutional layers in the base U-Net were chosen that extended feature dimensions from 16 in the first layer to 128 feature maps in the bottleneck; consequently, the connection module has four layers: three layers with the AGs and one layer with the ASPP. All of the convolutions used in the base U-Net are \( 3 \times 3 \), and in the gate layers, three \( 1 \times 1 \) convolutions were employed; in the ASPP layer, five atrous convolutions with the size of \( 3 \times 3 \) and dilation rates \{1, 2, 4, 8, 16\}, and one \( 1 \times 1 \) standard convolution were considered. To avoid overfitting, dropout rates were set to 0.1 in the first and second layers, and 0.2 in the third and fourth layers. And finally, our extended U-Net model was trained with a batch size of 10 and iterated for 100 epochs from scratch using the binary cross-entropy loss function. The weights were updated using the Adam optimizer with default parameters, and the learning rate was empirically set to \( 10^{-3} \).

The output of the proposed method is a probability map that indicates cyst prediction scores for each pixel. In the test phase, this probability map is thresholded with a threshold value of 0.5.

Recall, precision, and dice coefficient (DC) were used for performance evaluation defined as follows:

\[
\text{recall} = \frac{TP}{TP + FN}, \quad \text{precision} = \frac{TP}{TP + FP},
\]

### Table 1 Mean (standard deviation) of interobserver variability on SD-OCT images across different vendors in the OPTIMA dataset and Spectralis device in the UMN dataset.

|          | Cirrus | Nidek | Spectralis | Topcon | Spectralis |
|----------|--------|-------|------------|--------|------------|
| Interobserver | 0.95   | 0.85  | 0.87       | 0.91   | 0.98       |
| Variability  | (0.01) | (0.08) | (0.12)     | (0.02) | (0.03)     |
\[ DC = 2 \frac{|\text{Segmentation} - \text{result} \cap \text{Ground} - \text{truth}|}{|\text{Segmentation} - \text{result}| + |\text{Ground} - \text{truth}|}, \quad (8) \]

where \( TP \) and \( FP \) are, respectively, the number of image pixels that are correctly and wrongly classified as cystoid pixels; and \( FN \) indicates cystoid pixels that are not extracted by the automatic method.

### 3.3 Results

The efficiency of the presented method was evaluated quantitatively and qualitatively across two public OPTIMA and KERMANY datasets.

Table 2 shows the results of the presented method on the OPTIMA test set using mean (standard deviation) recall, precision, and DC measures across manual segmentations prepared by expert1 (GT1), expert2 (GT2), and the intersection of them (GT1 \( \cap \) GT2). As is reported in this Table, the highest performance of the proposed method was obtained on SD-OCT images from the Spectralis device with a mean DC of 0.87. And the lowest efficiency is related to the Nidek with a mean DC of 0.69 due to the more intensity variations and ambiguous boundaries compared to the other vendors.

Table 3 compares our method performance to the participants in the OPTIMA challenge,25 some other recent methods,19,22–24 and our previous method28 on OPTIMA image set, and Table 4 shows results on KERMANY dataset compared to our previous method28 and other recent method published in Ref. [27]. To show the efficiency of our extended U-Net model, three U-Net-based architectures including standard U-Net, U-Net with AGs (U-Net + AG), and U-Net with ASPP (U-Net + ASPP) were implemented in the framework of the proposed method and

| Table 2 | Mean (standard deviation) recall, precision, and DC results of the presented method on OPTIMA test images separately for every four vendors. |
|---------|----------------------------------------------------------------------------------------------------------|
|         | \( \text{GT1} \) | \( \text{GT2} \) | \( \text{GT1} \cap \text{GT2} \) |
| Recall | Precision | DC | Recall | Precision | DC | Recall | Precision | DC |
| Cirrus | 0.86 (0.03) | 0.84 (0.12) | 0.77 (0.11) | 0.86 (0.02) | 0.85 (0.12) | 0.77 (0.11) | 0.88 (0.02) | 0.84 (0.12) | 0.78 (0.11) |
| Topcon | 0.85 (0.21) | 0.83 (0.28) | 0.76 (0.15) | 0.88 (0.21) | 0.81 (0.28) | 0.76 (0.16) | 0.90 (0.21) | 0.79 (0.28) | 0.76 (0.16) |
| Spectralis | 0.85 (0.19) | 0.96 (0.20) | 0.86 (0.18) | 0.86 (0.16) | 0.96 (0.21) | 0.87 (0.17) | 0.87 (0.14) | 0.95 (0.24) | 0.87 (0.18) |
| Nidek | 0.88 (0.06) | 0.69 (0.16) | 0.73 (0.18) | 0.87 (0.05) | 0.67 (0.13) | 0.72 (0.16) | 0.92 (0.06) | 0.60 (0.09) | 0.69 (0.14) |
| All | 0.86 (0.12) | 0.83 (0.19) | 0.78 (0.16) | 0.87 (0.11) | 0.82 (0.19) | 0.78 (0.15) | 0.84 (0.11) | 0.79 (0.18) | 0.78 (0.15) |

| Table 3 | The presented method comparison in terms of mean (standard deviation) of DC on OPTIMA dataset with the optima cyst challenge,25 recent deep-based (22–24), and unsupervised methods (19,28), and the presented method implementation with three U-Net based models. |
|---------|----------------------------------------------------------------------------------------------------------|
| Method | Approach | year | \( \text{GT1} \) | \( \text{GT2} \) | \( \text{GT1} \cap \text{GT2} \) |
| Oguz et al. [21] | Unsupervised | 2016 | 0.48 (0.25) | 0.48 (0.22) | 0.48 (0.22) |
| Esmaeili et al. [26] | | 2016 | 0.46 (0.25) | 0.45 (0.24) | 0.45 (0.25) |
| Rashno et al. [19] | | 2018 | 0.70 (0.10) | 0.71 (0.11) | 0.72 (0.10) |
| Ganjee et al. [28] | | 2020 | 0.74 (0.14) | 0.73 (0.14) | 0.74 (0.14) |
| de Sisternes et al. [13] | Supervised | 2015 | 0.64 (0.14) | 0.63 (0.14) | 0.65 (0.15) |
| Venhuizen et al. [18] | | 2016 | 0.56 (0.20) | 0.55 (0.22) | 0.54 (0.20) |
| Haritz et al. [9] | | 2016 | 0.14 (0.08) | 0.14 (0.08) | 0.14 (0.08) |
| Venhuizen et al. [22] | Supervised (deep) | 2018 | – | – | 0.74 (0.16) |
| Grish et al. [23] | | 2018 | 0.71 (0.20) | 0.72 (0.19) | 0.72 (0.19) |
| Gopinath et al. [24] | | 2018 | 0.67 (0.17) | 0.68 (0.17) | 0.69 (0.18) |
| Presented method with U-Net | | 2021 | 0.71 (0.16) | 0.71 (0.16) | 0.72 (0.16) |
| Presented method with U-Net + AG | | 2021 | 0.74 (0.14) | 0.74 (0.14) | 0.74 (0.14) |
| Presented method with U-Net + ASPP | | 2021 | 0.76 (0.13) | 0.76 (0.12) | 0.76 (0.13) |
| Presented method with U-Net + connection module | | 2021 | 0.78 (0.16) | 0.78 (0.15) | 0.78 (0.15) |
their results were also included in Tables 3 and 4. As shown, on the OPTIMA dataset the proposed method succeeds to outperform the compared methods with a mean DC value of 0.78 on all ground truths. And on the KERMANY dataset, the proposed method can operate better than our previous method with 3% improvement of DC, although it is slightly inferior in comparison with the Lu et al.\textsuperscript{27} method due to the large difference in the training set size and also vendor-specific training strategy in Ref. [27]. It should be mentioned, the papers published in Refs. [19,26] only used the train set of the Spectralis device to evaluate their method on the OPTIMA dataset.

Qualitative evaluation of the proposed method is presented in Figures 3–7 by employing four samples provided by each vendor in the OPTIMA dataset (Figures 3–6) and a sample from the KERMANY dataset (Figure 7). In each Figures 3–6, (A) shows the preprocessed image; (A) and (C) show the first and the second ground truth provided by the expert1 and expert2, respectively; and in the third to sixth rows, each pair (D, E), (F, G), (H, I), and (J, K) show respectively the probability map and the final segmentation map obtained from the implementation of the proposed method with standard U-Net, U-Net+AG, U-Net+ASPP, and U-Net+connection module. In Figure 7, (B) show the ground truths intersection because separate grader maps were not published for KERMANY dataset; the second to fourth rows show the result of the proposed method with three mentioned U-Net based models and the proposed method with connection module. In qualitative evaluation, samples were chosen that can indicate the cysts segmentation challenges including segmentation of images with variable cyst size, low contrast cysts, and images with cysts located near each other with indistinct boundaries.

Samples containing various cyst sizes are seen in all figures except Figure 6, as it is observed the proposed model + connection module is more efficient in segmenting cysts with different sizes compared to others, especially in Figures 5 and 7 where small cysts appear with low contrast. Samples with low contrast cysts are also seen in Figure 6. As it is obvious, due to employing more effective contextual features through enlargement of the field-of-view, the presented method and U-Net+ASPP are more successful compared to the standard U-Net and U-Net+AG models, in addition, in comparison with the U-Net+ASPP, the segmentation result provided by the proposed method with connection module has a closer correlation with both ground truths. Figures 3 and 4 contain samples with ill-defined cysts boundary located near each other. As shown, totally the presented method has better performance in segmenting IRCs up to their true boundary with the least false positive detection. Although in Figure 4, it seems that U-Net+AG has more separability power in detecting the boundary between cyst areas, this model could not delineate lesions up to their true boundary. In addition, there is a big disagreement between the two graders in this sample so that the cyst map provided by our method is more correlated with the second ground truth (Figure 4C) while the U-Net+AG result has more correlation with the first ground truth (Figure 4B).

### 4 DISCUSSION

U-Net is a popular model in medical image segmentation due to its symmetric architecture that sees contextual features along with local information. However, U-Net-based models have limitations in learning information such as shape, size, and location (i.e., where the target object can happen); information that exists as prior knowledge in the context of problems, and if it is properly injected into the model, more accurate results will be provided.\textsuperscript{40}

In this article, we tried to use these effective features during the learning process compared to the standard model. In this way, we proposed a two-step approach that can receive this useful information both at the data and model levels. At the data level, we extended the number of input channels so that the first channel corresponds to the original image, and the extra channels can incorporate domain expertise with the data. By this strategy, we also could manage the limitation of the U-Net in receiving data of the same size; one of the challenges in deep-

| Method | Approach | year | GT |
|--------|----------|------|----|
| Ganjee et al. [28] | Unsupervised | 2020 | 0.79 (0.15) |
| Lu et al. [27] | Supervised (deep) | 2019 | 0.82 (0.15) |
| Presented method with U-Net | | 2021 | 0.78 (0.17) |
| Presented method with U-Net+AG | | 2021 | 0.80 (0.14) |
| Presented method with U-Net+ASPP | | 2021 | 0.79 (0.15) |
| Presented method with U-Net+connection module | | 2021 | 0.81 (0.15) |

\textsuperscript{27} The presented method comparison in terms of mean (standard deviation) of DC on KERMANY dataset with the recent deep-based method,\textsuperscript{27} our previous unsupervised method,\textsuperscript{28} and the presented method implementation with three U-Net based models.
based vendor-independent cyst segmentation methods. And at the model level, we extended the standard U-Net with a connection module enriched with AGs and ASPP that drive the network to find salient regions, avoid irrelevant features, and learn where a target object can happen, and also see more meaningful contextual features to segment lesions with various size and more accurate boundary.

As the results showed, our method is generalizable across different sites and vendors; because the presented method could obtain accurate results on the unseen dataset (KERMANY dataset) provided from a different site;
and it also can operate successfully on images provided by different vendors. In comparison with the previous vendor-independent deep learning-based methods, vendor independency has been evaluated more precisely in our method because we did not asymmetrically resize input images, and therefore, we did not lose small cysts or we did not connect nearby cysts. And the ability of the segmentation model was evaluated by the original ground truth, not the resized version. In addition, any post-processing step was not employed in our method that confirms the effectiveness of the prior information embedding as an extra input channel and connection module segment...
FIGURE 6  Topcon sample qualitative results. (A) Preprocessed image, (B) GT1, (C) GT2, (D) U-Net probability map, (E) U-Net segmentation result, (F) U-Net+AG probability map, (G) U-Net+AG segmentation result, (H) U-Net+ASPP probability map, (I) U-Net+ASPP segmentation result, (J) the proposed method + connection module probability map, and (K) the proposed method + connection module segmentation result.

FIGURE 7  KERMANY sample qualitative results. (A) preprocessed image, (B) GT1 ∩ GT2 ∩ GT3, (C) U-Net probability map, (D) U-Net segmentation result, (E) U-Net+AG probability map, (F) U-Net+AG segmentation result, (G) U-Net+ASPP probability map, (H) U-Net+ASPP segmentation result, (I) the proposed method + connection module probability map, and (J) the proposed method + connection module segmentation result.
module as a way to learn more effective features in managing false positive detections. However, the false positive rate of the proposed method is still significant due to the high complexity of cyst appearance and the variety of quality in imaging devices. It is expected that employing more input channels enriched by the relation between abnormalities and retinal layers alongside more efficient training data can lead to a decrease in this rate.

Our current approach has two limitations that can be explained as follows:

1. The inability of involving B-scans before and after the B-scan is being segmented:
Since cysts have a three-dimensional (3D) entity that appears in a series of consecutive B-scans in OCT images, it is expected that employing information from B-scans before and after the B-scan is being segmented can be useful in more accurate segmentation. This 3D analysis needs a sufficient number of B-scans in each OCT volume so that the information dependency between B-scans is kept. Currently, there is no public dataset for fluid segmentation with this property. Hence, unfortunately, we could not use this information in our segmentation model.

2. Needing to change the reference dimensions and retrain the U-Net if new OCT devices come to the market and provide larger B-scans:
In our proposed method, the reference size 640 x 1024 was computed based on the size of OCT volumes acquired by four well-known and top spectral SD-OCT devices from four different vendors including Cirrus HD-OCT (Zeiss Meditec), Spectralis (Heidelberg Engineering), Nidek (Nidek), and T-1000/T-2000 (Topcon). These devices are common ones that are widely used in hospitals and ophthalmology clinics (practical environment). However, if a new SD-OCT device (new vendor) with an OCT image size larger than this reference dimension comes to the market, it is needed to change the reference dimension to this new size and retrain the proposed model based on this new input image size.

5 CONCLUSION AND FUTURE WORKS

Here, we presented a vendor-independent cyst segmentation method based on the U-Net architecture that can operate on SD-OCT images obtained from various devices in their original size and provide acceptable results. Our solution summarizes in two parts: (1) prior information embedding and input data adjustment, and (2) the IRC segmentation model. In the first part, the asymmetrically resizing problem was solved by zero-padding; and then by adding an extra input channel with the information of target object location, both the model focuses are increased, and the impact of outliers on the model learning process is decreased. And in the second part, we applied a connection module comprised of AGs in the skip connection layers and atrous spatial pyramid pooling in the bottleneck layer to the core of the U-Net to transfer more useful and meaningful information from the encoder to the decoder part.

Two public datasets named OPTIMA and KERMANY with were used to evaluate the proposed method. Using the ground truths intersection, our method could outperform all of the previous vendor-independent IRC segmentation methods by a mean Dice value of 0.78 and 0.81 on OPTIMA and KERMANY datasets, respectively.

It is expected that injecting more powerful channels included with the domain expertise such as a weighted ROI mask or a map with more effective information about the correlation between lesions and retinal layer structures to the network can provide a more helpful introduction of the cystoid features to the network. It will be investigated in future work.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
Three public datasets OPTIMA, UMN, and KERMANY that support the finding of this study are, respectively, available at https://optima.meduniwien.ac.at/optima-segmentation-challenge-1/, https://people.ece.umn.edu/users/parhi/DATA/OCT/DME/, and https://data.mendeley.com/datasets/rscbjbr9sj/2, references [36, 19, 37].

ORCID
Razieh Ganjee https://orcid.org/0000-0003-1848-3832
Mohsen Ebrahim Moghaddam https://orcid.org/0000-0002-7391-508X

REFERENCES
1. Ganjee R, Moghaddam ME, Nourinia R. Automatic segmentation of abnormal capillary nonperfusion regions in optical coherence tomography angiography images using marker-controlled watershed algorithm. J Biomed Opt. 2018;23(9):1-16. doi:10.1117/1.jbo.23.9.096006
2. Nazir T, Irtaza A, Shabbir Z, Javed A, Akram U, Mahmood MT. Diabetic retinopathy detection through novel tetragonal local octa patterns and extreme learning machines. Artif Intell Med. 2019;99:101695. doi:10.1016/j.artmed.2019.07.003
3. Rotsos TG, Moschos MM. Cystoid macular edema. Clin Ophthalmol. 2008;2(4):919-930.
4. Wilkins GR, Houghton OM, Oldenburg AL. Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. *IEEE Trans Biomed Eng*. 2012;59(4):1109-1114. doi:10.1109/TBME.2012.2184759

5. Roychowdhury S, Koozekanani DD, Radwan S, Parhi KK. Automated localization of cysts in diabetic macular edema using optical coherence tomography images, in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 1426–1429. 2013. doi:10.1109/EMBC.2013.6609778

6. Wang J, Zhang M, Pechauer AD, et al. Automated volumetric segmentation of retinal fluid on optical coherence tomography. *Biomed Opt Express*. 2016;7(4):1577-1589. doi:10.1364/boe.7.001577

7. Quellec G, Lee K, Dolejsi M, Garvin MK, Abràmoff MD, Sonka M. Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in SD-OCT of the macula. *IEEE Trans Med Imaging*. 2010;29(6):1321-1330. doi:10.1109/TMI.2010.2047023

8. González A, Remeseiro B, Ortega M, Penedo MG, Charlón P. Automatic cyst detection in OCT retinal images combining region flooding and texture analysis. *Proceedings – IEEE Symposium on Computer-Based Medical Systems*. 2013;397-400. doi:10.1109/CSBMS.2013.6627825

9. Gopinath K, Sivaswamy J. Domain knowledge assisted cyst segmentation in OCT retinal images. *arXiv Prepr. arXiv:1612.02675*, 2016.

10. Chiu SJ, Allingham MJ, Mettu PS, Cousins SW, Izatt JA, Farsiu S. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed Opt Express*. 2015;6(4):1172-1194. doi:10.1364/boe.6.001172

11. Chen X, Niemeijer M, Zhang L, Lee K, Abramoff MD, Sonka M. Three-dimensional segmentation of fluid-associated abnormalities in retinal OCT: probability constrained graph-search-graph-cut. *IEEE Trans Med Imaging*. 2012;31(8):1521-1531. doi:10.1109/TMI.2012.2191302

12. Zhu W, Zhang L, Shi F, et al. Automated framework for intraretinal cystoid macular edema segmentation in three-dimensional optical coherence tomography images with macular hole. *J Biomed Opt*. 2017;22(7):076014. doi:10.1117/1.jbo.22.7.076014

13. de Sisternes L, Hong J, Leng T, Rubin DL. A machine learning approach for device-independent automated segmentation of retinal cysts in spectral domain optical coherence tomography images. Proceeding Optima Challenge-MICCAI; 2015.

14. Schlegl T, Waldstein SM, Vogl WD, Schmidt-Erfurth U, Langs G. Predicting semantic descriptions from medical images with convolutional neural networks, in Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9123, pp. 437–448. 2015. doi:10.1007/978-3-319-19992-4_34

15. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology*. 2018;125(4):549-558. doi:10.1016/j.jophtha.2017.10.031

16. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. 2017;8(7):3440-3448. doi:10.1364/boe.8.003440

17. Roy AG, Conjeti S, Karri SPK, et al. ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed Opt Express*. 2017;8(8):3627-3642. doi:10.1364/boe.8.003627

18. Venuhuizen CI, van Grinsven F, van Ginneken MJ, Hoyng B, Theelen CC, Sanchez T. Fully automated segmentation of intraretinal cysts in 3D optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2016;57(12):5949.

19. Rashno A, Koozekanani DD, Drayna PM, et al. Fully automated segmentation of fluid/cyst regions in optical coherence tomography images with diabetic macular edema using neurosophic sets and graph algorithms. *IEEE Trans Biomed Eng*. 2018;65(5):989-1001. doi:10.1109/TBME.2017.2734058

20. Girish GN, Kothari AR, Rajan J. Marker controlled watershed transform for intra-retinal cysts segmentation from optical coherence tomography B-scans. *Pattern Recognit Lett*. 2018;139:86-94. doi:10.1016/j.patrec.2017.12.019

21. Oguz I, Zhang L, Abràmoff MD, Sonka M. Optimal cyst segmentation from OCT images, Medical Imaging 2016: Image Processing. 2016, vol. 9784, 97841E. doi:10.1117/12.2217355

22. Venuhuizen FG, van Ginneken B, Liefers B, et al. Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography. *Biomed Opt Express*. 2018;9(4):1545-1569. doi:10.1364/boe.9.001545

23. Girish GN, Thakur B, Chowdhury SR, Kothari AR, Rajan J. Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model. *IEEE J Biomed Heal Informatics*. 2019;23(1):296-304. doi:10.1109/JBHI.2018.2810379

24. Gopinath K, Sivaswamy J. Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement. *IEEE J Biomed Heal Informatics*. 2019;23(1):273-282. doi:10.1109/JBHI.2018.2793554

25. Optima cyst segmentation challenge. 2015 [Online]. Available: https://optima.meduniewien.ac.at/research/challenges/

26. Esmaeili M, Dehnavi A, Rabbani H, Hajiзадeh F. 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*. 2016;6(3):166-171.

27. Lu D, Heisler M, Lee S, et al. Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network. *Med Image Anal*. 2019;54:100-110. doi:10.1016/j.media.2019.02.011

28. Ganjee R, Ebrahimi Moghaddam M, Nourinia R. An unsupervised hierarchical approach for automatic intra-retinal cyst segmentation in spectral-domain optical coherence tomography images. *Med Phys*. 2020;47(10):4872-4884. doi:10.1002/mp.14361

29. Bogunovic H, Venuhuizen F, Klimscha S, et al. RETOUCH: the retinal OCT fluid detection and segmentation benchmark and challenge. *IEEE Trans Med Imaging*. 2019;38(8):1858-1874. doi:10.1109/TMI.2019.2901398

30. Schlemper J, Oktay O, Schaap M, et al. Attention gated networks: learning to leverage salient regions in medical images. *Med Image Anal*. 2019;53:197-207. doi:10.1016/j.media.2019.01.012
31. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2018:11211;801-818.

32. Tomasi C, Manduchi R. Bilateral filtering for gray and color images. *Proceedings of the IEEE International Conference on Computer Vision,* pp. 839–846. 1998.

33. Chiu SJ, Li XT, Nicholas P, Toth CA, Izatt JA, Farsiu S. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt Express.* 2010;18(18):19413-19428. doi:10.1364/oe.18.019413

34. Dijkstra EW. A note on two problems in connexion with graphs. *Numer Math.* 1959;1(1):269-271. doi:10.1007/BF01386390

35. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation, Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics), vol. 9351, pp. 234–241. 2015. doi:10.1007/978-3-319-24574-4_28

36. Wu J, Philip AM, Podkowinski D, et al. Multivendor spectral-domain optical coherence tomography dataset, observer annotation performance evaluation, and standardized evaluation framework for Intraretinal cystoid fluid segmentation. *J Ophthalmol.* 2016;2016:1-8. doi:10.1155/2016/3898750

37. Kermany D, Zhang K, Goldbaum M. Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley Data.* 2018:2;651.

38. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. *Med Imaging, IEEE Trans.* 1994;13(4):716-724. doi:10.1109/42.363096

39. Kamnitsas K, Chen L-C, Ledig C, Rueckert D, Glocker B. Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI. *Proc MICCAI-ISLES 2015.* 2015;13:13-16.

40. El Jurdi R, Petitjean C, Honeine P, Abdallah F. Bb-unet: U-net with bounding box prior. *IEEE J Sel Top Signal Process.* 2020;14(6):1189-1198. doi:10.1109/JSTSP.2020.3001502

**How to cite this article:** Ganjee R, Ebrahimi Moghaddam M, Nourinia R. A generalizable approach based on the U-Net model for automatic intraretinal cyst segmentation in SD-OCT images. *Int J Imaging Syst Technol.* 2023;33(5):1647-1660. doi:10.1002/ima.22893