Heliyon 10 (2024) e33911

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

ScLNet: A cornea with scleral lens OCT layers segmentation dataset and new multi-task model

Yang Cao [a,1], Xiang le Yu [a,1], Han Yao [a], Yue Jin [a], Kuangqing Lin [a], Ce Shi [a], Hongling Cheng [a], Zhiyang Lin [a], Jun Jiang [a], Hebei Gao [a,b,**], Meixiao Shen [a,*]

[a] Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Eye Hospital and School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, 325000, China
[b] School of Artificial Intelligence, Wenzhou Polytechnic, Wenzhou, 325035, China

A B S T R A C T

Objective: To develop deep learning methods with high accuracy for segmenting irregular corneas and detecting the tear fluid reservoir (TFR) boundary under the scleral lens. Additionally, this study aims to provide a publicly available cornea with scleral lens OCT dataset, including manually labeled layer masks for training and validation of segmentation algorithms. This study introduces ScLNet, a dataset comprising cornea with Scleral Lens (ScL) optical coherence tomography (OCT) images with layer annotations, and a multi-task network designed to achieve rapid, accurate, automated segmentation of scleral lens with regular and irregular corneas.

Methods: We created a dataset comprising 31,360 OCT images with scleral lens annotations. The network architecture includes an encoder with multi-scale input and a context coding layer, along with two decoders for specific tasks. The primary task focuses on predicting ScL, TFR, and cornea regions, while the auxiliary task, aimed at predicting the boundaries of ScL, TFR, and cornea, enhances feature extraction for the main task. Segmentation results were compared with state-of-the-art methods and evaluated using Dice similarity coefficient (DSC), intersection over union (IoU), Matthews correlation coefficient (MCC), Precision, and Hausdorff distance (HD).

Results: ScLNet achieves 98.22 % DSC, 96.50 % IoU, 98.13 % MCC, 98.35 % Precision, and 3.6840 HD (in pixels) in segmenting ScL; 97.78 % DSC, 95.66 % IoU, 97.71 % MCC, 97.70 % Precision, and 3.7838 HD (in pixels) in segmenting TFR; and 99.22 % DSC, 98.45 % IoU, 99.15 % MCC, 99.14 % Precision, and 3.5355 HD (in pixels) in segmenting cornea. The layer interfaces recognized by ScLNet closely align with expert annotations, as evidenced by high IoU scores. Boundary metrics further confirm its effectiveness.

Conclusion: We constructed a dataset of corneal OCT images with ScL wearing, which includes regular and irregular cornea patients. The proposed ScLNet achieves high accuracy in extracting ScL, TFR, and corneal layer masks and boundaries from OCT images of the dataset.

1. Introduction

Numerous ocular conditions including severe keratoconus, post–LASIK ectasia, and severe dry eyes, can cause elevated irregular astigmatism, which result in poor visual acuity when corrected with spectacles and soft contact lenses. Modern Scl, which vault the cornea of the eye and have great benefit in correcting irregular astigmatism, have emerged as a new treatment option to restore vision for patients with irregular astigmatism suffering from complex ocular diseases [1–3]. In an effort to achieve a successful lens-to-corneal

* Corresponding author. School of Ophthalmology and Optometry, Wenzhou Medical University, 270 Xueyuan Road, Wenzhou, Zhejiang, 325027, China.
** Corresponding author. Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Eye Hospital and School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, 325000, China.
E-mail addresses: gaohebei@ojlab.ac.cn (H. Gao), smx77@sohu.com (M. Shen).
[1] Yang Cao and Xiang le Yu co-first authors and have contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e33911
Received 16 January 2024; Received in revised form 28 June 2024; Accepted 28 June 2024
Available online 29 June 2024

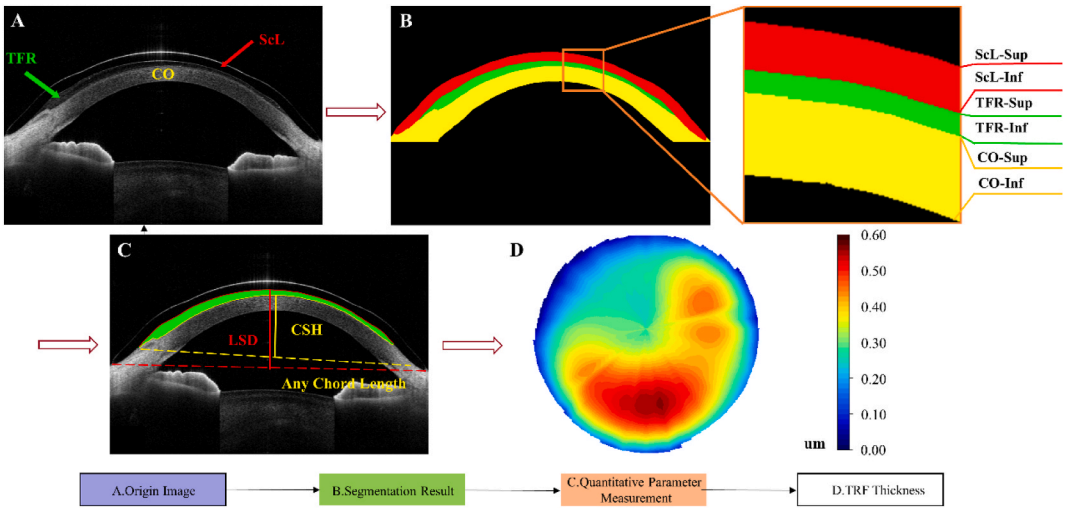

Fig. 1. The proposed pipeline: A. An example of corneal OCT image with ScL wearing. B. Segmentation Result. (red: ScL, green: TRF, yellow: cornea.) C. The quantitative parameters, including lens sagittal depth (LSD), corneal sagittal height (CSH) and thickness of the TRF etc., are obtained according to the segmentation results. D. According to the segmentation results, the real coordinates of the anterior and posterior surfaces of TRF in a scleral lens wearer were reconstructed in three dimensions to obtain the thickness of the TRF at a certain time. ScL-Sup: scleral lenses superior boundary. ScL-Inf: scleral lenses inferior boundary. TRF-Sup: tear film superior boundary. TRF-Inf: tear film inferior boundary. CO-Sup: cornea superior boundary. CO-Inf: cornea inferior boundary.

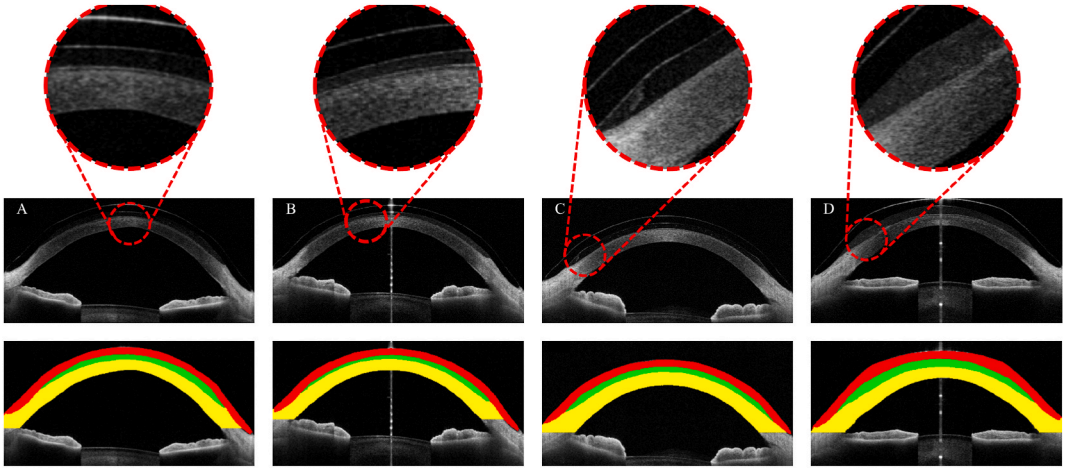

Fig. 2. Typical OCT images with the TFR under the scleral lens and theirs ground truth images with manually labeled. (A) Typical OCT image showing a keratoconus eye, (B) Typical OCT image showing a thin TFR, (C) Typical OCT image showing the asymmetric and uneven distribution of the TFR in a keratoconus eye with an irregular corneal surface, (D) Typical OCTimage showing midday fogging in the TFR.

fitting relationship, a variety of quantitative parameters, including the shape of ocular surface, the thickness mappings of the tear fluid reservoir (TFR) under the scleral lens and assessment of the whole cornea thickness during lens wear have been widely studied, employing different devices and methods [4].

Optical coherence tomography (OCT) is the standard-of-care imaging modality in ophthalmology. OCT is used to examine the whole cornea and visualize the lens-to-corneal fitting relationship [5]. Fig. 1(A-B) shows the typical example of a corneal OCT image with a ScL, as well as the main steps to extract the quantitative parameters for evaluating the lens fitting. As depicted in Fig. 1(C-D), the evaluation of quantitative parameters for lens-to-corneal fitting in OCT images necessitates the simultaneous segmentation of the interfaces of the scleral lens, TRF, and corneal layer. However, manual segmentation is subjective and time-consuming, thus calling for the development of automated segmentation techniques. Although several conventional segmentation algorithms based on thresholding and edge detection can yield reliable outcomes, they often encounter challenges posed by various artifacts present in the OCT images, such as saturation artifact, speckle noise, and regions characterized by a low signal-to-noise ratio (SNR).

Recently, a lot of deep learning (DL) networks have been proposed for the segmentation of the retinal and the corneal layers in OCT image [6–16]. For example, CorNet, a CNN-based segmentation model was proposed by Mathai et al. for detecting corneal tissue

interfaces in different OCT images. It utilized dense connections to establish connections among each layer via feature map concatenation, then generated the corneal interface segmentation in the final image [8]. Many researchers have focused on the end-to-end CNN framework based on U-NET [6,8,9,11–16]for corneal and retinal layer segmentation. Nevertheless, these methods have two major drawbacks. First, U-Net's architecture, while achieving improved performance in layer segmentation, often leads to inaccurate boundaries and unsatisfactory segmentations. This limitation is especially pronounced in OCT images with ScL on irregular corneas, given their unique morphological characteristics. Moreover, there's a limited amount of research and data available for training networks to accurately delineate OCT images with ScL, making it challenging to ensure the reliability of the results. At the time of writing, to our best knowledge, few study using a deep learning approach has been reported to automatically delineate OCT images with ScL on eyes due to the lack of dataset to train the network. Second, the multi-pathway outputs segmentation areas and region contours immediately for assessment of the whole cornea thickness. The dense connections in each pathway also significantly boost computational complexity. This results from every layer being interconnected, elevating the computation load in contrast to conventional architectures. Such intricacies lead to extended training durations, greater hardware demands, and potential overfitting risks, especially for smaller datasets. Additionally, extracting discernible features across multiple scale layers intensifies computation and memory use, inhibiting the network's scalability and its efficacy with larger images.

In addition, due to OCT images with ScL have complex characteristics in Fig. 2(A-D). First corneal morphology in patients with keratoconus may change significantly due to the presence of lesions. Second, the TFR presents with different morphological features with arbitrary shapes, sizes duo to special settling pattern of ScL. Morevoer, the pixel intensity of the TRF similar to that of the cornea results in blurred boundaries due to the presence of midday fogging that appears during lens wear. In our previous work [17], we proposed a method for segmenting the TRF from corneal OCT image with ScL. The design is targeted based on the characteristics of the TRF, but it can only be used to divide the TRF, without considering the lens above the TRF and the cornea below, which can not meet the corresponding clinical needs. Considering the process of fitting the lens, a comprehensive analysis of the position relationship between the lens and the ocular surface requires calculate and evaluate ScL fitting parameters such as TRF thickness distribution, corneal central thickness, corneal sagittal depth, Primary Functional Sagittal Depth of the lens (PFSD). For this reason, it is necessary to segment lens, TRF and cornea in OCT images with ScL and obtain their boundaries. The three regions of the OCT images with ScL are closely fitted, so it is very important to obtain contextual information. The boundary information of the three regions is also not utilized, which makes the powerful feature understanding of deep learning wasted.

Building on that foundation and the limitation mentioned above, we proposed ScLNet, a multi-pathway framework for the automatic delineation of OCT images with ScL on patients with regular and irregular corneas. It includes a primary task to predict Region of interest (ROI) regions and an auxiliary task to predict ROI boundaries. The auxiliary task is used to enhance the learning capability of the primary task. ScLNet is a convolutional neural network with the multi-scale input with context coding layer and a dual-pathway architecture for task-specific output. Our contributions can be summarized as follows:

- In order to verify current Scl segmentation methods, we constructed a dataset of corneal OCT images with ScL wearing. To our knowledge, it is the first dataset of corneal OCT images with ScL.
- We also designed a multi-task network with multi-scale input. The network used a context coding layer before the different scale images input layer in the encoder. It combined context information of different resolutions to encode the salient features of different layers. This allows the network to consider both global and local context, enhancing its understanding of the input data. And, for the semantic segmentation decoder side, to form precise target mask.
- We designed a boundary-guided construction combined by residual blocks and convolutional layers followed with a subtraction operation. It utilized important discriminative features to help the network capture fine-grained details necessary for accurate segmentation. This is the first network that can perform both semantic segmentation and edge detection on OCT images with ScL.

2. Related work

2.1. Multi-scale input

Creating multi-scale input is often achieved by generating an image pyramid [18,19], where an original image is processed at different resolutions or scales. These variously scaled images are then fed into the network either independently or in combination to ensure the model can capture features at multiple scales. Where features from different scales are commonly concatenated along the depth dimension before being passed to subsequent layers [20]. There may be challenges in aligning features from different scales properly, especially when the features are of different dimensions or aspects ratios [21]. Combining features from multiple scales can significantly increase the computational burden on the system. This might make it less suitable for real-time applications or environments with limited computational resources. Drozdzal et al. [22] employ a fully convolutional network for data normalization, and then iteratively use the preprocessed data for accurate segmentation through a fully convolutional residual network. This iterative segmentation approach falls short in various areas, including performance metrics, graphical processing memory requirements, and computational time.

2.2. Multi-pathway neural networks

In an image segmentation task, one pathway could focus on extracting low-level features such as edges and textures, while another could specialize in capturing high-level features like object categories. The information learned across tasks is leveraged to enhance the

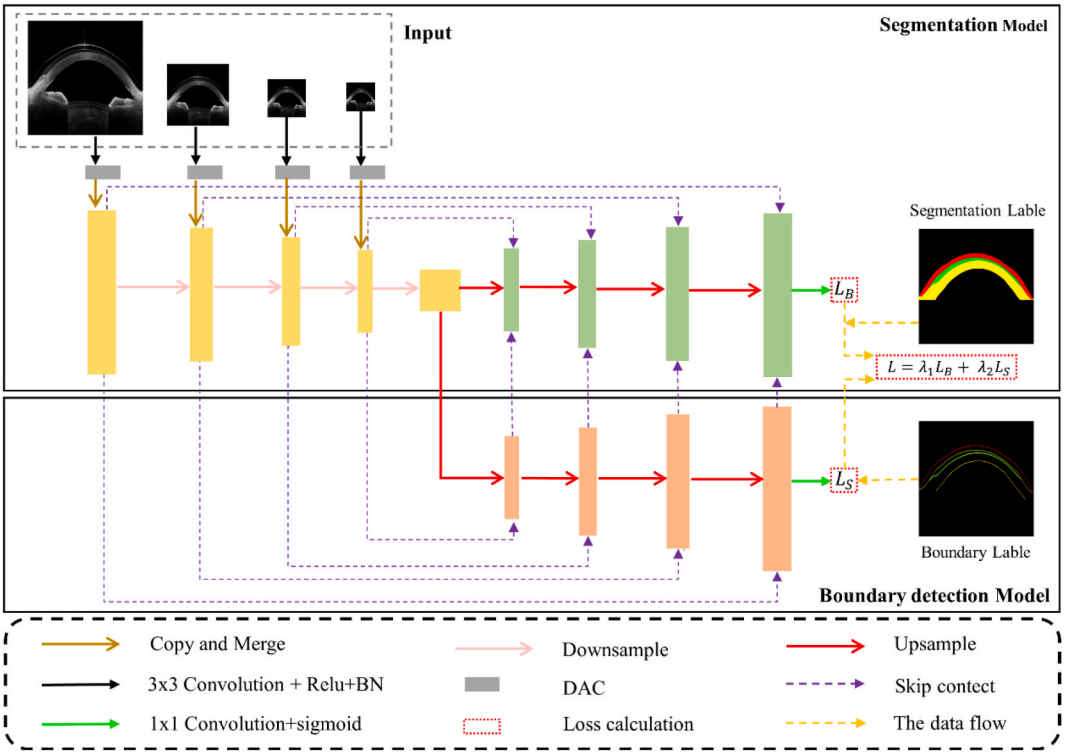

Fig. 3. Structure of the ScLNet framework. We first use the multi-scale input with context coding layer. It includes a primary task aimed at predicting scleral lenses, tear film, and cornea regions and an auxiliary task to predict their boundaries. The auxiliary task is used to enhance the learning capability of the primary task.

model's performance. Zeng et al. [23] introduces a dual-tracer encoder-decoder network SSIM, built on multi-task learning principles, designed to directly reconstruct two single-tracer dynamic images from a dual-tracer dynamic sinogram. However, in some regions of the images, the activities reconstructed by the SSIM deviated from the ground truth, and the images produced by the method generated by the method appeared overly smoothed. Wachinger et al. [24]. introduced DeepNAT, a multi-task learning framework for brain segmentation, which simultaneously learns a generalized feature representation and performs multi-class classification. While it is difficult to obtain enough precise segmentation for young, old, or diseased brains.

2.3. Atrous convolution

Atrous convolution [25], also known as dilated convolution, is a variation of the standard convolution operation used commonly in image processing and deep learning tasks. The key difference is the introduction of a dilation rate that adjusts the spacing between the values in the kernel. Essentially, this dilation rate creates "holes" in the kernel that are filled with zeros, thus making the kernel larger without adding extra parameters or computational complexity. DeepLabV2 [21] sets itself apart by introducing the Atrous Spatial Pyramid Pooling (ASPP) module. This module amalgamates feature maps created through the convolution of atrous layers with varying dilation rates. By doing so, it captures richer contextual information, thereby enhancing segmentation accuracy. HRNet [26] employs parallel stacking of atrous convolutions in its multi-scale branch to learn high-resolution representations. While, HRNet maintains high-resolution feature maps throughout the network, it increases the computational complexity. Furthermore, the number of parameters can be quite large for the multiple branches working in parallel across different resolutions. It makes the model heavier and potentially more prone to overfitting in smaller datasets.

3. Methodology

Problem Statement. Given a corneal OCT image $\mathscr{I}$, the task is to find a function $\mathscr{F} : \mathscr{I} \rightarrow \mathscr{L}$ that maps every pixel in I to a label $\mathscr{L} \in \{0,1,2,3\}$.The corneal interfaces to be segmented are: (1) Scleral lens, (2) Tear fluid reservoir, and (3) Cornea, with 0 being the background.

3.1. Network architecture

As shown in Fig. 3, our proposed ScLNet consists of one encoder module and two decoder modules with similar structures. The

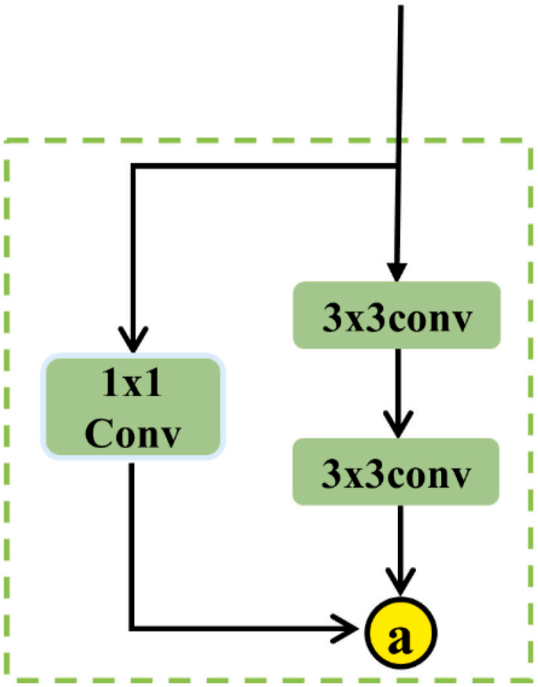

Fig. 4. Detail of resblock.

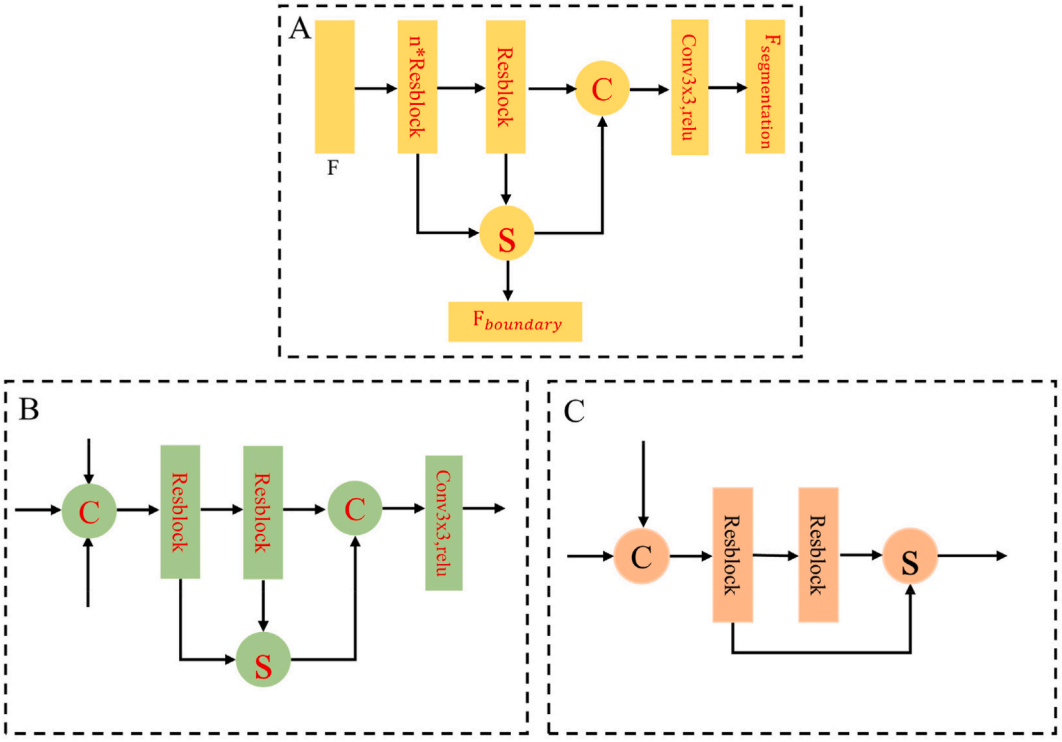

Fig. 5. The introduced encoding (A) and Segmentation decoding (B) in the developed ScLNet architecture, both of which included n resblock and a convolutional layer, the element-wise subtraction, and concatenation operations to obtain the image features. The developed Boundary decoding (C) in the boundary module.

encoder layer is comprised of convolutional blocks, a modified combination of resblocks (as seen in Fig. 4), and convolutional blocks inspired by ResNet, which serve as the backbone. The features obtained from this encoder are then summed with the backbone and passed on to the decoders. Each decoder module is designated for a specific task: one for semantic segmentation and the other for edge detection. The semantic segmentation task aims to classify each pixel as either belonging to the object of interest or the background, while the edge detection task guides semantic segmentation by initially identifying object boundaries and subsequently outputting them. The network loss is calculated as the weighted sum of the losses incurred in these two tasks.

A Backbone Encoder-Decoder

Encoder architecture is designed to start with a modified combination of resblock and convolutional blocks inspired by Res34Net [27]. Feature maps between resblocks and convolutional blocks are processed using element-wise subtraction and concatenation operations, along with the data flow, to capture the edge and object convolutional features, respectively. Each layer of the encoder adopts the subtraction operation of adjacent feature maps, which can detect a large number of feature information closely related to the target boundary, and the obtained information difference at the corresponding position can be used to guide the segmentation of the target region. These information differences effectively improve the detection sensitivity of the segmentation network to the target boundary and improve the segmentation accuracy of the boundary. Each layer contains varying numbers of resblocks, which helps to mitigate the problem of vanishing gradients. All layers maintain consistent convolutional kernels size and strides. As a result, the encoder's backbone retains five sets of convolution and varying numbers of residual convolution blocks for primary feature extraction as shown in Fig. 5(A). The object feature maps that were captured were down-sampled by using a 3×3 convolutional kernel with stride 2 operations to reduce the amount of computation.

After performing the encoding and down-sampling operations on the input images, the resulting region and boundary feature maps with varying dimensions and hierarchies. These feature maps were then used to perform region semantic segmentation and boundary delineation, respectively. The semantic feature maps were further processed using a 2×2 UpSampling layer and fed into a segmentation decoding block. A total of three features were utilized in the segmentation decoding block: the encoder's upsampling and skip connection, as well as a feature from the skip connection of the boundary decoding block. As shown in Fig. 5(B), this combination of feature maps was used to generate the final decoding features, which were then fed into a 1×1 convolution layer with a sigmoid activation function. This resulted in the output of a probability map for the segmented objects present in the entire image region.

To better identify object boundaries, the processed boundary feature maps underwent a 2×2 upsampling layer. This layer effectively increased the resolution of the maps by doubling their dimensions, there by enhancing the precision and accuracy of the boundary detection. They were then fed into a boundary decoding block, as well as another features that skip connection operations. These operations are illustrated in Fig. 5 (C). The process was repeated several times to effectively locate the object boundaries. Finally, the Conv 1×1 layer was utilized to apply a sigmoid activation function to the output, resulting in precise boundary detection.

B multi-scale input with context coding layer

Generally speaking, the deeper the network, the richer the nonlinear feature extraction capability of the network. However, with the deepening of the network, part of the spatial details lost in the process of multiple downsampling cannot be recovered through the oversampling, thus reducing the ability of the network to segment the edge of large targets and identify small targets. To solve this problem, the network introduces multi-scale inputs to increase the low-level detail features in the high-level semantic features. This method facilitates the network to learn small regional features of interest and local spatial features.

Input layer. The input layer consist of multiple images laid out horizontally in Fig. 3. Each image can be represented in an S-scale. The horizontal layout of the structure can be interpreted as a compact convolutional network resembling "S-layers" (S = 4 in Fig. 3). The image is first reduced to half, quarter, and eighth of the original image by means of the average pooled subsampling method, The output feature maps at layer l and scale s are denoted as f_l^s, while the original input image is represented as f_0^s.

By doing so, we obtain multi-scale input images that contain rich spatial information at multiple scales. and the ScL-Net can then use these multi-scale images to further refine its feature maps and classify complex patterns with greater accuracy. Then, the number of channels is expanded to the number of channels at the corresponding layer scale of different layers of the encoder through the convolution block. Finally, they are spliced together separately by the DAC module in the encoder corresponding to the scale.

In traditional convolution operations, the receptive field of the convolution kernel in the operation is fixed size, so it may be difficult to capture context information over a large range CE-Net [28] proposes a DAC block, while DAC can increase the effective receptive field of the convolution kernel by introducing different atrous rates into the convolution kernel in, and the output of each convolution kernel will produce a feature map as it uses. The DAC block is especially effective in capturing rich semantic features due to its use of four cascaded branches with varying receptive fields. This design enables the encoder to grasp the different levels of abstraction in the input data, thereby yielding more accurate and relevant representations.

$$f_l = H_{K,d_l}\left(\left[f_{l-1}, f_{l-2}, \cdots, f_0\right]\right) \tag{1}$$

d_l represents the dilation rate of the first layer. [.] denotes the concatenation operation. H_{K,d_l} represent the atrous convolution, [f, f] signifies that the feature map is obtained by concatenating the output of the upper layer through a concatenation operation, as shown in Fig. 6.

For tasks involving large objects or a global context, convolutions with larger receptive fields are advantageous. Conversely,

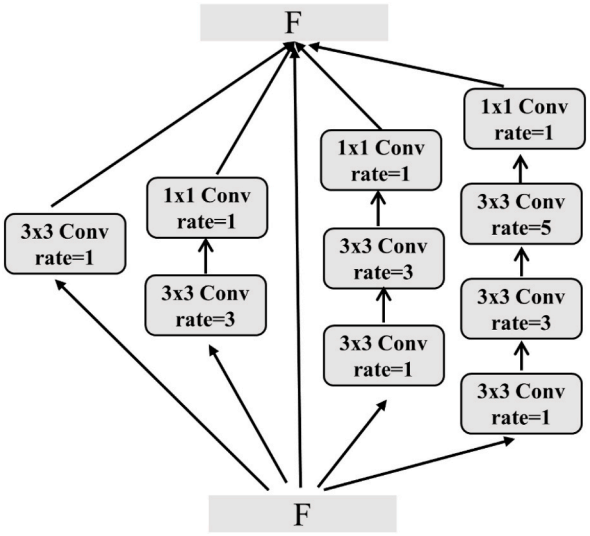

Fig. 6. The dense atrous convolution (DAC).

convolutions with smaller receptive fields are better suited for smaller objects or finer details [29]. In our specific task, TFR displays a broad spectrum of morphological features, including varying shapes and sizes that are inconsistent across different samples. This variability presents a challenge for accurate feature extraction and segmentation. The DAC block employs atrous convolution with different atrous rates to adeptly extract features from TFRs of diverse sizes, thereby enhancing the model's pixel-wise classification accuracy.

3.2. Loss function

Segmenting small objects has long been a challenging task in the field of semantic segmentation. A major obstacle in this regard is the notable disparity between positive and negative samples within the distribution. This imbalance detrimentally affects the performance of region-of-interest segmentation, as pixel-level labeling is a fundamental aspect of image semantic segmentation [30,31]. Hence, it is imperative to address the aforementioned concern in order to enhance the precision and efficacy of segmenting small objects. In our case, the background area occupies a large proportion of the entire image, while the region of interest occupies a small proportion of the entire image. This class imbalance problem causes the network to perform well in background area detection but poor in the analysis of Scl, TRF and cornea. Specifically, their boundary only takes about 1/100 of the whole OCT image area. Our network consists of two tasks, we used a hybrid loss consisting of contributions from two logarithm dice loss, which used as the loss function for the target objects and boundaries. The loss function $LOSS_{total}$ is defined by:

$$LOSS_{total} = \lambda\, LOSS_{Boundary} + (1-\lambda)\, LOSS_{Segmentation} \tag{2}$$

where λ is a non-negative weighting factor. In this study, we set $\lambda = 0.4$.where $LOSS_{Segmentation}$ is the loss of the semantic segmentation task, $LOSS_{Boundary}$ is the boundary depict task. logarithm dice loss is defined by log ($£_{\mathrm{Dice}}$) with the dice loss $£_{\mathrm{Dice}}$ defined as:

$$£_{\mathrm{Dice}}(y, \hat{y}) = 1 - \frac{2\sum_{i=1}^{n} y_i \hat{y}_i}{\sum_{i=1}^{n} y_i^2 + \sum_{i+1}^{n} \hat{y}_i^2} \tag{3}$$

where $\hat{y}$ is the predicted value; y is the ground truth.

3.3. Evaluation metrics

We validated the performance of the ScLNet using the independent testing dataset. We used five metrics for quantitative evaluation, including equations (4)–(8). The agreement between the computerized results and the manual results was quantitatively assessed using the DSC, intersection over union (IoU), Matthews correlation coefficient (MCC), Precision, and Hausdorff distance (HD) [32]. The metrics can be calculated using the following equations:

$$\mathrm{DSC} = \frac{2TP}{2TP + \mathrm{FP} + FN} \tag{4}$$

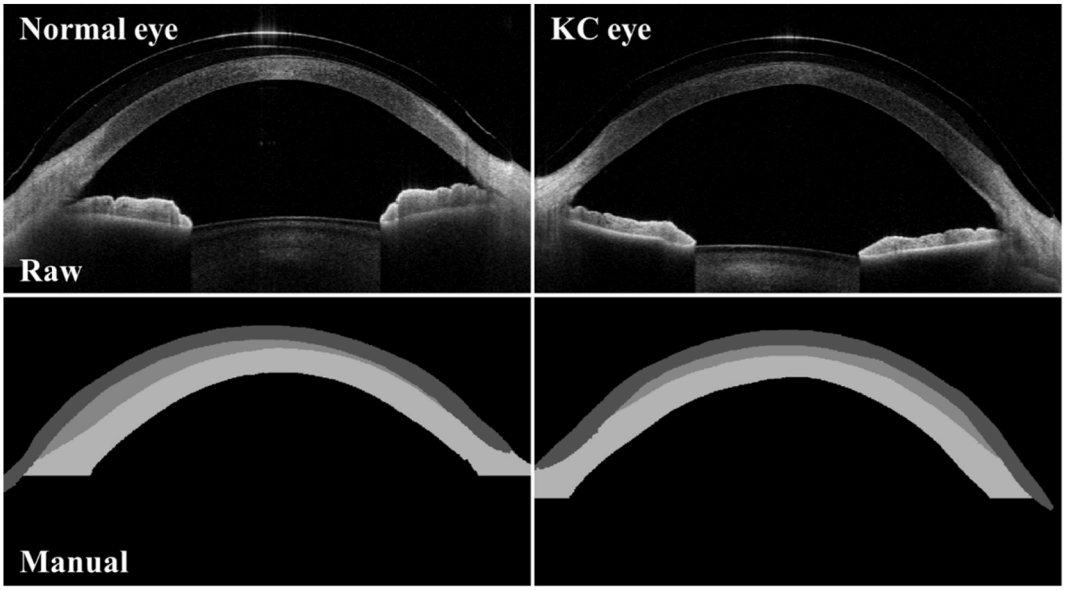

Fig. 7. The top row displays the OCT images from our collected dataset, while the bottom row shows their corresponding manual annotations.

$$IoU = \frac{Pix_{seg} \cap Pix_{gt}}{Pix_{seg} \cup Pix_{gt}} \tag{5}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6}$$

$$Precision = \frac{TP}{TP + FN} \tag{7}$$

$$HD = max\left(h(A,B), h(B,A)\right) \tag{8}$$

Where $h(A,B)$ and $h(B,A)$ are defined as

$$h(A,B) = \underset{a \ni A}{max} \left\{ \underset{b \ni B}{min} \| a - b \| \right\} \tag{9}$$

$$h(B,A) = \underset{b \ni B}{max} \left\{ \underset{a \ni A}{min} \| b - a \| \right\} \tag{10}$$

where Pix_{seg} is the pixel of predicted segmentation and Pix_{gt} is the ground truth. TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively. $h(A,B)$ is the directed Hausdorff distance from point set A to point set B. In addition to the HD, the larger the value, the better the performance.

3.4. The dataset for development and validation

We created to our knowledge the first of database corneal OCT images with ScL wearing. The dataset used in this study was obtained using a commercial SS-OCT (Tomey Casia 2, Tomey Corporation, Nagoya, Japan) system. The SS-OCT system operates at a central wavelength of 1310 nm, providing an axial resolution of 10 μm. The scanning speed of the system is 50,000 A-scans per second [33].

The dataset consisted of 52 healthy eyes with regular corneal surfaces and 46 eyes with keratoconus and irregular corneal surfaces. Imaging was performed immediately after lens wear and at multiple time points: 30 min, 1 h, 2 h, and 4 h. Each 3D-OCT dataset contained 18 B-scan images centered on the corneal apex. The size of each B-scan was 796 × 1160 pixels (height × width), corresponding to an actual dimension of 13 mm × 16 mm. Therefore, the dataset comprised a total of 31, 360 images.

To evaluate the performance of the network, a subset of 1678 images was randomly selected from all groups in the dataset. These images were further divided into three subsets: 70 % for training, 20 % for validation, and 10 % for testing purposes. In order to facilitate the network design, both the OCT images and corresponding masks were resized to a dimension of 512 × 512 pixels.

This study has obtained ethical approval from the ethics committee of Wenzhou Medical University (Ethics Approval Number: Y2018-071). All procedures follow the Helsinki Declaration. The study subjects were informed of the relevant precautions and signed

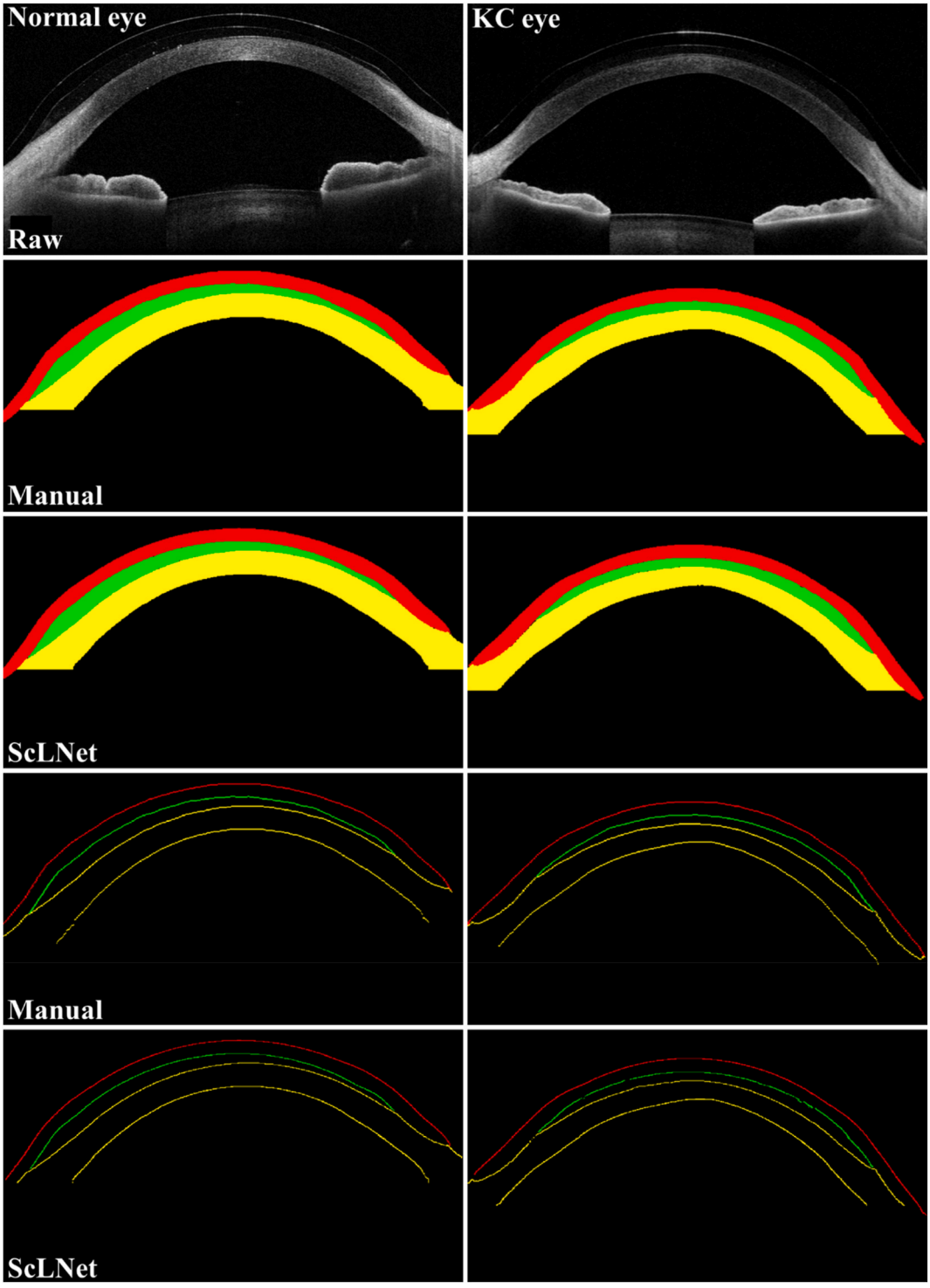

Fig. 8. Illustration of the performance of the developed ScLNet network on normal and KC eye OCT images. The first three rows are the OCT images and their segmentation results by manual annotations and the ScLNet, respectively. The last two rows are part of boundaries (the superior and inferior boundary of the ScL, the superior and inferior boundary of the Cornea) of different layers obtained by manual annotations and the ScLNet, respectively.

Table 1
Evaluation metrics for pathological images and normal images.

District	Method	DSC	IoU	MCC	HD	Precision
Scleral lens	Normal	0.9813	0.9633	0.9805	3.8282	0.9809
	Keratoconus	0.9825	0.9657	0.9817	3.6229	0.9846
	All	0.9822	0.9650	0.9813	3.6840	0.9835
tear fluid reservoir	Normal	0.9754	0.9521	0.9748	3.8052	0.9762
	Keratoconus	0.9788	0.9584	0.9781	3.7748	0.9774
	All	0.9778	0.9566	0.9771	3.7838	0.9770
cornea	Normal	0.9935	0.9871	0.9929	3.349	0.9934
	Keratoconus	0.9917	0.9835	0.9909	3.5864	0.9908
	All	0.9922	0.9845	0.9915	3.5355	0.9914

Table 2
The performance of the developed ScLNet and other five classical segmentation networks on the collected dataset.

District	Method	DSC	IoU	MCC	HD	Precision
Scleral lens	HRNet	0.9716	0.9447	0.9699	4.4126	0.9706
	Res34-Unet	0.9760	0.9542	0.9749	3.7338	0.9825
	CorneaNet	0.9668	0.9414	0.9651	3.9908	0.9733
	DeepLabV3+	0.9769	0.9549	0.9756	3.8337	0.9824
	BGCNN	0.9771	0.9567	0.9762	3.8129	0.9818
	ScLNet	**0.9822**	**0.9650**	**0.9813**	**3.6840**	**0.9835**
tear fluid reservoir	HRNet	0.9651	0.9327	0.9640	4.3215	0.9663
	Res34-Unet	0.9653	0.9381	0.9654	4.0147	0.9634
	CorneaNet	0.9606	0.9276	0.9601	4.1638	0.9690
	DeepLabV3+	0.9729	0.9472	0.9720	3.9643	**0.9797**
	BGCNN	0.9725	0.9473	0.9718	3.8915	0.9764
	ScLNet	**0.9778**	**0.9566**	**0.9771**	**3.7838**	0.9770
cornea	HRNet	0.9851	0.9707	0.9838	4.0835	0.9868
	Res34-Unet	0.9885	0.9773	0.9875	3.7766	0.9881
	CorneaNet	0.9817	0.9664	0.9807	3.9991	0.9730
	DeepLabV3+	0.9873	0.9749	0.9861	3.8984	0.9861
	BGCNN	0.9843	0.9726	0.9837	3.7282	0.9781
	ScLNet	**0.9922**	**0.9845**	**0.9915**	**3.5355**	**0.9914**

an informed consent form. The optimization of $Loss_{total}$ was performed using the stochastic gradient descent (SGD) algorithm with an initial learning rate (LR) set to 0.1. To ensure effective learning, the LR was reduced by a factor of ten every 20 epochs. The network training spanned 100 epochs, and a batch size of 2 was chosen to accommodate the GPU's capabilities. To prevent unnecessary optimization, an early stopping scheme was implemented, halting the process if the loss failed to decrease for a continuous period of 50 epochs.

To enhance the robustness of the trained models, image augmentation techniques were employed during training. These operations encompassed random vertical flips, translations ranging from −10 to 10 percent per axis, rotations within the range of −10 to 10°, and scaling between 0.9 and 1.1. By incorporating these augmentation strategies, the models became more resilient and adaptable to variations in the input data.

All the networks were implemented using python 3.6 and keras library version 2.4 with the Tensorflow platform. The experiments were performed on the Windows 10 operating system using CPU Intel(R) Core(TM) i9-10940K 3.3 GHz, GPU NVIDIA GeForce GTX TITAN, and a RAM of 128 GB.

To segment the Scl, TFR, and cornea, as well as to generate labeled images, we employed an algorithm developed in Matlab R2019 (The MathWorks Inc., Natick, MA, USA). This algorithm is based on the graph theory approach we have previously described [34] to semi-aumatically segment the boundaries of TRF. After that, two experienced graders (Dr.S Ce & Dr. HL Cheng) checked the boundaries of the segmented TFR under the scleral lens for each image and corrected it manually if necessary. At last, a labeled normal and KC eye image containing 4 classes, including ScL, TRF, cornea and background, was generated as shown in Fig. 7.

3.5. Statistical analysis

SPSS 25.0 was used for all statistical analyses. Quantitative data conforming to a normal distribution are expressed as mean and standard deviation. The T test was used for statistical analysis between the normal cornea and irregular cornea groups, and one-way analysis of variance was used for statistical analysis of 2 groups. $P < 0.05$ indicated statistical significance.

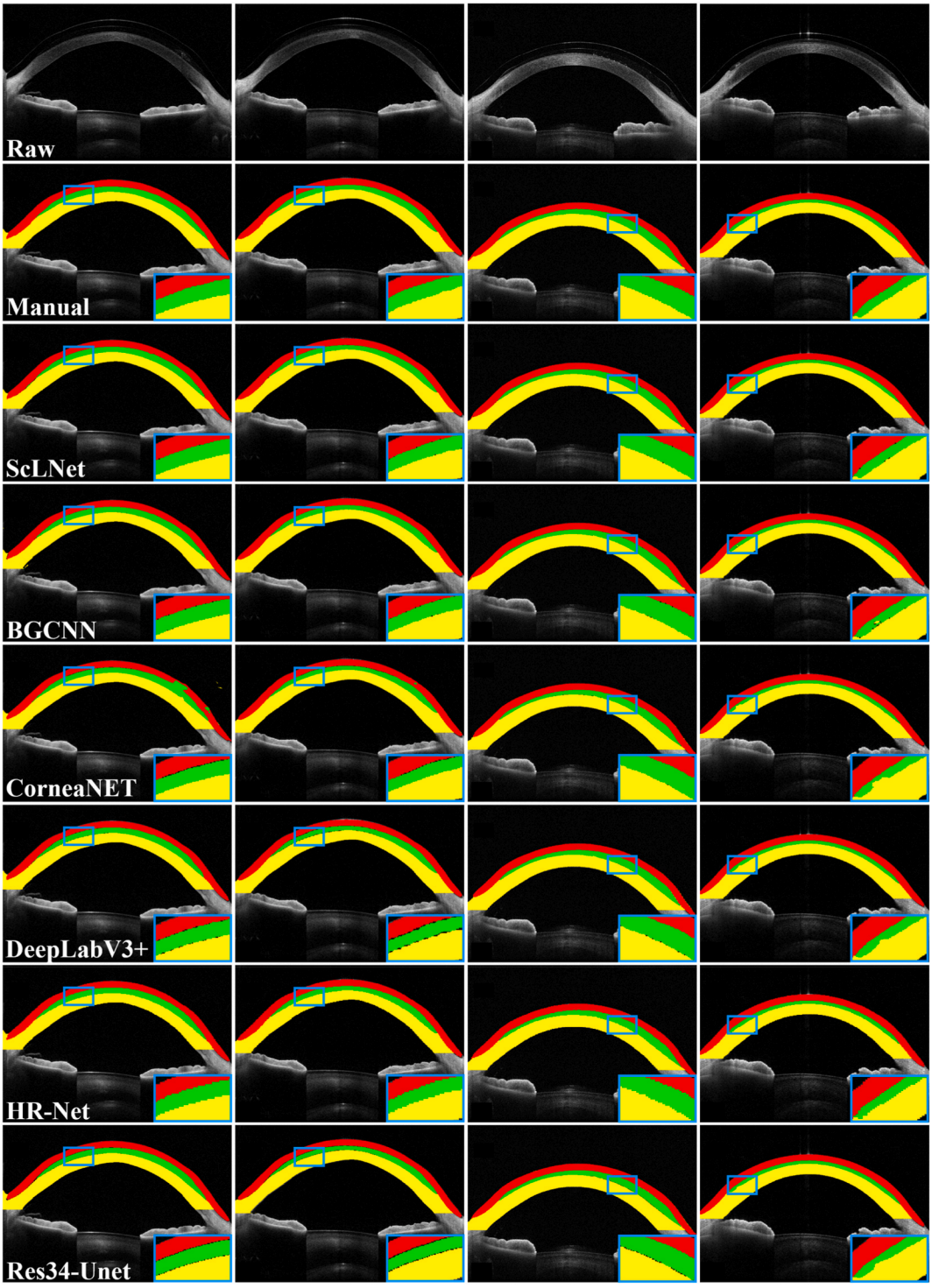

Fig. 9. Variability in the performance of three segmentation networks: ScLNet, CorneaNET, and DeepLabV3+, in identifying the ScL, TRF, and cornea layers is demonstrated using six representative OCT images. The first row presents the original OCT images. The second row shows manual annotations of the OCT images, carried out using a Matlab algorithm. Subsequent rows display segmentation results as achieved by ScLNet, CorneaNET, DeepLabV3+, HR-Net, and Res34-Unet, in that order.

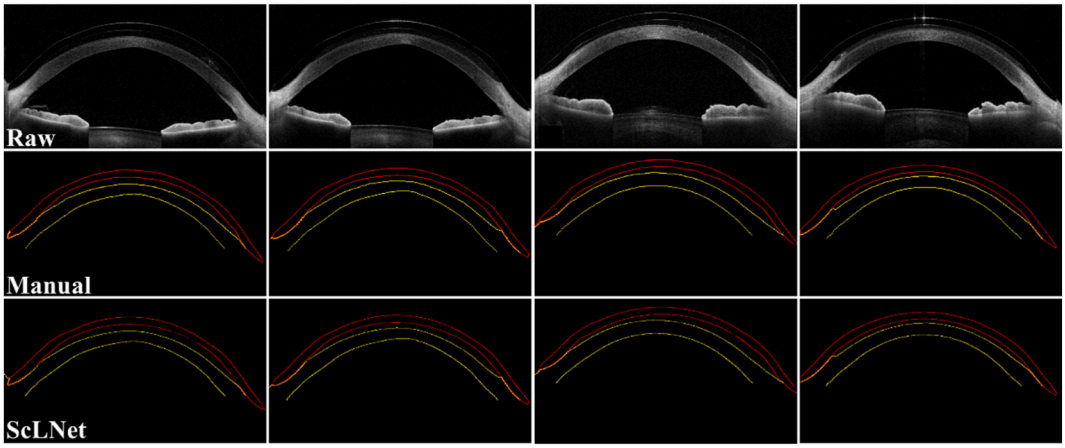

Fig. 10. The section illustrates the boundaries of the ScL, TRF, and cornea layers as identified by the ScLNet network for the OCT images displayed in Fig. 9. The first two rows show the original OCT images along with partial manual annotations of their boundaries. The final row presents the corresponding partial results generated by the ScLNet network.

4. Results

4.1. Overall performance

Our proposed ScLNet has two outputs: pixel-wise segmentation map and boundary. Fig. 8 visually displayed three tissue layers and four boundaries were automatically segmented using the performance of the ScLNet on normal and KC eye OCT images. These values suggest that the developed network exhibits promising potential in extracting the TFR under the scleral lens from OCT images, compared to the manual annotations. As the TRF lies adjacent to the ScL and cornea, we present four boundaries: the both superior and inferior boundary of the ScL, the both superior and inferior boundary of the Cornea. The segmentation of layers in pathological images poses challenges due to substantial changes in corneal morphology caused by conditions like keratoconus or other lesions. To evaluate the performance of ScLNet on pathological cases, we partition the testing set into two groups: normal and pathological, based on expert diagnoses. Table 1 illustrates the comparison of performance between keratoconus (118 images) and normal (50 images) on independent datasets. Although keratoconus cases typically exhibit irregular shapes and obscure boundaries, the evaluation metrics for pathological images indicate only a tiny percentage gap lower than the normal cornea. Furthermore, the evaluation metrics for pathological images are slightly higher than those of normal images for ScL and TRF regions. This comparison demonstrates the high robustness of our ScLNet in segmenting ScL wear in pathological images.

4.2. Performance comparison

Extensive comparative studies have shown the superior performances of ScLNet in depicting the enclosed TFR under the scleral lens in OCT images compared to some models of segmented corneas and other state-of-the-art segmentation models, including CorneaNet, BG-CNN, HR-Net, Res34-Unet [35], Deeplabv3+ [36]. The experimental results are shown in Table 2. We trained all the compared methods using the same training and testing strategy. The ScLNet achieves 98.22 % DSC, 96.50 % IoU,98.13 % MCC, and 98.35 % Precision and 3.6840 HD (in pixel) in segmentation ScL, 97.78 % DSC,95.66%IoU,97.71%MCC and 97.70 % Precision and 3.7838 HD (in pixel) in segmentation TRF,99.22 % DSC,98.45%IoU,99.15%MCC, and 99.14 % Precision and 3.5355HD (in pixel) in segmenting cornea. Fig. 9 shows the computerized results from the developed algorithm, CorneaNet, DeepLabV3+, HR-Net and Res34-Unet in identifying the ScL, TRF, cornea depicted in the OCT images. Fig. 10 shows the boundaries of the ScL, TRF, and cornea layers as identified by the ScLNet network for the OCT images displayed in Fig. 9.

4.3. Ablation study

We perform ablation studies to analyze the effects of removing certain design aspects in our proposed model, while keeping the hyperparameters constant. The whole framework includes an encoder block, semantic segmentation decoder and boundary detection decoder, we added each module separately to the backbone to analyze its effectiveness by conducting experiments on six different architectures, namely backbone, backbone + multi-scale, backbone + context, ScLNet, ScLNet w/o atrous and ScLNet w/o branch task. The qualitative and quantitative evaluation results of the ablation study are shown in Fig. 11 and Table 3, respectively.

Only the first layer of the backbone has an image input. The multi-scale input module is added to the backbone to enrich the spatial information. After adding this module, the first four coding layers of the backbone use image input of different scales. Segmentation results are improved by the Baseline + multi-scale input which increases the DSC from 97.28 % to 98.05 % in the segmentation Scleral lens, from 96.63 % to 97.59 % in segmentation TFR, from 98.39 % to 99.18 % in segmentation cornea. Experiments show that the key

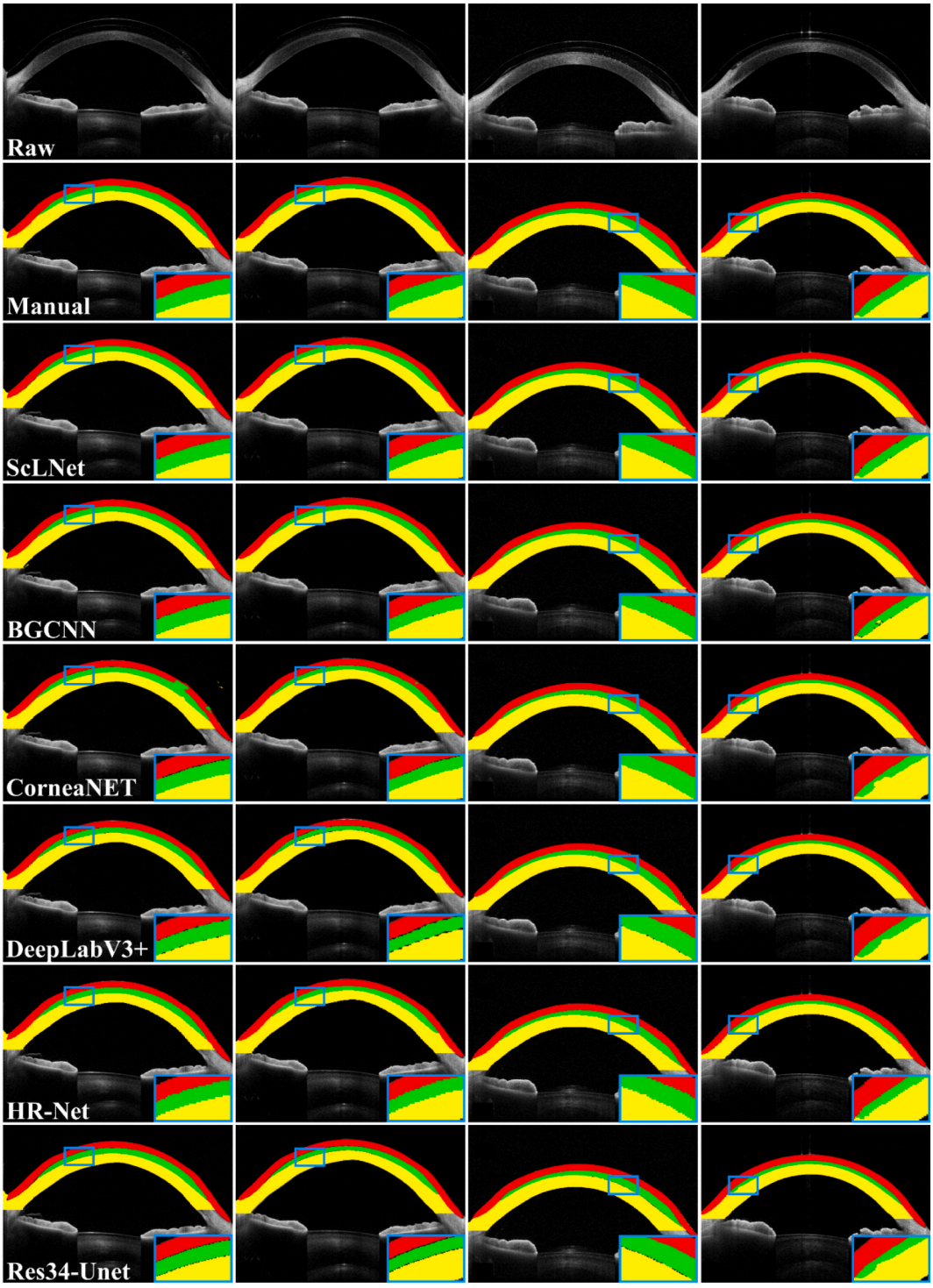

Fig. 11. Visualizes the segmentation results of different ablation models. We can see that our method achieves the smoother result, and the visualization result of our model is better than other ablation models.

modules of the ScLNet are effective.

The DAC block is added after the image has undergone a 3×3 convolutional operation. After adding this module. Segmentation results are improved by the Baseline + DAC. The DAC block increases the DSC from 97.28 % to 97.87 % in segmentation Scleral lens, from 96.63 % to 97.55 % in segmentation TRF, from 98.39 % to 99.18 % in segmentation cornea. Meanwhile, The used DAC block

Table 3
Performance of each ablation model compared to manual delineation on the collected dataset.

District	Method	DSC	IoU	MCC	HD	Precision
Scleral lens	Backbone	0.9728	0.9525	0.9724	3.8630	0.9770
	Backbone + multi-scale	0.9805	0.9620	0.9797	3.7090	0.9830
	Backbone + DAC	0.9787	0.9589	0.9778	3.8397	0.9799
	ScLNet	**0.9822**	**0.9650**	**0.9813**	**3.6840**	**0.9835**
	ScLNet w/o atrous	0.9820	0.9648	0.9812	3.7380	0.9828
	ScLNet w/o branch task	0.9790	0.9590	0.9781	3.7748	0.9834
tear fluid reservoir	Backbone	0.9663	0.9413	0.9668	4.0571	0.9630
	Backbone + multi-scale	0.9759	0.9530	0.9752	3.8246	0.9769
	Backbone + DAC	0.9755	0.9525	0.9749	3.8116	0.9728
	ScLNet	**0.9778**	**0.9566**	**0.9771**	**3.7838**	**0.9770**
	ScLNet w/o atrous	0.9774	0.9558	0.9767	3.8829	0.9746
	ScLNet w/o branch task	0.9741	0.9496	0.9734	3.9034	0.9734
cornea	Backbone	0.9839	0.9726	0.9834	3.8095	0.9788
	Backbone + multi-scale	0.9918	0.9838	0.9911	3.5483	0.9905
	Backbone + DAC	0.9918	0.9837	0.9911	3.5610	0.9901
	ScLNet	**0.9922**	**0.9845**	**0.9915**	**3.5355**	0.9914
	ScLNet w/o atrous	0.9920	0.9842	0.9913	3.5923	**0.9915**
	ScLNet w/o branch task	0.9906	0.9815	0.9898	3.6083	0.9906

employs the atrous convolution with different rates. We use regular convolution to replace the atrous convolution in DAC block (referred to ScLNet without atrous). As shown in Table 3, the DSC is reduced to 98.20 % in the segmentation of Scleral lens, 97.74 % in the segmentation of TRF, 99.20 % segmentation of the cornea. This indicates that atrous convolution helps to extract high-level semantic features, compared to the regular convolution.

Then, we evaluated the performances based on multi-tasking. The auxiliary task is used to predict the edge of the object, which can improve the feature extraction capabilities of the primary task. We trained the ScLNet without an auxiliary task for comparison. As shown in Table 3, its DSC is reduced to 97.90 % in segmentation of the Scleral lens, 97.41 % in segmentation of TRF, 99.06 % in segmentation of the cornea. If the auxiliary task is deleted, the boundary between the cornea and the TRF layer, the boundary between the TRF layer and the ScL will be misclassified in Fig. 12. Therefore, all improvements enhanced the final segmentation performance effectively.

Finally, We investigate the effect of the weight λ on the performance of the ScLNet (Eq. (2)). The weight was set to 0.2,0.4, 0.6 and 0.8 separately for the same segmentation experiments. The experiments showed that theScLNet had the best performance when the weight was set to 0.4, as compared with those of the other values in image segmentation, as shown in Table 4.

5. Discussion

This paper introduces a novel network called ScLNet, which is designed for two tasks, the segmentation of ScL, TFR, and cornea layers in OCT images and the descript the ragion shape. To achieve the simultaneous generation of layer masks and boundaries, a multi-pathway architecture is proposed: one multi-scale input and two task special outputs. The experimental results of the ScLNet without the branch task demonstrate its effectiveness in accurately identifying the contours and boundaries of scleral lenses, tear film, and cornea layers in OCT images, thereby enabling the automatic generation of masks and boundaries with a high level of accuracy. In addition to generating masks and boundaries with a high level of precision, this study introduces a multi-scale layer that constructs an image pyramid of multi-scale inputs to retain spatial information in the feature maps extracted by the encoder. Ablation experiments demonstrate that removing the multi-scale inputs layer leads to misclassification of objects as background. Furthermore, the context coding layer, which comprises a DAC block combined with a residual block followed by a subtraction operation, serves as the fundamental building block of the model encoder. DAC employs hybrid atrous convolution to increase the receptive field and improve the quality of feature representation. The inclusion of residual blocks in the construction provides advantages such as improved gradient flow and easier optimization. This promotes better feature extraction and helps the network to learn more robust representations. The subtraction operation plays a role in emphasizing the boundary information. By subtracting relevant feature maps, the network can highlight the differences or changes in pixel values, which often correspond to layer boundaries [13,15]. A boundary decoding block is introduced to the model decoder, which helps refine the upsampled feature maps after convolution operations leading to even more precise segmentation. Experimental results indicate that the proposed network is capable of segmenting ScL, TFR, and cornea layers from OCT images and detected the boundaries automatically. The thickness of the TRF is very important when fitting the ScLfor patients with irregular corneas. When dividing the test set into normal and pathological, the evaluation metrics for pathological cases align more closely with manual annotation levels than normal cases in TRF. However, keratoconus cases, known for their irregular shapes and indistinct boundaries in images, exhibit a slight percentage gap in corneal thickness compared to normal cases. To enhance the algorithm's segmentation efficiency for corneas in pathological images, future studies will focus on optimizing the edge-aware loss and boundary refinement module.

When dealing with saturation artifacts, speckle noise, and regions with low Signal-to-Noise Ratio (SNR), our method achieves the best results compared with other corneal segmentation models (CorneaNet and BG-CNN)and State-of-the-Art (SOTA) models (HR-Net, Res34-Unet, DeepLabV3+) evaluated in our experiments.

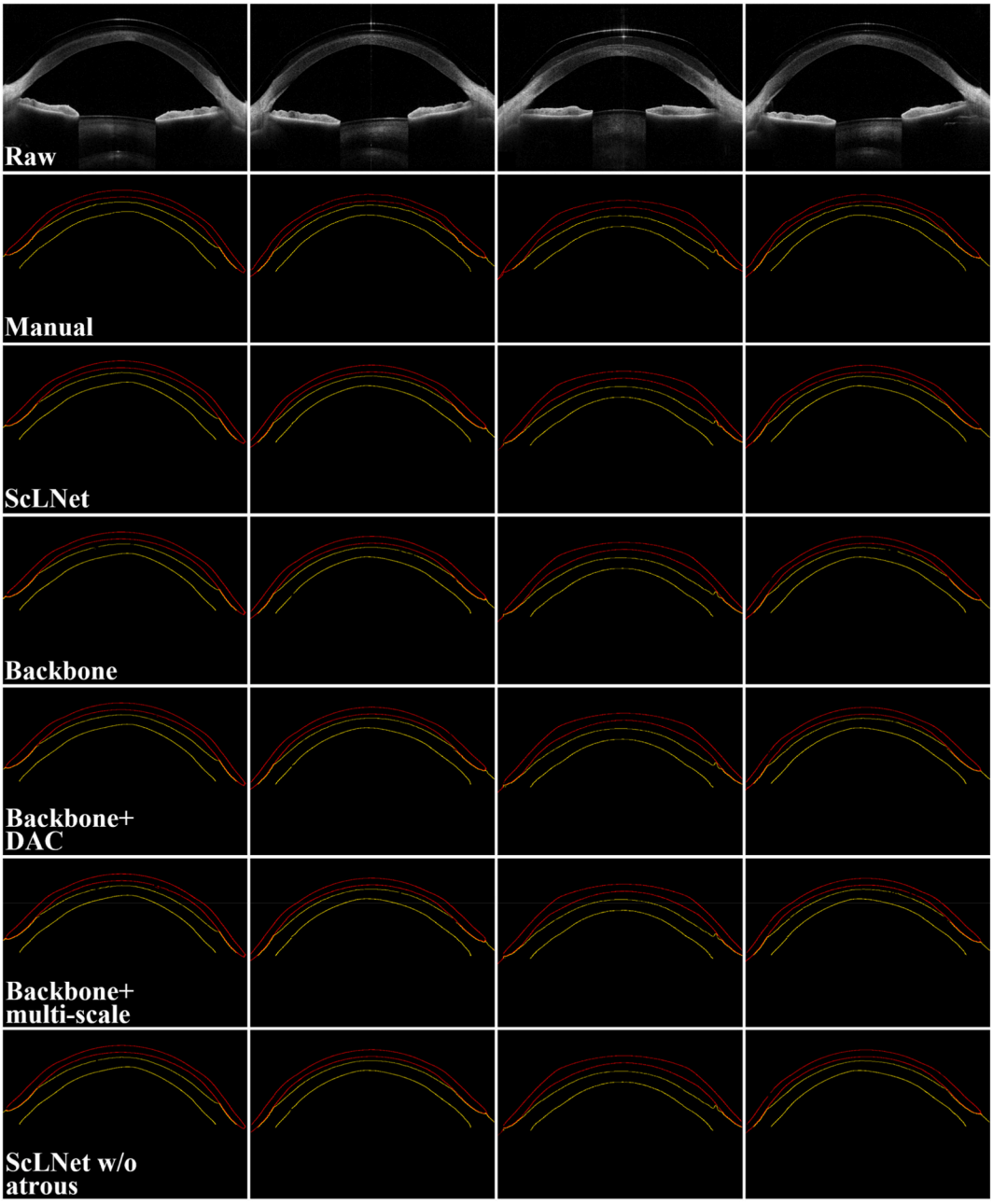

Fig. 12. Performance differences of our ablation models in identifying the tear liquid layers from four typical OCT images. The first row is the raw OCT images. The next six rows are manual and the segmentation results obtained by ScLNet, Backbone, Backbone + multi-scale, Backbone + DAC and ScLNet w/o atrous, respectively.

Table 4
HD of the developed network with the dice loss function and different λ in the hybrid loss functions.

District	**0.2**	**0.4**	**0.6**	**0.8**
Scleral lens	3.7185	**3.6840**	3.6914	3.7507
tear fluid reservoir	3.8255	**3.7838**	3.7939	3.8477
cornea	3.5753	**3.5355**	3.5420	3.5484

It should be noted that in our study, there is a limitation regarding OCT image data sourced from a single OCT machine. A key challenge for deep learning is domain shift, where models trained on one dataset may struggle when applied to slightly different datasets from various sources. The issue of generalizability is shown by a commercial AI device used to screen for diabetic retinopathy, which was only approved for use with one specific camera due to concerns about performance with other cameras [37,38]. Various methods have been suggested to tackle this problem, including transfer learning and style transfer. Transfer learning involves pre-training a network on a large general image database and fine-tuning it on a smaller, more specific dataset. Style transfer can be done directly on paired images using encoder and decoder networks or by transferring styles between images [39,40]. Recently, a foundational CV model for Segment Anything [41] has been released. These models are designed to segment or delineate various types of objects or regions within images, regardless of their complexity or diversity. In our upcoming study, we plan to leverage these foundational models to effectively adapt segmentation across domains, addressing the challenges.

6. Limitations

Our work is still limited to some degree, as only our dataset is performed in our experiments, which can limit the model's potential use in the multi-center application. Multi-center data segmentation is expected in the future to further evaluate the performance of the proposed model. The dataset was obtained using a single type of OCT device, the validation of which has not been extended to other OCT models, as previously mentioned. In addition, ScL can be used for refractive correction after corneal transplantation, severe dry eye, and neurotrophic keratitis, and more types of patients will be included in our dataset in the future. And sclera vasculature are an important feature for scleral recognition. Currently, our segmentation network is unable to segment sclera vasculature, but this will be considered in the future.

7. Conclusion

Here, we introduce a meticulously designed corneal segmentation dataset with scleral lenses in OCT images, representing a large and systematically annotated collection. To the best of our knowledge, this is the first OCT dataset released to the research community specifically for the task of TFR segmentation. These images were obtained from both regular and irregular cornea patients, divided into two subsets.

Furthermore, our proposed ScLNet demonstrates a high level of accuracy in simultaneously extracting ScL, TFR, and corneal layer masks and boundaries from OCT images. This capability is essential for generating quantitative parameters to evaluate the fitting of scleral lenses. In ScLNet, the utilization of DAC layers integrates images at different scales, enabling the network to capture content information from various layers at multiple resolutions. The incorporation of boundary-guided construction enhances the network's ability to learn high-frequency features. This high-frequency feature learning helps the network generalize well to different edges of the cornea and improves its performance in discontinuous boundaries.

We have demonstrated that multi-task learning with task-wise early stopping allows each model to be more robust to different irregular interfaces with severe, highly blurred margins compared to a single-task method.

Data availability

The online webserver is available at https://82q411i088.goho.co/, while the codes and Dataset of ScLNet have been made publicly available at https://github.com/smallJude/ScLNet.

CRediT authorship contribution statement

Yang Cao: Software, Resources, Formal analysis, Conceptualization. **Xiang le Yu:** Software, Methodology, Data curation. **Han Yao:** Visualization, Software, Funding acquisition. **Yue Jin:** Software, Data curation. **Kuangqing Lin:** Software, Data curation. **Ce Shi:** Software, Investigation, Data curation. **Hongling Cheng:** Investigation, Data curation. **Zhiyang Lin:** Investigation, Data curation. **Jun Jiang:** Data curation, Conceptualization. **Hebei Gao:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Meixiao Shen:** Writing – original draft, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grants 82171016.

References

[1] A.S. Parra, et al., Assessment of the Prosthetic Replacement of Ocular Surface Ecosystem (PROSE) scleral lens on visual acuity for corneal irregularity and ocular surface disease, Ocul. Surf. 16 (2) (2018) 254–258, https://doi.org/10.1016/j.jtos.2018.01.003.
[2] R. Laballe, et al., Preclinical assessment of scleral lens as a reservoir-based ocular therapeutic system, Contact Lens Anterior Eye 39 (5) (2016) 394–396, https://doi.org/10.1016/j.clae.2016.04.008.
[3] J.D. Ling, A. Gire, S.C. Pflugfelder, PROSE therapy used to minimize corneal trauma in patients with corneal epithelial defects, Am. J. Ophthalmol. 155 (4) (2013) 615–619.e2, https://doi.org/10.1016/j.ajo.2012.09.033.
[4] B. Tan, et al., Effects of scleral-lens tear clearance on corneal edema and post-lens tear dynamics: a pilot study, Optom. Vis. Sci. 95 (6) (2018) 481–490, https://doi.org/10.1097/opx.0000000000001220.
[5] S.J. Vincent, D. Alonso-Caneiro, M.J. Collins, Optical coherence tomography and scleral contact lenses: clinical and research applications, Clin. Exp. Optom. 102 (3) (2019) 224–241, https://doi.org/10.1111/cxo.12814.
[6] C. Shi, et al., A recurrent skip deep learning network for accurate image segmentation, Biomed. Signal Process Control 74 (2022), https://doi.org/10.1016/j.bspc.2022.103533.
[7] A. Elsawy, M. Abdel-Mottaleb, PIPE-Net: a pyramidal-input-parallel-encoding network for the segmentation of corneal layer interfaces in OCT images, Comput. Biol. Med. 147 (2022) 105595, https://doi.org/10.1016/j.compbiomed.2022.105595.
[8] T. Sudharshan Mathai, K. Lathrop, J. Galeotti, Learning to segment corneal tissue interfaces in OCT images. https://doi.org/10.48550/arXiv.1810.06612, 2018.
[9] Y. Liu, et al., Hierarchy-constrained network for corneal tissue segmentation based on anterior segment OCT images, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, https://doi.org/10.1109/ISBI45749.2020.9098648.
[10] L. Wang, et al., EE-Net: an edge-enhanced deep learning network for jointly identifying corneal micro-layers from optical coherence tomography, Biomed. Signal Process Control 71 (2022), https://doi.org/10.1016/j.bspc.2021.103213.
[11] Y.F. He, et al., Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT, Biomed. Opt Express 10 (10) (2019) 5042–5058, https://doi.org/10.1364/boe.10.005042.
[12] V.A. Dos Santos, et al., CorneaNet: fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning, Biomed. Opt Express 10 (2) (2019) 622–641, https://doi.org/10.1364/boe.10.000622.
[13] B. Wang, et al., Boundary aware U-net for retinal layers segmentation in optical coherence tomography images, Ieee Journal of Biomedical and Health Informatics 25 (8) (2021) 3029–3040, https://doi.org/10.1109/jbhi.2021.3066208.
[14] L.Y. Fang, et al., Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search, Biomed. Opt Express 8 (5) (2017) 2732–2744, https://doi.org/10.1364/boe.8.002732.
[15] L. Wang, et al., Automated delineation of corneal layers on OCT images using a boundary-guided CNN, Pattern Recogn. 120 (2021), https://doi.org/10.1016/j.patcog.2021.108158.
[16] J. Ouyang, et al., Accurate tissue interface segmentation via adversarial pre-segmentation of anterior segment OCT images, Biomed. Opt Express 10 (10) (2019) 5291–5324, https://doi.org/10.1364/BOE.10.005291.
[17] Y.H. Zhou, et al., Deep learning segmentation of the tear fluid reservoir under the sclera lens in optical coherence tomography images, Biomed. Opt Express 14 (5) (2023) 1848–1861, https://doi.org/10.1364/BOE.480247Journal.
[18] X. Li, et al., MINet: multi-scale input network for fundus microvascular segmentation, Comput. Biol. Med. 154 (2023) 106608, https://doi.org/10.1016/j.compbiomed.2023.106608.
[19] W. Liu, et al., A pyramid input augmented multi-scale CNN for GGO detection in 3D lung CT images, Pattern Recogn. 136 (2023) 109261, https://doi.org/10.1016/j.patcog.2022.109261.
[20] Y. Ye, et al., MFI-net: multiscale feature interaction network for retinal vessel segmentation, IEEE Journal of Biomedical and Health Informatics 26 (9) (2022) 4551–4562, https://doi.org/10.1109/JBHI.2022.3182471.
[21] L.-C. Chen, et al., DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848, https://doi.org/10.1109/tpami.2017.2699184.
[22] M. Drozdzal, et al., Learning normalized inputs for iterative estimation in medical image segmentation, Med. Image Anal. 44 (2018) 1–13, https://doi.org/10.1016/j.media.2017.11.005.
[23] F. Zeng, et al., Direct reconstruction for simultaneous dual-tracer PET imaging based on multi-task learning, EJNMMI Res. 13 (1) (2023), https://doi.org/10.1186/s13550-023-00955-w.
[24] C. Wachinger, M. Reuter, T. Klein, DeepNAT: deep convolutional neural network for segmenting neuroanatomy, Neuroimage 170 (2018) 434–445, https://doi.org/10.1016/j.neuroimage.2017.02.035.
[25] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, ICLR (2016) arXiv 2015, arXiv:1511.07122.
[26] K. Sun, et al., Deep high-resolution representation learning for human pose estimation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, https://doi.org/10.1109/CVPR.2019.00584.
[27] K. He, et al., Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, https://doi.org/10.1109/CVPR.2016.90.
[28] Z. Gu, et al., CE-net: context encoder network for 2D medical image segmentation, IEEE Trans. Med. Imag. 38 (10) (2019) 2281–2292, https://doi.org/10.1109/TMI.2019.2903562.
[29] Z. Dong, M. Xie, X. Li, Multi-scale receptive fields convolutional network for action recognition, Appl. Sci. 13 (6) (2023) 3403, https://doi.org/10.3390/app13063403.
[30] W.T. Zhu, et al., AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy, Med. Phys. 46 (2) (2019) 576–589, https://doi.org/10.1002/mp.13300.
[31] P. Boonyakitanont, A. Lek-Uthai, J. Songsiri, ScoreNet: a neural network-based post-processing model for identifying epileptic seizure onset and offset in EEGs, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 2474–2483, https://doi.org/10.1109/tnsre.2021.3129467.
[32] D.P. Huttenlocher, G.A. Klanderman, W.A. Rucklidge, Comparing images using the Hausdorff distance, IEEE Trans. Pattern Anal. Mach. Intell. (1993), https://doi.org/10.1109/34.232073.
[33] S. Takuhei, et al., In vivo crystalline lens measurements with novel swept-source optical coherent tomography: an investigation on variability of measurement, BMJ Open Ophthalmology 1 (1) (2017) e000058, https://doi.org/10.1136/bmjophth-2016-000058.
[34] M.X. Shen, et al., Agreement of corneal epithelial profiles produced by automated segmentation of SD-OCT images having different optical resolutions, Eye Contact Lens 40 (2) (2014) 99–105, https://doi.org/10.1097/icl.0000000000000017.
[35] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional Networks for Biomedical Image Segmentation, Springer International Publishing, Cham, 2015.
[36] L.-C. Chen, et al., Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-030-01234-2_49.
[37] M.D. Abràmoff, et al., Automated and computer-assisted detection, classification, and diagnosis of diabetic retinopathy, Telemedicine and e-Health 26 (4) (2020) 544–550, https://doi.org/10.1089/tmj.2020.0008.
[38] F. Wang, L.P. Casalino, D. Khullar, Deep learning in medicine—promise, progress, and challenges, JAMA Intern. Med. 179 (3) (2019) 293–294, https://doi.org/10.1001/jamainternmed.2018.7117.

[39] I.A. Viedma, et al., OCT retinal image-to-image translation: analysing the use of CycleGAN to improve retinal boundary semantic segmentation, in: 2021 Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2021, https://doi.org/10.1109/DICTA52665.2021.9647266.
[40] D. Romo-Bucheli, et al., Reducing image variability across OCT devices with unsupervised unpaired learning for improved segmentation of retina, Biomed. Opt Express 11 (1) (2020) 346–363, https://doi.org/10.1364/BOE.379978.
[41] A. Kirillov, et al., Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 4015–4026.