Expert Systems With Applications 207 (2022) 117968

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Artificial Intelligence-based computer-aided diagnosis of glaucoma using retinal fundus images

Adnan Haider, Muhammad Arsalan, Min Beom Lee, Muhammad Owais, Tahir Mahmood, Haseeb Sultan, Kang Ryoung Park *

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, Republic of Korea

ARTICLE INFO

Keywords:
Artificial intelligence
Optic cup and optic disc segmentation
Glaucoma screening
Computer-aided diagnosis
SLS-Net and SLSR-Net

ABSTRACT

Glaucoma is one of the most common chronic diseases that may lead to irreversible vision loss. The number of patients with permanent vision loss due to glaucoma is expected to increase at an alarming rate in the near future. A considerable amount of research is being conducted on computer-aided diagnosis for glaucoma. Segmentation of the optic cup (OC) and optic disc (OD) is usually performed to distinguish glaucomatous and non-glaucomatous cases in retinal fundus images. However, the OC boundaries are quite non-distinctive; consequently, the accurate segmentation of the OC is substantially challenging, and the OD segmentation performance also needs to be improved. To overcome this problem, we propose two networks, separable linked segmentation network (SLS-Net) and separable linked segmentation residual network (SLSR-Net), for accurate pixel-wise segmentation of the OC and OD. In SLS-Net and SLSR-Net, a large final feature map can be maintained in our networks, which enhances the OC and OD segmentation performance by minimizing the spatial information loss. SLSR-Net employs external residual connections for feature empowerment. Both proposed networks comprise a separable convolutional link to enhance computational efficiency and reduce the cost of network. Even with a few trainable parameters, the proposed architecture is capable of providing high segmentation accuracy.

The segmentation performances of the proposed networks were evaluated on four publicly available retinal fundus image datasets: Drishti-GS, REFUGE, Rim-One-r3, and Drions-DB which confirmed that our networks outperformed the state-of-the-art segmentation architectures.

1. Introduction

Glaucoma has become one of the major causes of vision loss, and in this disease, the optic nerve head (ONH) is damaged (Tham et al., 2014). Glaucoma causes gradual vision loss, and the patient has no abrupt considerable symptoms; hence, its early detection and screening are crucial. Many advanced imaging methods are employed by experts for retinal disease diagnosis and assessment. Fundus imaging is widely used in glaucoma detection tasks because it is fast, affordable, and non-invasive (Edupuganti et al., 2018). Color fundus imaging best serves the glaucoma detection in both advanced glaucoma or early glaucoma detection cases (Ahn et al., 2018). Fundus imaging also enables researchers and experts for computational analysis like cup-to-disc ratio (CDR) computation which significantly helps in glaucoma detection (Orlando et al., 2020).

Several methods have been used for the assessment of glaucoma; however, owing to numerous clinical and resource problems, they could not fill the gap of its early diagnosis (Baum et al., 1995). Compared to other methods, the ONH assessment is more commonly used. Automated ONH assessment methods are gaining popularity over manual methods these days because the manual assessment of ONH is a challenging task and requires specific expertise. Automated glaucoma detection can be broadly divided into two categories of machine learning-based methods, and deep learning-based methods. Machine learning is the subfield of artificial intelligence that deals with developing algorithms capable of pattern recognition from given data with minimal dictation of a human developer for recognition assistance (Thompson et al., 2020). However, in the conventional forms of machine learning, they require additional algorithm to extract features with optimal feature selection such as principal component analysis (PCA) and linear discriminant analysis (LDA), etc, from the raw input data. Traditional machine learning techniques may play an effective role in severe glaucoma cases but its

* Corresponding author.
E-mail address: parkgr@dongguk.edu (K.R. Park).

https://doi.org/10.1016/j.eswa.2022.117968
Received 10 March 2021; Received in revised form 2 May 2022; Accepted 22 June 2022
Available online 25 June 2022

performance for early glaucoma detection is not satisfactory (Ahn et al., 2018). Early glaucoma detection requires an efficient learning system therefore, deep learning-based methods show superior glaucoma screening accuracy compared to those of traditional machine learning methods (Ahn et al., 2018). In contrast with traditional machine learning, deep learning methods can be directly applied to raw data, and its algorithms enable to learn most relevant spatial features automatically (Thompson et al., 2020).

Deep learning-based systems learn from the training data and generate the parameters for feature extraction, classification, and regression while working on testing data (Gao et al., 2020). Deep learning-based techniques have caught the attention of researchers because of their accuracy, robustness, and cost effectiveness in the medical sector (Arsalan et al., 2020; Nguyen et al., 2020; Owais et al., 2019). In addition, computer-aided diagnosis using deep learning has been recently employed in many applications (Arsalan et al., 2019; Mahmood et al., 2020; Owais et al., 2019). Many studies have also been conducted on glaucoma screening using deep learning-based OD and OC segmentation (Thakur and Juneja, 2019).

Deep learning-based segmentation methods have been widely used for glaucoma detection because of its pixel-level accurate predictions (Al-Bander et al., 2018; Edupuganti et al., 2018; Wang et al., 2019). However, deep learning-based glaucoma detection approaches are not limited to segmentation only. Other deep learning-based methods are also used for glaucoma detection. Glaucoma screening can be performed using classification-based approaches, although there are many challenges associated to it. Cognizant of the morphology of the optic disc (OD) and optic cup (OC), numerous computations are performed on clinical indicators for glaucoma detection (Meier et al., 2007). The morphological changes in OD and OC with glaucoma progression can also be best analyzed with the help of segmentation. Accurate segmentation of OD and OC provides correct diameters for OD and OC, which results in accurate vertical CDR and it is widely accepted as the computational biomarker for glaucoma detection (Orlando et al., 2020; Soh et al., 2020; Wang et al., 2019; Yu et al., 2019). Vertical CDR is the ratio between the vertical diameter of the OC and that of the OD. Vertical CDR is one of the core measures for glaucoma detection. A low vertical CDR value indicates a low probability of occurrence of glaucoma, whereas a high vertical CDR value indicates a high probability of occurrence of glaucoma (Fu et al., 2018a).

The following are the problem statements resulting from previous research. First, the segmentation performance for both OD and OC in previous methods needs to be improved, which would consequently enhance the reliability of computer-aided diagnosis for glaucoma. The boundary of OC is quite non-distinctive; therefore, improving the corresponding segmentation performance is substantially challenging (Wang et al., 2019). Moreover, the computational efficiency of the network is one of the critical problems directly related to the memory requirements and robustness of the network (Sarkar and Das, 2017). Existing networks that achieve good segmentation performance usually use such a large number of trainable parameters, which makes the network computationally inefficient. Lastly, most of the existing developed networks require initialization and preprocessing in training to achieve the desired performance.

To overcome these problems, this study proposes two networks, separable linked segmentation network (SLS-Net) and separable linked segmentation residual network (SLSR-Net), for accurate pixel-wise segmentation of the OC and OD.

The contributions of the proposed work can be summarized as follows:

- Two novel networks, SLS-Net and SLSR-Net, are developed that are capable of achieving good segmentation performance with enhanced computational efficiency. Both networks maintain a large final feature map to avoid spatial information loss from minor details of the image.
- The proposed networks introduce a separable convolutional link (SCL) unit to enhance the computational efficiency of the networks. Our work best settles the challenging tradeoff between segmentation performance and network cost. Therefore, even with a few trainable parameters, the proposed architecture can provide high segmentation accuracy.
- External residual connections are developed in our final network (SLSR-Net) for feature empowerment to provide high segmentation performance using a shallow network. In addition, we make our networks available to other researchers (Dongguk SLS-Net and SLSR-Net., 2020) for fair comparisons.

The remainder of this paper is organized as follows. We explain related work on the segmentation of OD and OC in Section 2. The proposed work and experimental results with analyses are given in Sections 3 and 4. Finally, the conclusions are presented in Section 5.

2. Related work

Intelligent and expert systems play a vital role in the smart social development of the society. Deep learning is helping in the development of these expert systems by enabling detection, segmentation, and classification in many applications. Deep learning-based systems learn from the training data and generate parameters for feature extraction while working on testing data (Gao et al., 2020). We developed an expert system capable of aiding the glaucoma screening process using deep learning. In recent years, research trends abruptly increased along the lines of computer-aided diagnosis using deep learning. Deep learning-based techniques have attracted the attention of researchers owing to their accuracy, robustness, and cost effectiveness in the medical sector (Arsalan et al., 2020; Nguyen et al., 2020; Owais et al., 2019). Previous methods of segmentation of OD and OC for glaucoma detection in retinal fundus mages can be divided into two major categories: handcrafted-based and deep learning-based methods.

2.1. Segmentation based on conventional handcrafted features

Some approaches deal with hand-crafted characteristics for the segmentation purpose of OC and OD. Moreover, it is also reported that the segmentation of OC is more complex than OD segmentation because of the former's non-distinctive boundary. However, the overall system performance and robustness of these approaches are not up to expectations (Xu et al., 2014). We can further categorize segmentation based on handcrafted features in two parts.

2.1.1. Handcrafted features segmentation based on clustering and symmetry

Many techniques related to morphology were also introduced for the segmentation of OC and OD in retinal fundus images. These techniques are mainly dependent on the intensity values of an image; consequently, they cannot achieve good results in a variety of images (Srivastava et al., 2015). To mitigate the performance degradation problem, some algorithms based on thresholding were developed to find the desired information of the OC considering its symmetry. Borderlines of OC were finally detected using mathematical models. This methodology is also used in clustering-related approaches (Mittapalli and Kande, 2016). There has been a common problem of variation handling in handcrafted-based methods therefore a technique of combining clustering and level-set is used for the segmentation of OD intuitionally. This combination of different approaches can improve variation handling and results. However, this type of methodology may face problems because of the variety in OD and OC cases. Some of these examples include different morphologies, intensity values, and complicated shapes of target classes (Thakur and Juneja, 2019).

Table 1
Comparison between proposed and existing methods for OD and OC segmentation.

Type	Method	Strength / Motivation	Weakness
Handcrafted features-based	Low-rank superpixel representation for OC segmentation (Xu et al., 2014; Srivastava et al., 2015)	Considers morphology and image intensity values. Can aid in non-distinctive object (OC) segmentation.	Bound to unsupervised learning, and post-processing is also used.
	Clustering and level-set combination (Mittapalli and Kande, 2016; Thakur and Juneja, 2019)	Improved results with variation handling along with symmetry consideration were achieved using clustering and level-set combination.	Faced problems in different morphology, intensity values, and complicated shapes of OD and OC.
	The template-based matching technique (Roychowdhury et al., 2016; Septiarini et al., 2018)	Considers the correlation, texture, and morphology of OD and OC. A relatively simpler approach for the segmentation of OD and OC segmentation.	OD and OC shapes are not always regular, circular, or oval. Consequently, their boundary sometimes cannot be accurately extracted. Requires additional pre-processing.
	Contemplating the pixel-level features (Saeed et al., 2019; Tan et al., 2015; Zhou et al., 2017)	Extraction of minor details in images is always challenging. Examining pixel-level minor details. Overall segmentation performance was improved.	Unable to show remarkable results in a lesser time while working with large datasets.
Deep learning-based	U-Net-based CNN (Sevastopolsky, 2017; Orlando et al., 2017))	The employed encoder-decoder-based model improves the segmentation performance by taking less prediction time even using limited data.	Requires additional pre-processing.
	Adversarial learning (Wang et al., 2019; Shankaranarayana et al., 2017)	Encoder-decoder structure addresses the key problem of domain-shifting using patch-based adversarial learning.	Deficient in OC segmentation performance.
	U-net + ResNet-34 and Disc-aware method (Yu et al., 2019; Fu et al., 2018b)	Less training time with acquiring more relevant details using encoder-decoder architectures	Fine-tuning is used with prior weight initialization.
	M−Net multi-label approach (Fu et al., 2018a)	Multi-level input can be fed, and a better prediction can be obtained using this encoder-decoder architecture	Faced with performance degradation on blurred and low-contrast image segmentation.
	Two-stage Mask-RCNN (Almubarak et al., 2020)	This is a simple two-step method. First, the optic nerve head is detected. In the second stage, segmentation is performed using a weighted loss.	Pre-processing is used with prior weight initialization.
	neural network + ensemble (Jiang et al., 2019; Zilly et al., 2017)	GL-Net shows good performance with noisy and lesion images. Ensemble-based use fewer parameters	Requires additional pre-processing.
	ImageNet–trained CNN architectures (Diaz-Pinto et al., 2019).	Five separate CNNs trained on ImageNet produce improved results with fusion for glaucoma screening.	Pre-processing is required.
	Transfer learning-based google Inception v3 model (Ahn et al., 2018)	Inception v3 provides relatively better results than GoogleNet with fewer parameters.	Needs fine-tuning for optimal results.
	FC-DenseNet approach (Al-Bander et al., 2018)	The segmentation performance of DenseNet for OC is better with reduced parameters.	Pre-trained weights are required for this technique.
	deep learning network with post-processing (Edupuganti et al., 2018)	This method improved the segmentation performance using FCN without resizing images.	Post-processing is accomplished for final results.
	Adaptive region-based edge smoothing model (Haleem et al., 2017)	The adaptive edge smoothing update model enables the model to capture a variety of image shapes with different variants and also helps to find initial boundaries.	Post-processing is accomplished for final results.
	SLS-Net and SLSR-Net (Proposed)	Minimize spatial information loss by maintaining large feature map size. Optimize the computational efficiency and segmentation performance trade-off with the help of separable convolutional link. Enhance segmentation performance by external residual connections.	Prior data augmentation is required for the segmentation process.

2.1.2. *Handcrafted features segmentation based on template, texture, and pixel-level details.*

Template-based matching techniques were developed for the segmentation of OD and OC. In one of the techniques, a prior circular template was used for the segmentation of OD considering the morphology and shape of the OD. However, the OD shape is not always circular or oval; hence, the OD boundary cannot be accurately extracted (Roychowdhury et al., 2016). Since glaucoma brings changes in the

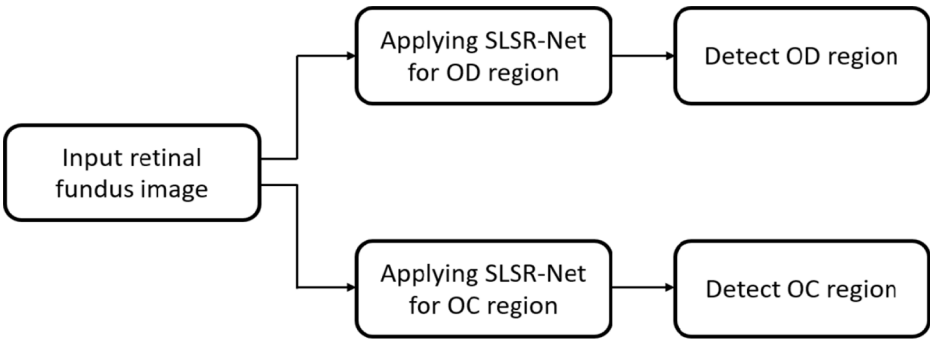


Fig. 1. Flowchart of the proposed method.

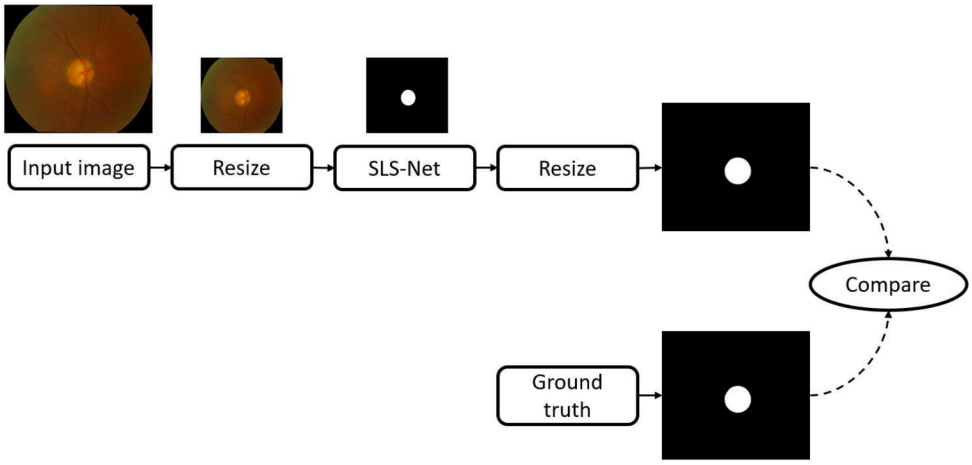


Fig. 2. SLS-Net overview.

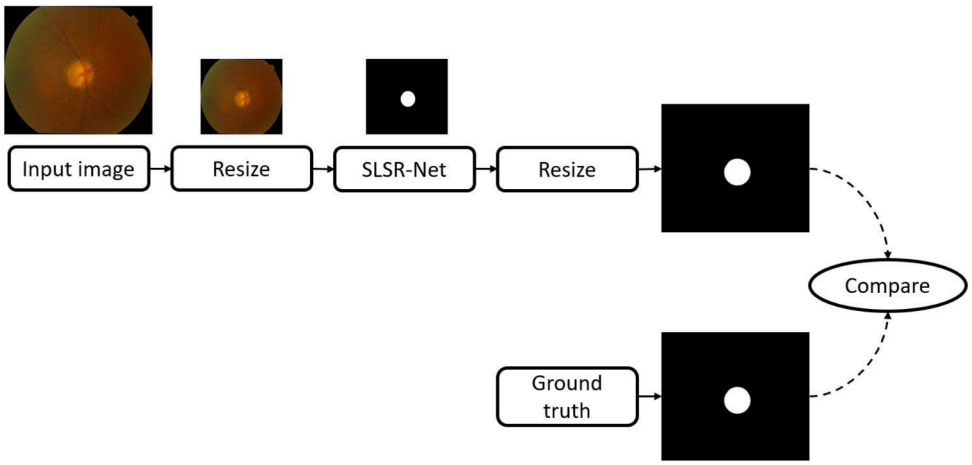


Fig. 3. SLSR-Net overview.

morphology of OD and OC which produce changes in the retinal nerve fiber layer (RNFL), and early detection of these changes helps in glaucoma screening. In this study, a technique was introduced which considers the texture and correlation trends for features selection. This method achieved good accuracy and performance, working with only two features; nonetheless, its results can be further improved (Septiarini et al., 2018). Considering the textures helps in extracting the global information whereas extraction of minor details in images is always challenging. To address this problem, a considerable amount of research has been conducted on contemplating the pixel-level minor details of an image for the extraction of OD. However, the approach was unable to show remarkable results with a low processing time while working with large datasets (Saeed et al., 2019; Tan et al., 2015; Zhou et al., 2017).

2.2. *Segmentation based on deep features*

Traditional approaches have been widely replaced with convolutional neural networks (CNNs) and deep learning. Generally, deep learning-based approaches follow the same sequence for the detection of glaucoma. Segmentation is normally performed, followed by clinical

Table 2
Key differences of SLS-Net and SLSR-Net with U-Net (Ronneberger et al., 2015) and SegNet (Badrinarayanan et al., 2017).

Models	SLS-Net	SLSR-Net	U-Net	SegNet
Number of convolution layers	Total 14 convolution layers + 2 separable convolution layers.		Total 23 convolution layers	Total 13 convolution layers
With or without separable convolution	Separable convolution link to reduce the overall cost of the network.		No	
Final feature map size	62 × 62	62 × 62	31 × 31	31 × 31
Number of max channel	512 (used once)	512 (used once)	1024	512 (used twice)
Number of pooling layer	3	3	4	5
With or without skip connection	No skip connection	External residual skip connections	Dense skip connection	No skip connection
Number of parameters	4.6 million		31 million	29.4

analysis for assisting the glaucoma screening process.

2.2.1. Segmentation based on deep features using encoder-decoder architectures

In segmentation, many architectures use encoder-decoder-based architectures inspired by famous U-Net and SegNet. In glaucoma detection, vertical CDR is one of the basic measures for which the segmentation of OC and OD is performed. Modification in U-Net was proposed to improve these CDR results, which consequently enhanced the segmentation performance as a state-of-the-art framework (Sevastopolsky, 2017). Adversarial learning also improved the segmentation performance of retinal fundus images. An adversarial learning-based network, ResU-net, was developed for the detection of glaucoma. This approach eliminates the preprocessing procedure and uses fully convolutional networks (FCNs) to improve segmentation performance for the joint segmentation of OC and OD (Shankaranarayana et al., 2017). Few frameworks train the network on non-medical images before formal training. This type of automation for glaucoma screening requires extensive preprocessing such as cropping, contrast adaptation, and histogram equalization (Orlando et al., 2017). Traditional approaches

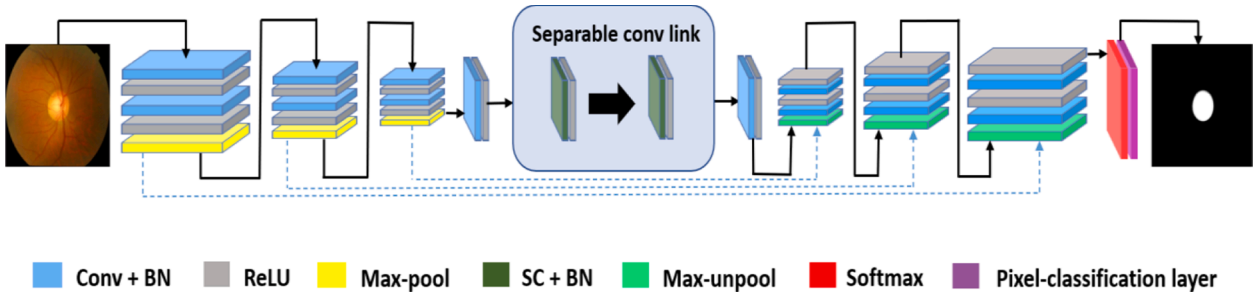


Fig. 4. Proposed SLS-Net.

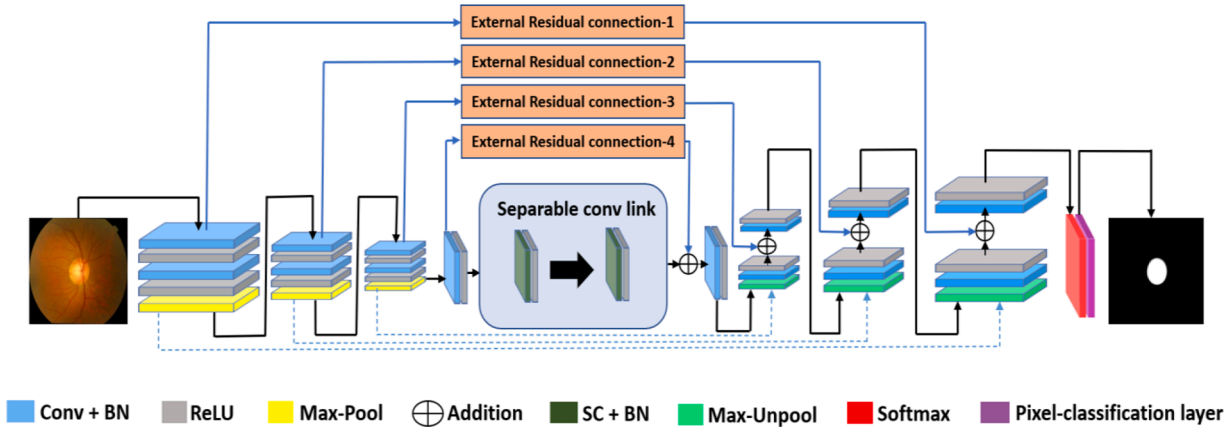


Fig. 5. Proposed SLSR-Net.

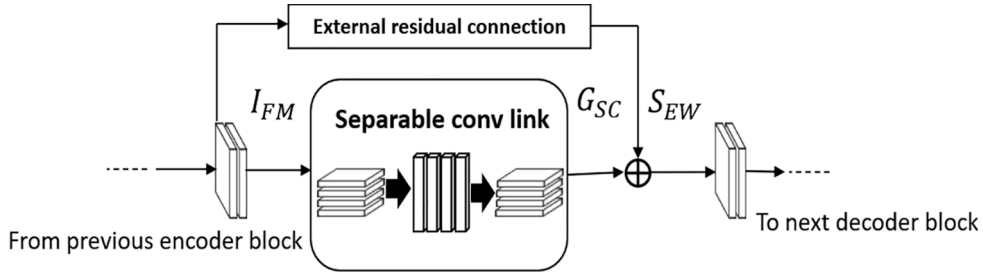


Fig. 6. Schematic of SCL with the external residual connection.

Table 3
Comparison of parameters between SCL layers and expected standard convolution layers.

Layer Name	Conv + ReLU-4 (Expected layer)	Separable-Conv_A	Conv + ReLU4 (Expected layer)	Separable-Conv_B	Total parameters saved
Unit	Pre-SCL-unit-4	SCL-unit	Post-SCL-unit-4	SCL-unit	–
Parameters	$(512 \times 512 \times 3 \times 3) + 512$ $= 2{,}359{,}808$	$(512 \times 3 \times 3)$ $+ 512$ $= 5{,}120$	$(512 \times 512 \times 3 \times 3) + 512$ $= 2{,}359{,}808$	$(512 \times 3 \times 3)$ $+ 512$ $= 5{,}120$	$(2{,}359{,}808 \times 2)$ $-(5{,}120 \times 2)$ $= 4{,}709{,}376$

Table 4
SLSR-Net computational details of all units along with respective layers, output feature map size, and filters. The network architecture consists of four pre-SCL units, one SCL-unit itself, and four post-SCL units. The number of parameters used by corresponding layers is given in the last column. SLSR-Net uses images of dimensions 500×500 pixels for Drishti-GS dataset. SCL enables the network to work cost effectively by using minimum parameters.

Unit	Name/Size	No. of filters		Output feature map size (width × height × number of channels)	No. of parameters
Pre-SCL-unit-1	Conv1_A + ReLU-1_A/3 × 3 × 3	64		500 × 500 × 64	1,792
	BN 1_A	–			128
	Conv1_B + ReLU-1_B /3 × 3 × 64	64			36,928
	BN 1_B	–			128
Max-pool_1	Pool 1/ 2 × 2	–		250 × 250 × 64	–
Pre-SCL-unit-2	Conv2_A + ReLU-2_A/3 × 3 × 64	128		250 × 250 × 128	73,856
	BN 2_A	–			256
	Conv2_B + ReLU-2_B /3 × 3 × 128	128			147,584
	BN 2_B	–			256
Max-pool_2	Pool 2/ 2 × 2	–		125 × 125 × 128	–
Pre-SCL-unit-3	Conv3_A + ReLU-3_A/3 × 3 × 128	256		125 × 125 × 256	295,168
	BN 3_A	–			512
	Conv3_B + ReLU-3_B /3 × 3 × 256	256			590,080
	BN 3_B	–			512
Max-pool-3	Pool 3/ 2 × 2	–		62 × 62 × 256	–
Pre-SCL-unit-4	Conv4_A + ReLU-4_A/3 × 3 × 256	512		62 × 62 × 512	1,180,160
	BN 4_A	–			1,024
SCL-unit	Sep-Conv_A/3 × 3 × 1	512		62 × 62 × 512	5,120
	BN_A	–			1,024
	Sep-Conv_B/3 × 3 × 1	512			5,120
	BN_B	–			1,024
Post-SCL-unit-4		Conv4_A + ReLU-4_A/3 × 3 × 512	256	62 × 62 × 256	1,179,904
		BN 4_A	–		512
Un-pooling 3		Un-pool 3	–	125 × 125 × 256	–
Post-SCL-unit-3		Conv3_A + ReLU-3_A/3 × 3 × 256	256	125 × 125 × 256	590,080
		BN 3_A	–		512
		Conv3_B + ReLU-3_B /3 × 3 × 256	128	125 × 125 × 128	295,040
		BN 3_B	–		256
Un-pooling 2		Un-pool 2	–	250 × 250 × 128	–
Post-SCL-unit-2		Conv2_A + ReLU-2_A/3 × 3 × 128	128	250 × 250 × 128	147,584
		BN 2_A	–		256
		Conv2_B + ReLU-2_B /3 × 3 × 128	64	250 × 250 × 64	73,792
		BN 2_B	–		128
Un-pooling 1		Un-pool 1	–	500 × 500 × 64	–
Post-SCL-unit-1		Conv1_A + ReLU-1_A/3 × 3 × 64	64	500 × 500 × 64	36,928
		BN 1_A	–		128
		Conv1_B + ReLU-1_B /3 × 3 × 64	2	500 × 500 × 2	1,154
		BN 1_B	–		4
Total number of trainable parameters					4,666,950

have been widely replaced with convolutional neural networks (CNNs) and deep learning. Generally, deep learning-based approaches follow the same sequence for the detection of glaucoma. Segmentation is normally performed, followed by clinical analysis for glaucoma screening. In an approach, deep learning-based algorithms can acquire more relevant details from the image for direct detection of glaucoma using different streams. This framework was extensively evaluated on Singapore Chinese eye study (SCES) and the Singapore Indian Eye Study (SINDI) datasets instead of commonly used ones such as Drishti and RIM-ONE databases (Fu et al., 2018b). In this study, a famous M−Net model was developed by modifying U-net to jointly segment the OC and OD. A multi-scale input layer and side output layers were also introduced in this model. Multi-level input can be fed, and as a result, better prediction can be obtained for scaling of layers (Fu et al., 2018a). In deep learning, networks are commonly trained and tested on the same dataset. However, when the same trained network is used for testing with different datasets, domain shifting problems are produced. This problem was addressed using the patch-based output space adversarial learning method for domain adaptation. In this work, results were obtained by training the network on one dataset and testing on different datasets. Furthermore, one of the networks was also trained and tested on the same dataset. Preprocessing was also used to achieve the presented results. However, there is a gap in the improvement in the segmentation performance of OC (Wang et al., 2019). Another method is proposed, which uses a combination of two separate networks in an encoder–decoder manner: the residual network-34 (ResNet-34) architecture was used as an encoder, whereas the famous U-net was employed as the decoder of the model. The training time for this combined model is very less, and it yields improved results. Moreover, this method uses fine-tuning to produce the presented results. In this research, two stages

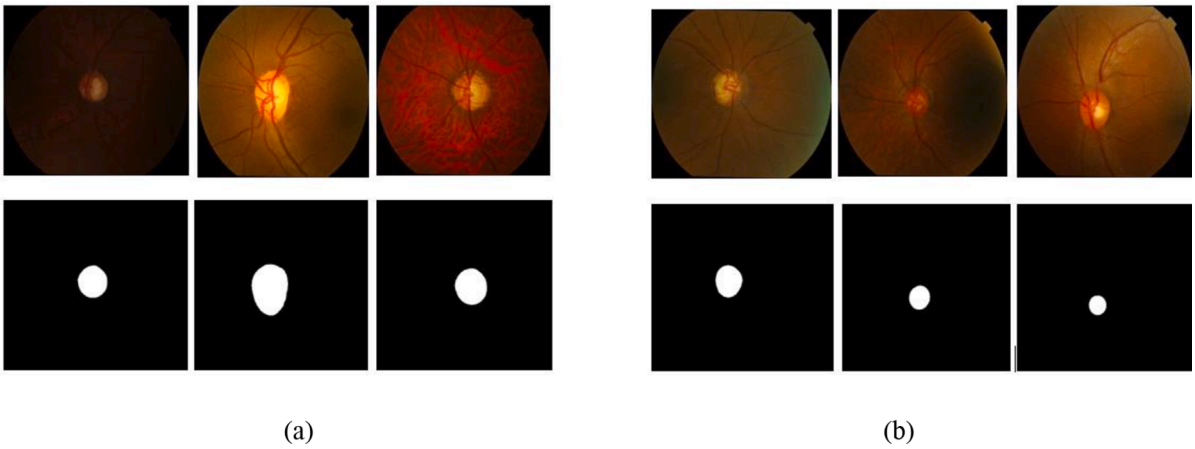


Fig. 7. Example images of the Drishti-GS database with the ground truth for (a) OD and (b) OC.

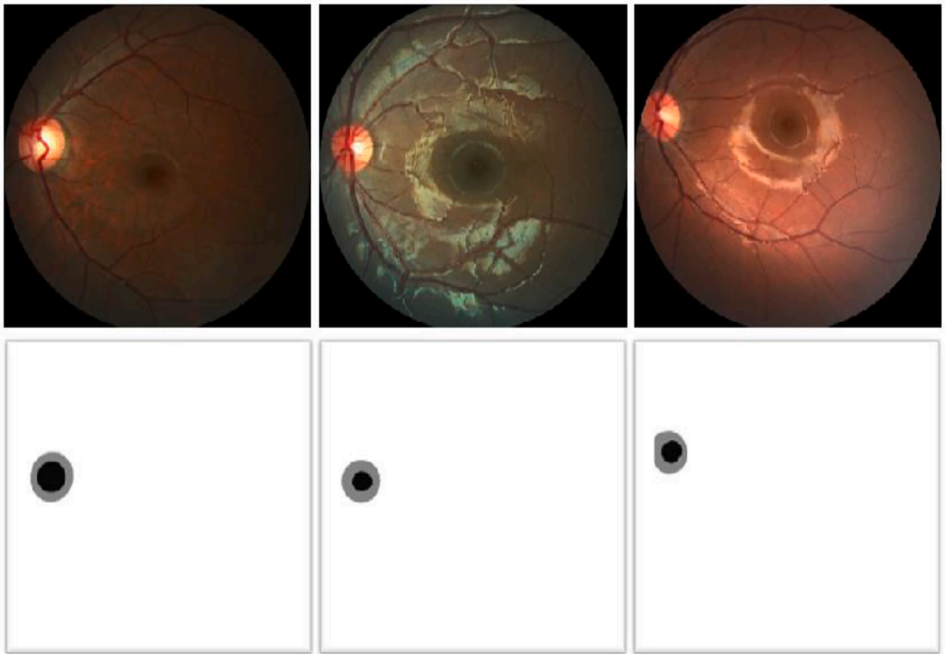


Fig. 8. Example images of the REFUGE database with ground truth. The black region in the ground truth image represents OC while the outer gray region of ground truth represents OD.

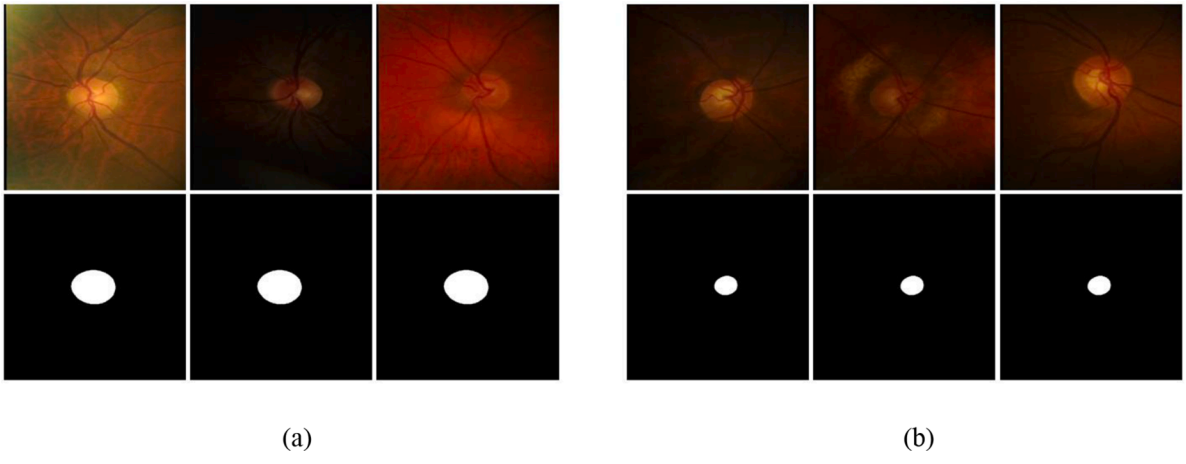


Fig. 9. Example images of the Rim-One-r3 database with the ground truth for (a) OD and (b) OC.

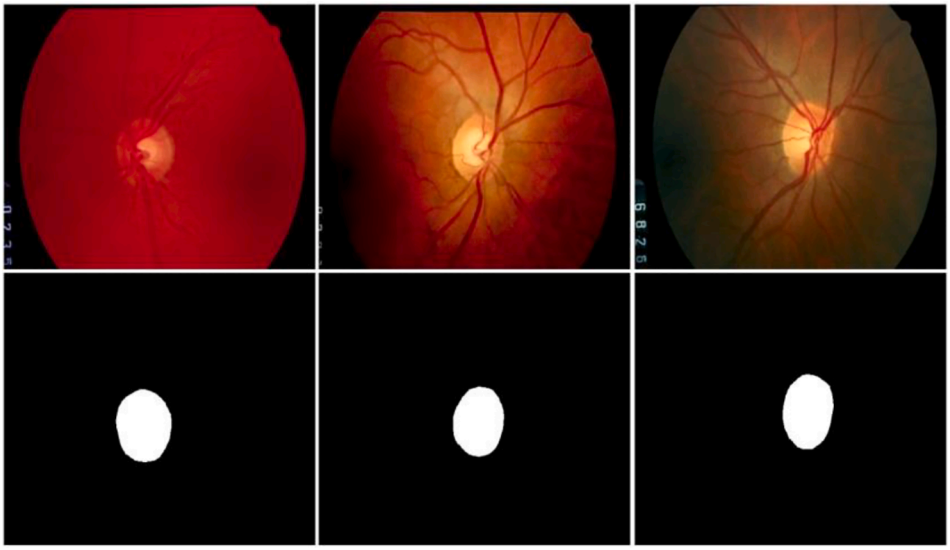


Fig. 10. Example images of the Drions-DB with ground truth. The white region in the ground truth image represents OD.

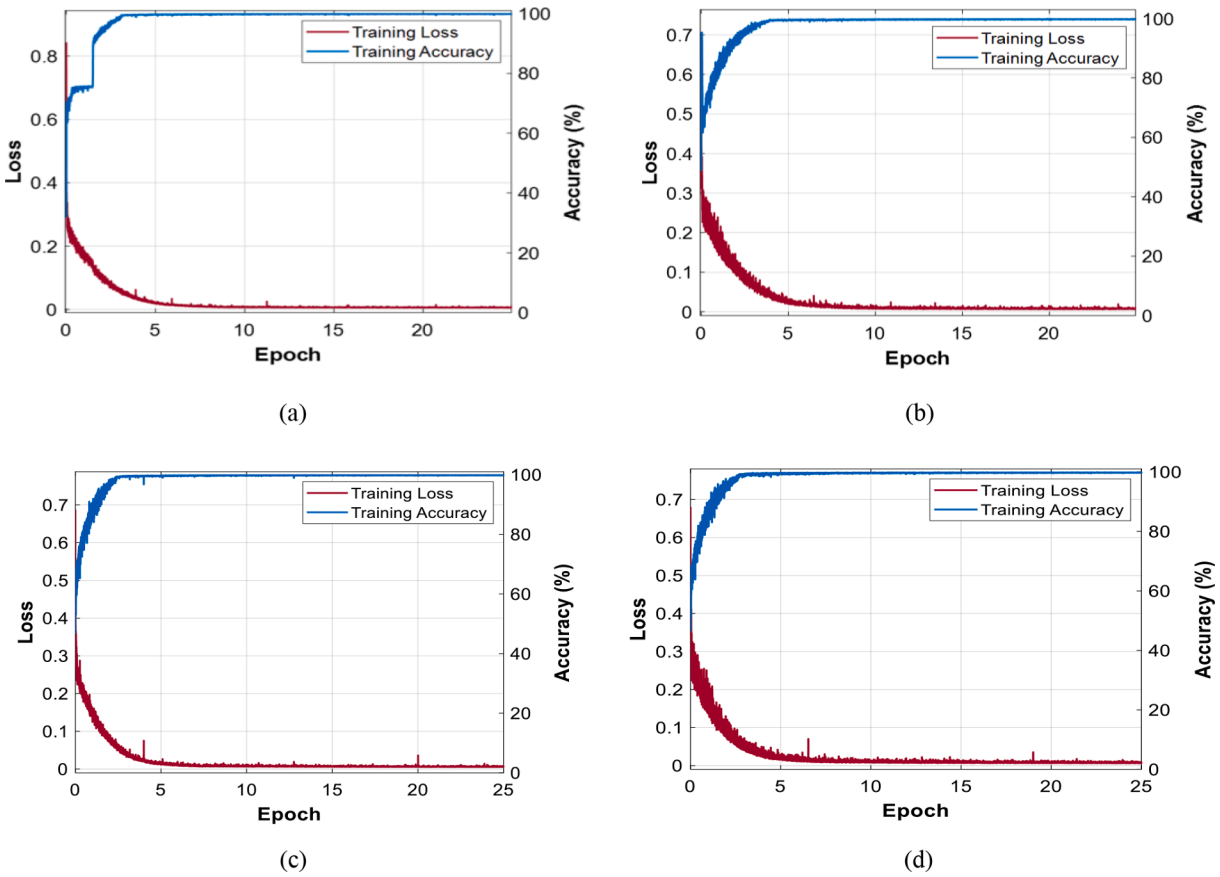


Fig. 11. Training loss and accuracy curves with SLS-Net for (a) OD (b) OC and those with SLSR-Net for (c) OD (d) OC.

Table 5
Comparison between SLS-Net and SLSR-Net with respect to Drishti-GS dataset.

Methods	OC					OD				
	Acc	Sen	Spec	DC	JI	Acc	Sen	Spec	DC	JI
SLS-Net	0.996	0.930	0.998	0.90	0.828	0.998	0.957	0.999	0.974	0.950
SLSR-Net	0.997	0.947	0.998	0.915	0.853	0.998	0.986	0.999	0.980	0.962

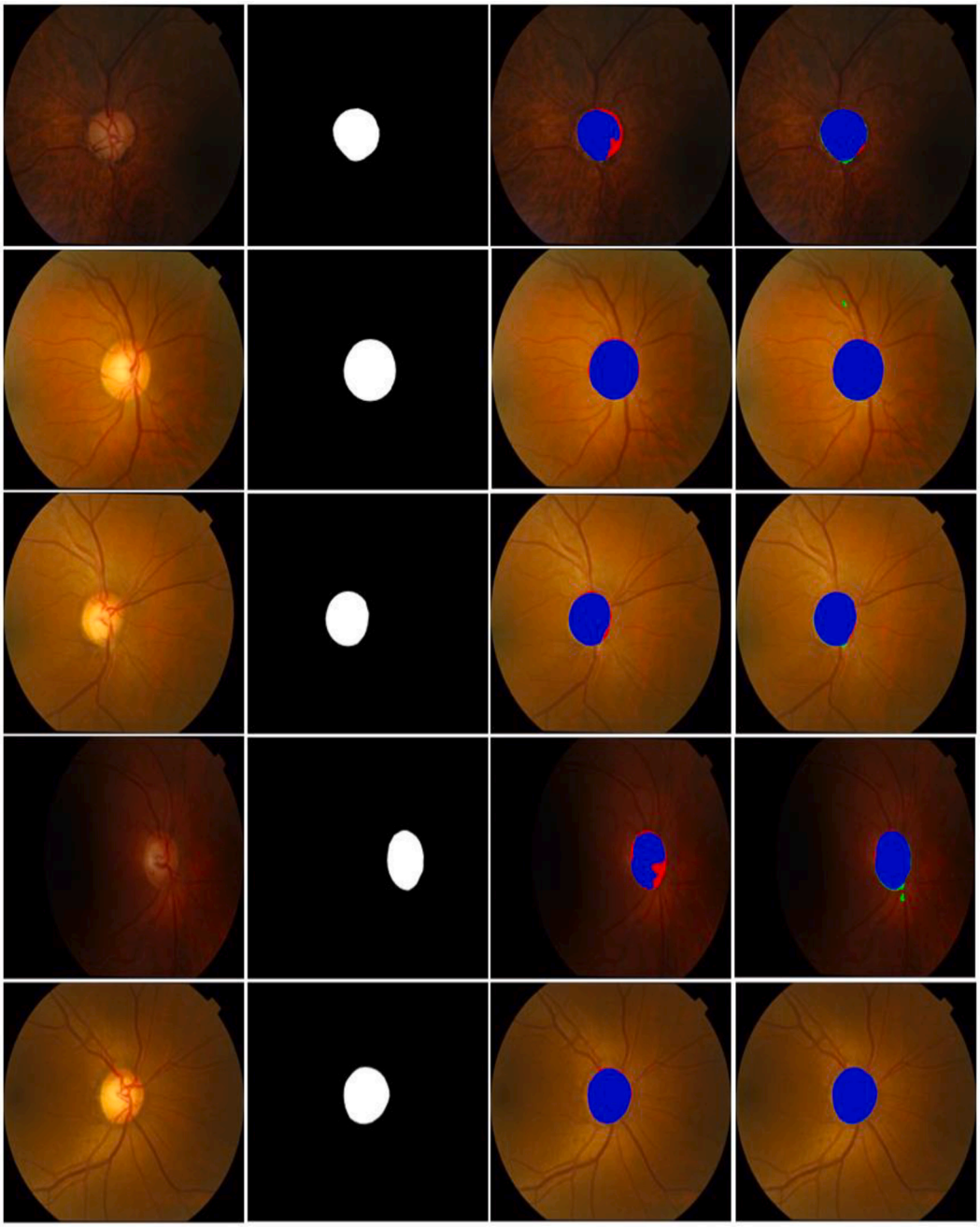


Fig. 12. Good segmentation results for OD with Drishti-GS dataset, obtained using SLS-Net and SLSR-Net. (a) Original image, (b) ground truth image, and segmented images obtained using (c) SLS-Net and (d) SLSR-Net (*tp*, *fp*, and *fn* pixels are shown in blue, green, and red, respectively).

are proposed to localize the ONH and segment the OD/OC for glaucoma detection. In the first stage, cropping is performed around the ONH, and the output image is provided to the second stage. In this stage, training is carried out by a weighted loss function to finally segment the OD and OC. After training with the one dataset, this developed framework also shows remarkable performance for other datasets (Yu et al., 2019). Another method is proposed, which uses a combination of two separate networks in an encoder-decoder manner: the residual network-34 (ResNet-34) architecture was used as an encoder, whereas the famous U-net was employed as the decoder of the model. The training time for this combined model is very less, and it yields improved results (Almubarak et al., 2020).

2.2.2. Segmentation based on deep features using pre-processing

Many segmentation-based methods use pre-processing to achieve the desired segmentation performance. Ensemble learning is usually used to enhance the computational efficiency of the network. Desired spots from the image are only selected through sampling for learning, which enables the network to work with a few parameters. Such approaches are used when the dataset contains a few images (Zilly et al., 2017). In traditional deep learning, restoring the context information becomes difficult due to insufficient feature information extraction, which ultimately deteriorates the segmentation results. Another model, GL-Net, was developed, which works on the generative adversarial approach. On the generator side of the model, low-level and high-level features are combined using skip connections. In this way, the loss of information is

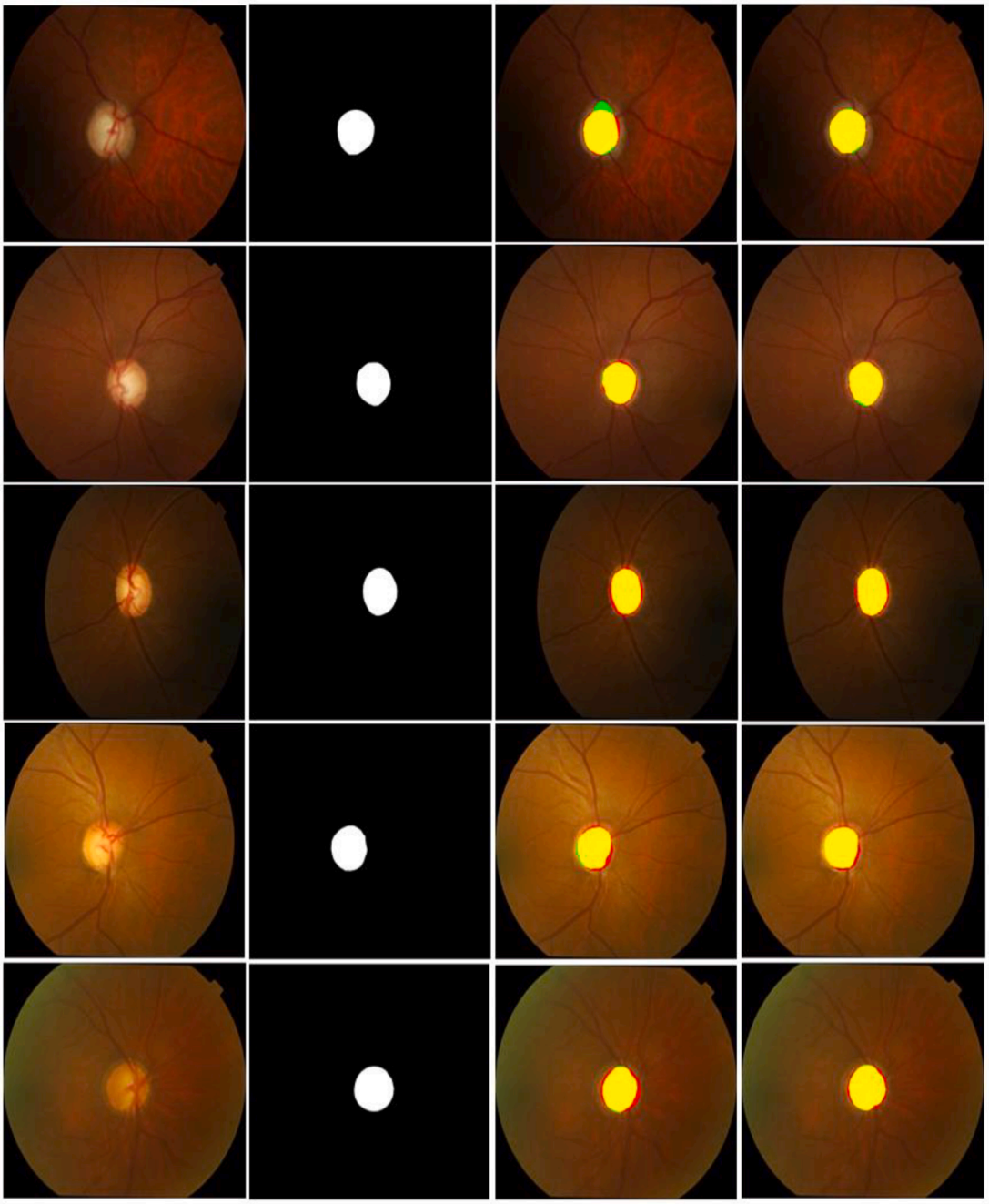


Fig. 13. Good segmentation results for OC with Drishti-GS dataset, obtained using SLS-Net and SLSR-Net. (a) Original image, (b) ground truth image, and segmented images obtained using (c) SLS-Net and (d) SLSR-Net (*tp*, *fp,* and *fn* pixels are shown in yellow, green, and red, respectively).

avoided, and feature extraction is improved because detailed feature information is usually lost while passing through the upsampling and downsampling stages. The performance of GL-Net with noisy and lesion images was remarkably good, though the method has some limitations in the high-level processing of features. By improving this, the aggregate performance of the model can be improved (Jiang et al., 2019). In previous studies, five separate CNNs were trained on ImageNet for glaucoma screening, and the network was evaluated on four publicly available datasets. The results produced by this architecture were promising with a few limitations. Although preprocessing is accomplished, slight performance degradation is also observed when training and testing datasets are not the same (Diaz-Pinto et al., 2019).

2.2.3. Segmentation based on deep CNN with fully connected layers.

In some of the deep learning-based approaches fully connected networks are used for predictions. Transfer learning was used in some of the deep learning models for glaucoma detection at any stage. In this work, the Inception V3 model was used for transfer learning. The change in the last layer of Inception V3 yielded better results than GoogleNet (inception V1) with a slightly smaller number of parameters. However, the model is overall complex, and for a small dataset, its robustness and performance need to be further improved (Ahn et al., 2018). The DenseNet core was used with FCNs for OD and OC segmentation. The efficiency of the developed methodology was assessed by applying it to retinal datasets for the detection of glaucoma. The model requires a long training time, which is one of the major limitations of this architecture

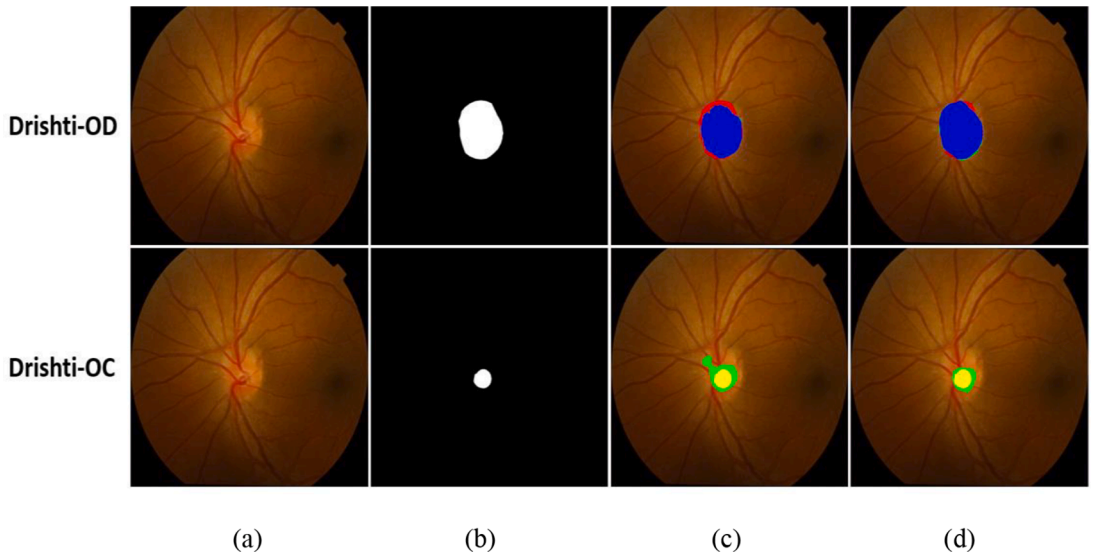


Fig. 14. Bad segmentation results for OD and OC with Drishti-GS dataset, obtained using SLS-Net and SLSR-Net. (a) Original image, (b) ground truth image, and segmented images obtained using (c) SLS-Net and (d) SLSR-Net (for OD and OC, *tp* pixels are presented in blue and yellow, respectively, whereas *fp* pixels are shown in green, and *fn* pixels are shown in red).

(Al-Bander et al., 2018).

2.2.4. Segmentation based on deep features with post-processing

Many segmentation-based methods use post-processing to achieve the desired segmentation performance. In a work, visual geometry group-16 (VGG16), was designed to segment retinal images of the Drishti-GS dataset. Using FCNs, prioritized training was incorporated, and CDR was computed for OD and OC segmentation. However, this network is complex and uses many parameters (Edupuganti et al., 2018). A deformable approach is used to reduce the error in the classification. In this approach, edge detection and smoothing of edges are performed for different objects, and an artificial neural network is used for result refinement. This scheme enables the model to capture a variety of image shapes using post-processing (Haleem et al., 2017).

Although the segmentation performance for glaucoma detection has improved in the last few years, it is still not up to expectations. Most of the architectures delivering good segmentation performance use many parameters, which results in excessive memory needs and computational overheads. To overcome these problems, we proposed two novel networks, SLS-Net and SLSR-Net, optimizing the trade-off between computational efficiency and segmentation performance. Comprehensive comparisons between the previous and proposed methods are presented in Table 1.

3. Proposed methods

3.1. Overview of the proposed methods

A flowchart of the proposed method for OC and OD segmentation is shown in Fig. 1. Developed architectures take retinal fundus images as input through the image input layer. In this study, SLS-Net is the base network whereas SLSR-Net is the final network which is the extended form of SLS-Net. SLSR-Net is developed for enhancing the segmentation performance of SLS-Net. As shown in Fig. 1, both networks detect the corresponding pixels of OD and OC regions.

3.1.1. SLS-Net overview

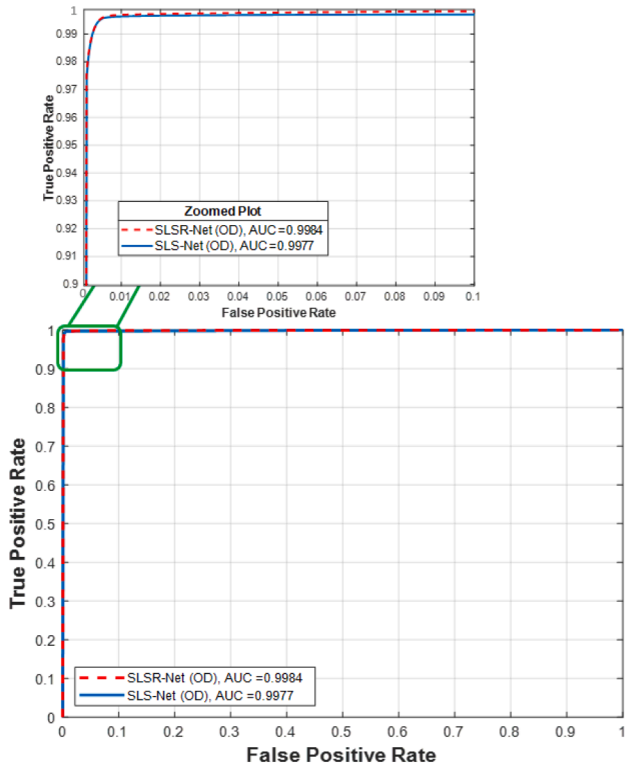
As stated, SLS-Net serves as a base network for aiding the automatic glaucoma detection process. In SLS-Net, a large final feature map is retained which reduces the information loss and helps in achieving high segmentation accuracy. Since network cost is also one of the key aspects for implementation therefore in our base network SCL is established between the encoder side (pre-SCL unit) and decoder side (post-SCL unit) to reduce the overall cost of the network. In other methods, usually segmentation performance is achieved with expensive networks. We employed depth-wise separable convolution in SCL that can reduce the number of parameters compared with conventional layers, which is explained in detail in subsection 3.4. As shown in Fig. 2, original images are resized before feeding to the SLS-Net. SLS-Net generates a prediction mask of the same size which is later resized to the original size using bilinear interpolation for a fair comparison. This original size prediction mask is compared with the original sized ground truth.

3.1.2. SLSR-Net overview

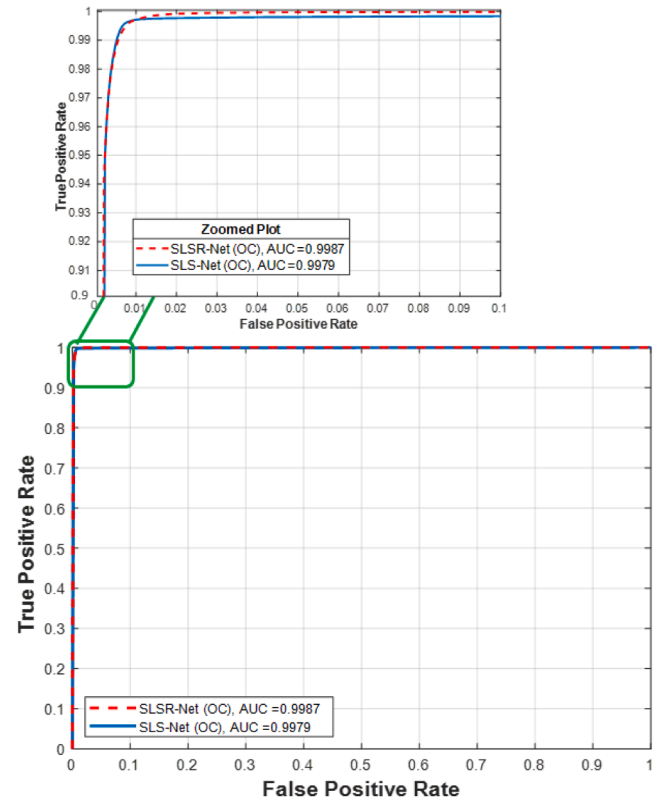
SLSR-Net is an advanced network and it uses the base of SLS-Net. SLS-Net was needed to improve segmentation performance by fixing the feature degradation problem caused by the series of operating layers. This problem is fixed using our final network, SLSR-Net. In SLSR-Net, external residual skip connections are developed to empower features and overcome the downsampling loss. This feature empowerment enables SLSR-Net to acquire minor details of the image, which results in the improvement of segmentation performance. The predicted pixels belonging to the OC and OD boundaries are finally detected in the last stage as an output of the networks. The overview of the SLSR-Net evaluation is shown in Fig. 3. SLSR-Net receives resized input fundus images and generates prediction masks for the desired class. This generated mask is of the same size therefore it is resized back to its original size using bilinear interpolation. This original-sized mask is finally compared with the original ground truth for a fair comparison.

3.2. Motivation for developing SLS-Net and SLSR-Net

Both SLS-Net and SLSR-Net use 14 standard convolutions and 2 separable convolutions (SC) layers. The inspiration for encoder-decoder-based network design was taken from famous U-Net (Ronneberger et al., 2015) and SegNet (Badrinarayanan et al., 2017). However, there are several limitations associated with U-Net and SegNet, which motivates the other researchers to play their part. Because the final feature map size of U-Net and SegNet is very small, the detection of indistinctive objects with high accuracy becomes challenging. OC boundaries in different datasets are indistinctive therefore its segmentation accuracy using U-Net and SegNet is usually not good. As given in Table 2, U-Net



(a)



(b)

Fig. 15. Graphs of Sen according to 1 – Spec obtained using SLS-Net and SLSR-Net on Drishti-GS dataset for the segmentation of (a) OD and (b) OC. Green box shows the selected area of graph for zoomed plot.

and SegNet require a large number of trainable parameters which makes them computationally less efficient. In U-Net, a maximum depth of 1024 is used whereas in SLSR-Net a maximum depth of 512 is used to save trainable parameters. SegNet uses a depth of 512 for two times whereas SLSR-Net uses a depth of 512 for only one time. Moreover, we also use SCL which saves a huge number of parameters. Similarly, some key differences of SLS-Net and SLSR-Net with U-Net and SegNet are presented in Table 2.

3.3. SLS-Net-based segmentation of OC and OD

Significant research has been conducted for glaucoma screening through OC and OD segmentation. In most cases, an improvement in segmentation performance was achieved with expensive networks. The cost of the network is one of the basic criteria for evaluating network efficiency. In the proposed SLS-Net, we introduced SCL to overcome this problem and make the network computationally inexpensive. Depth-wise separable convolution in SCL can reduce the number of parameters compared to that with conventional convolution layers, which is explained in detail in Section 3.3.

In traditional CNNs, the final feature map size is set to be small during downsampling, leading to spatial information loss. In particular, the minor features of the image are lost when passing through a series of convolution layers, which results in degraded network performance (Yu et al., 2017). However, the final feature map size of SLS-Net is maintained substantially large to overcome this critical problem, and as a result, information loss is significantly minimized. The size of the final feature map in the proposed model is $62 \times 62 \times 512$, which is large enough to retain the minor features of the image. The network diagram of SLS-Net in Fig. 4 shows that the input image is directly fed to the convolution layer through the input image layer. The encoder section before the SCL-unit is named the pre-SCL unit, while the decoder section after the SCL-unit is named the post-SCL unit. In the pre-SCL unit, the input image is processed in four stages. Each stage contains 3×3 convolution layers with batch-normalization (BN) layers. Rectified linear unit (ReLU) is the activation function, and the max-pooling (Max-pool) layers downsample the feature map size. There are seven convolution layers and three max-pooling layers in the pre-SCL unit. The network is designed such that every Max-pool layer reduces the feature map size, and the last convolution layer of the pre-SCL unit possesses the smallest feature map. Previous researchers have concluded that OC segmentation is comparatively more challenging than OD segmentation (Wang et al., 2019). Most of the networks could not achieve significant segmentation performance for OC because of the extraordinarily small final feature map size. Such a small final feature map size loses minor details of the non-distinctive OC boundary. SLS-Net does not drop the final feature map size to this extent. In the post-SCL unit, there are seven convolution layers and three Max-unpool layers used to upsample the feature map size to its normal level. The output is finally extracted using the softmax and pixel classification layers.

In the SCL unit, depth-wise SC is used to enhance the computational efficiency of the system. This SC is a spatial convolution that works on entire channels; thereafter, point-wise convolution is performed (Chollet, 2017). The development of SC has enabled researchers to replace standard convolution with SC if applicable and feasible. In many famous architectures, SC was introduced to make the network cost effective. In SLS-Net, the SCL output is fed to the post-SCL unit where convolution and unpooling operations are applied, and the image is upsampled to its original size. A segmentation mask is obtained as a result of using the network with the help of softmax and pixel classification layers.

However, owing to the feature degradation problem, the SLS-Net performance was further improved with the development of SLSR-Net. In encoder–decoder-based CNNs, the features of images are usually lost/degraded during several layer operations; however, this problem is successfully addressed using our advanced SLSR-Net.

Table 6
Comparisons of segmentation accuracies for OC and OD obtained using the proposed SLS-Net and SLSR-Net with those of state-of-the-art methods on Drishti-GS dataset. "-" means that there is no accuracy reported.

Methods	OC					OD				
	Acc	Sen	Spec	DC	JI	Acc	Sen	Spec	DC	JI
pOSALseg-T (Wang et al., 2019)	–	–	–	0.901	–	–	–	–	0.974	–
U-Net with VGG16 encoder (Sarhan et al., 2010)	–	–	–	–	–	99.79	97.54	–	0.965	0.931
Mod U-Net (Yu et al., 2019)	–	–	–	0.887	0.804	–	–	–	0.973	0.949
RACE-Net (Chakravarty and Sivaswamy, 2019)	–	–	–	0.87	–	–	–	–	0.97	–
Multi-modal (Hervella et al., 2020)	–	–	–	0.902	0.822	–	–	–	0.960	0.924
U-shaped CNN (Xu et al., 2019)	–	0.915	0.998	0.892	0.823	–	0.979	0.999	0.978	0.949
RetinaGAN (Son et al., 2019)	–	–	–	–	–	–	–	–	0.967	–
GL-Net (Jiang et al., 2019)	–	–	–	0.905	–	–	–	–	0.971	–
FCN-based method (Edupuganti et al., 2018)	–	–	–	0.897	–	–	–	–	0.967	–
Entropy sampling (Zilly et al., 2017)	–	–	–	0.871	–	–	–	–	0.973	–
FC-DenseNet (Al-Bander et al., 2018)	0.994	–	–	0.828	0.711	0.996	–	–	0.949	0.904
Joint OD-OC (Chakravarty and Sivaswamy, 2017)	–	–	–	0.83	–	–	–	–	0.97	–
Modified U-Net CNN (Shankaranarayana et al., 2017)	–	–	–	0.85	0.75	–	–	–	–	–
Edge smoothing (Haleem et al., 2017)	–	–	–	0.81	–	–	–	–	0.95	–
Modified U-Net (Sevastopolsky, 2017)	–	–	–	0.850	–	–	–	–	–	–
Coupled shape (Sedai et al., 2016)	–	–	–	0.85	–	–	–	–	0.95	–
Multi-Stage (Joshi et al., 2011)	–	–	–	0.84	–	–	–	–	0.97	–
Drishti-GS Challenge (Sivaswamy et al., 2014)	–	–	–	0.79	–	–	–	–	0.96	–
SegNet (Badrinarayanan et al., 2017)	0.996	0.920	0.998	0.889	0.824	0.998	0.986	0.999	0.979	0.960
U-Net (Ronneberger et al., 2015)	0.981	0.865	0.983	0.702	0.578	0.992	0.928	0.994	0.913	0.858
SLS-Net (Proposed)	0.996	0.930	0.998	0.900	0.828	0.998	0.957	0.999	0.974	0.950
SLSR-Net (Proposed)	0.997	0.947	0.998	0.915	0.853	0.998	0.986	0.999	0.980	0.962

Table 7
The *t*-test analysis between SLSR-Net to U-Net (Ronneberger et al., 2015) and SLSR-Net to SegNet (Badrinarayanan et al., 2017) results given in Table 6.

Dataset	Between	Evaluation	Acc	Sen	Spec	DC	JI
Drishti-GS (OD)	SLSR-Net and U-Net	P-value	0.00968	0.00014	0.04345	0.00021	0.00002
		Confidence score	99.03%	99.99%	95.65%	99.98%	99.998%
Drishti-GS (OC)		P-value	0.0004	0.0016	0.0002	0.00002	0.00001
		Confidence score	99.96%	99.84%	99.98%	99.998%	99.999%
Drishti-GS (OD)	SLSR-Net and SegNet	P-value	0.28	0.08	0.32	0.04	0.05
		Confidence score	72%	92%	68%	96%	95%
Drishti-GS (OC)		P-value	0.034	0.10	0.12	0.050	0.03
		Confidence score	96.6%	90%	88%	95%	97%

Table 8
Comparisons between SLS-Net and SLSR-Net with respect to REFUGE dataset.

Methods	OC					OD				
	Acc	Sen	Spec	DC	JI	Acc	Sen	Spec	DC	JI
SLS-Net (Proposed)	0.999	0.826	0.999	0.880	0.795	0.998	0.986	0.999	0.962	0.928
SLSR-Net (Proposed 4 SC layers)	0.999	0.827	0.999	0.884	0.801	0.998	0.930	0.999	0.953	0.911
SLSR-Net (Proposed 2 SC layers)	0.999	0.947	0.999	0.895	0.815	0.998	0.969	0.999	0.965	0.933

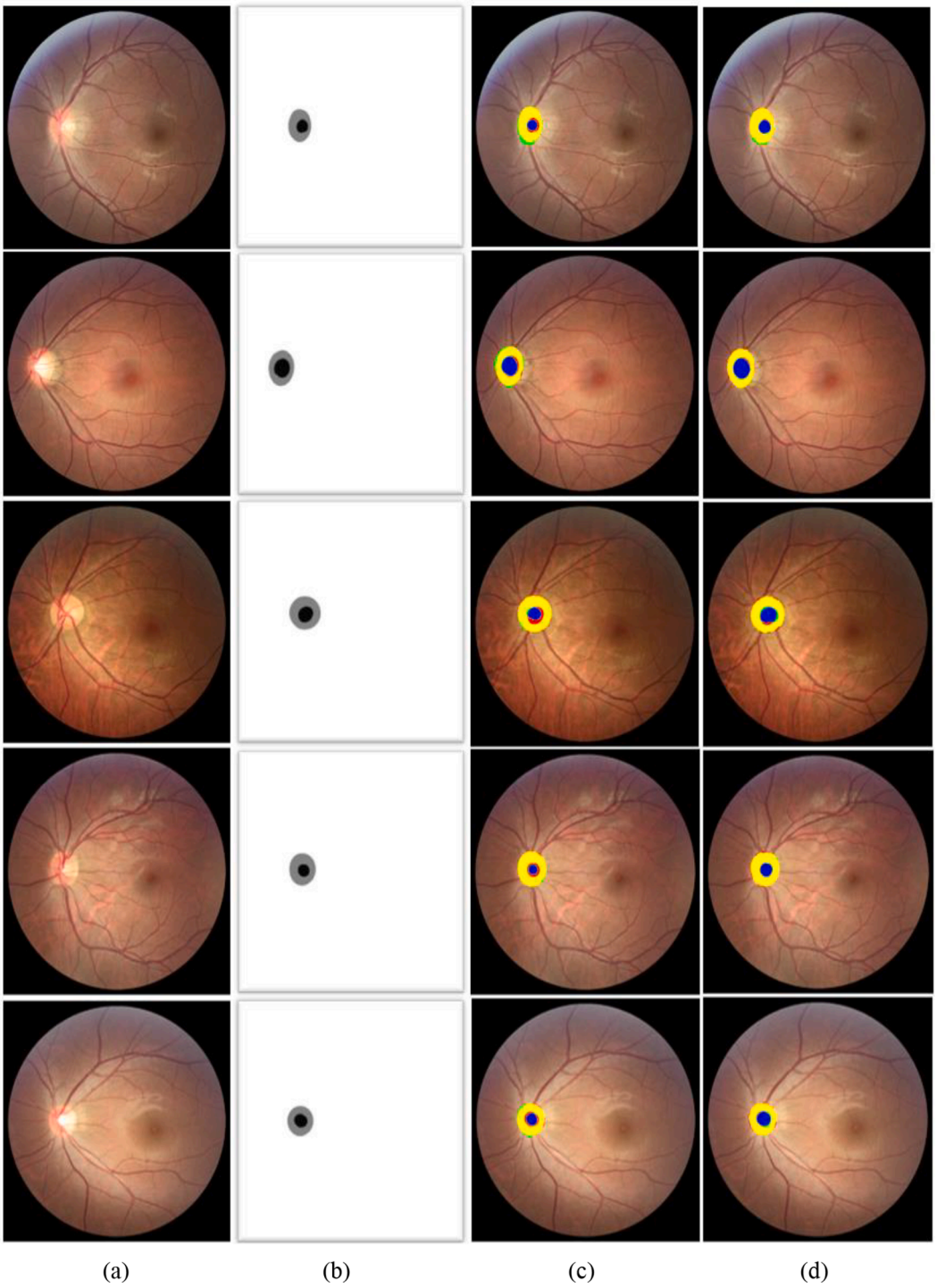


Fig. 16. Good segmentation results for OC and OD with REFUGE dataset, obtained using SLS-Net and SLSR-Net. (a) Original image, (b) ground truth image, and segmented images obtained using (c) SLS-Net and (d) SLSR-Net (for OC and OD, *tp* pixels are presented in blue and yellow, respectively, *fp* pixels are shown in green, and *fn* pixels are shown in red).

3.4. SLSR-Net-based segmentation of OC and OD

In the SLSR-Net architecture, external residual connections are developed for feature empowerment, whereas element-wise addition is performed using add layers. There is no preprocessing involved; thus, the input image is provided directly to SLSR-Net, as shown in Fig. 5.

External residual connections originate from the convolution layers of the pre-SCL unit and are terminated in add layers of the post-SCL unit. Add layers perform element-wise addition and feed the output to the corresponding convolution layer. The establishment of external residual connections minimizes the feature degradation problem. SCL is based on the depth-wise separable convolution, which filters and combines inputs using two of its distinct layers. In contrast to SC, standard convolution performs filtering and combining operations in a single step without splitting the input into layers (Howard et al., 2017). For a better insight into the process, the schematic image of the proposed SCL is shown in Fig. 6.

The last convolution layer of the pre-SCL unit provides an input

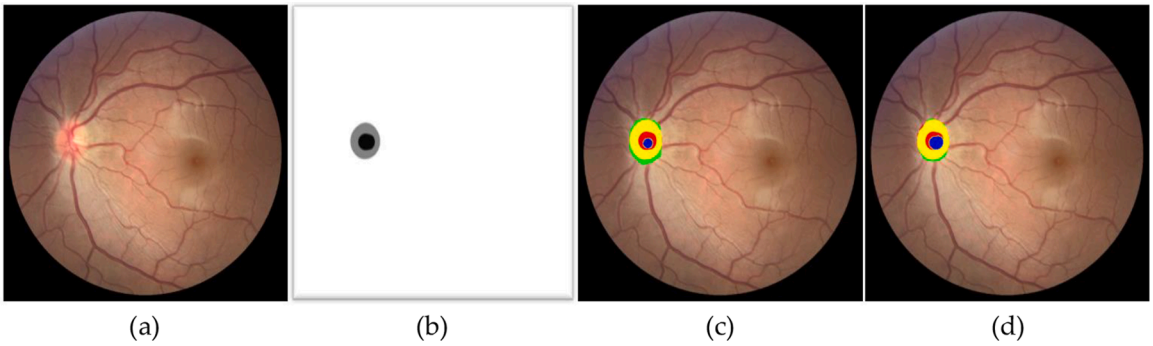


Fig. 17. Bad segmentation results for OC and OD with REFUGE dataset, obtained using SLS-Net and SLSR-Net. (a) Original image, (b) ground truth image, and segmented images obtained using (c) SLS-Net and (d) SLSR-Net (for OC and OD, *tp* pixels are presented in blue and yellow, respectively, *fp* pixels are shown in green, and *fn* pixels are shown in red).

feature map (I_{FM}) to the SCL via the ReLU layer, and the same feature map (I_{FM}) is provided to add a layer through an external residual connection. This feature map after the SCL process becomes G_{SC} and is fed to the add layer. The add layer performs element-wise addition for both of its inputs. The output of the add layer is denoted by S_{EW}, which can be computed using the following expression:

$$S_{EW} = G_{SC} + I_{FM} \tag{1}$$

Many research studies referred that standard convolution uses nine to ten times more parameters than the depth-wise separable convolution. (Howard et al., 2017). In our network, we use a maximum depth of 512, and we replace the standard convolution layers of this depth with SC layers to reduce the number of parameters. As shown in Table 3, if SC layers are not replaced with standard convolution layers, each of the expected convolution layers would have used approximately 2.359 million parameters. Each SC layer uses only 5,120 parameters. For further comparative interpretation, we can see that both expected standard convolution layers would have used 4,719,616 (2,359,808 × 2) parameters. However, both separable convolution layers actually use only 10,240 (5,120 × 2) parameters. That is, the total number of expected parameters of the proposed SLSR-Net in Fig. 5 using a conventional convolution layer is estimated to be 9.3 million. SCL helps save almost half (4.7 million) the total number of parameters, as given in Table 3, and the proposed network uses only 4.67 million trainable parameters, as shown in Table 4.

The main difference between SLS-Net and SLSR-Net lies in external residual connections. Final feature maps are fed to the pixel classification layer, which marks the pixels according to the learning. Our network produces two masks as its final output. One of the masks belongs to OC/OD, whereas the other mask belongs to the background. The overall architecture of SLSR-Net is presented in Table 4, containing the three main units of the network. The name of the layers along with the size and output feature map is also presented in Table 4. The parameters for each layer are given separately for computational analysis. We can see that in the pre-SCL unit, filters are doubled periodically while the feature map size is reduced after each stage. The SCL unit retains the same final feature map size, and its output is fed to the post-SCL unit. In this unit, the final feature size increases periodically and finally achieves the original size. The pre-SCL unit and post-SCL unit contain the same number of layers as in the reverse operation. There are three pooling layers in the pre-SCL unit and three unpooling layers in the post-SCL unit.

The architectural design of both SLS-Net and SLSR-Net is novel and not based on any other network/architecture. SCL placement in the maximum depth of the network significantly enhanced the computational efficiency of the proposed networks. Unlike famous U-Net and SegNet, the final feature map size in the proposed method is maintained large enough to minimize spatial information loss. Likewise, feature degradation problem is addressed by external residual connections in SLSR-Net. External residual connections originated from different stages of the pre-SCL unit and terminated in corresponding post-SCL stages. The effectiveness of external residual connections is evident from ablation studies. To the best of our knowledge, no other study proposed a similar design for deep learning-based automated glaucoma diagnosis.

4. Experimental results

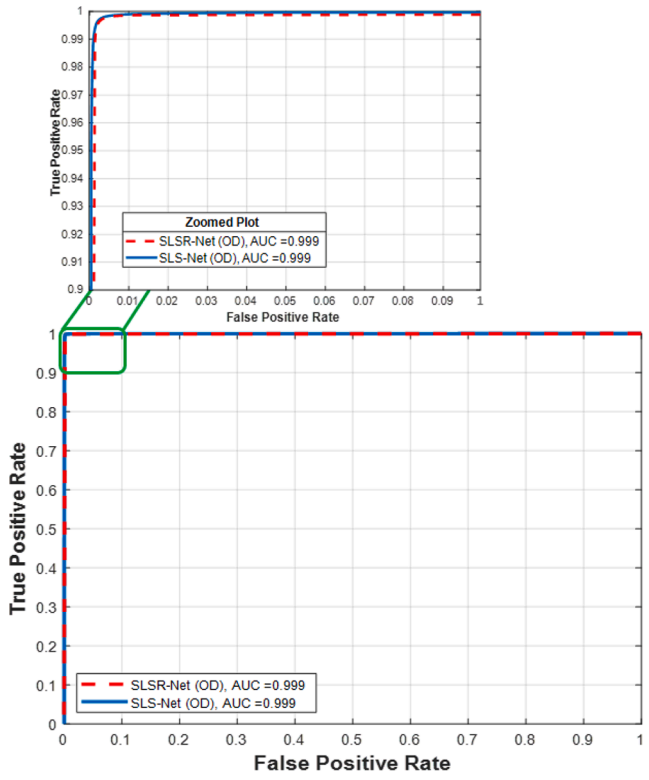
4.1. Experimental data and environment

Experiments were performed on four publicly available datasets of retinal fundus images: Drishti-GS (Sivaswamy et al., 2014), REFUGE (Orlando et al., 2020), reference image database (Rim-One-r3) (Fumero et al., 2011), and digital retinal images for optic nerve segmentation database (Drions-DB) (Carmona et al., 2008). The training and testing data were already split by the providers of both REFUGE and Drishti-GS. The Drishti-GS dataset was launched in 2014 and contains 50 images for training and 51 images for testing with their corresponding ground truth images. In the Drishti-GS dataset, the original training images do not have the same size; therefore, we resized the images to 500 × 500 pixels. Three sample images for both OD and OC in Drishti-GS are shown in Fig. 7(a) and (b), respectively. The white region in the ground truth image represents OC or OD, whereas the black region shows the background.

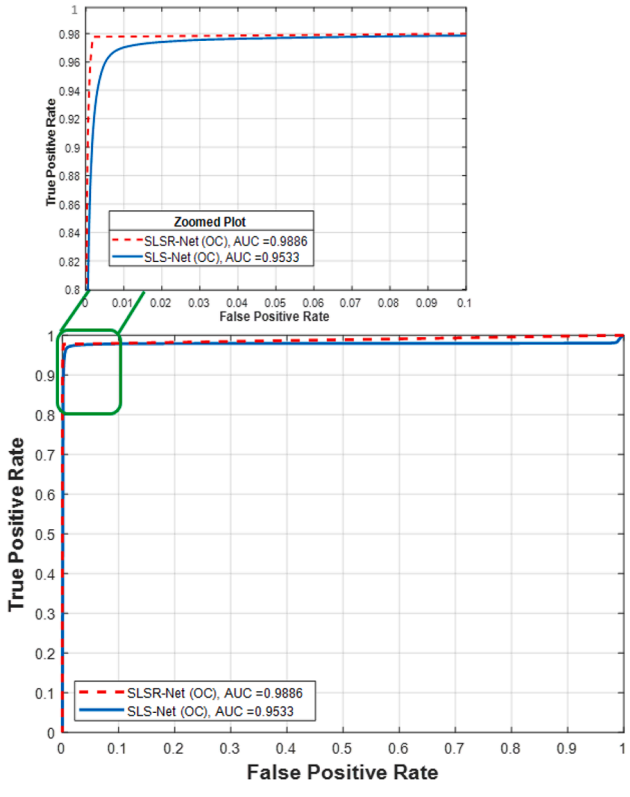
REFUGE is one of the latest datasets of retinal fundus images and was launched in 2018 by Medical Image Computing and Computer-Assisted Intervention (MICCAI). The REFUGE dataset contains 400 training images, 400 validation images, and 400 testing images. In the REFUGE dataset, the size of the original image is 2124 × 2056 pixels, and there are joint ground truth images for both OD and OC. Three sample images of the REFUGE dataset are shown in Fig. 8.

Rim-One-r3 is also one of the challenging datasets for glaucoma detection. It has total of 159 images with expert annotations and labeling. We followed the same data splitting criteria as used in (Wang et al., 2019) for fair comparisons. Moreover, we resized training images on 512 × 512 for our experiments. Sample images of the Rim-One-r3 dataset are shown in Fig. 9. To further check the effectiveness of the proposed models we selected Drions-DB for evaluation of our networks. Drions-DB has a total of 110 images with OD annotations. All images are of 600 × 400 size, and we used the same size for our experiments. In addition, we followed the same data splitting criteria as given in (Son et al., 2019) for a fair comparison. Sample images of Drions-DB are provided in Fig. 10.

The proposed models were trained and tested by using a desktop computer having 24 GB random access memory (RAM), Intel ® Core™ i7 CPU950@3.7 GHz, and NVIDIA GeForce GTX 1070 graphics processing unit (GPU) card with 1920 compute unified device architecture (CUDA) cores and a graphics memory of 8 GB (GeForce, 1070). We developed our network using MATLAB 2020b (MATLAB R, 2020) for



(a)



(b)

Fig. 18. Graphs of Sen according to 1 – Spec obtained using SLS-Net and SLSR-Net on the REFUGE dataset for the segmentation of (a) OD and (b) OC. Green box shows the selected area of graph for zoomed plot.

training and testing. Moreover, any pre-trained network was not used, and training from scratch was carried out with training data. Computer-aided diagnosis usually requires a large amount of medical data labeled by experts for appropriate learning of the network, which needs extensive medical resources and time. Consequently, using the original dataset images, we created artificial training images via data augmentation. We employed horizontal and vertical flipping, random rotation, cropping, and X-Y translation for augmentation. In addition, the different combination of the aforementioned techniques is used to produce a variety of training set for diverse learning. Data augmentation is performed for all four datasets used in this study.

4.2. Training of proposed model

Training images are resized for efficient training using limited GPU memory resources. In the training, neither a pre-trained network nor any sort of fine-tuning or weight migration is employed. Training of SLSR-Net is carried out from scratch without any initialization. Considering the computational efficiency, adaptive learning, and large data handling ability, the Adam optimizer is used for training, which helps in fast convergence. The Adam optimizer adapts to learning the subject considering its first moment (weighted sum of gradients) and second moment (moving estimate of the square gradient) and takes large optimization steps in the beginning of the training (Kingma and Ba, 2015). Training was continued for 25 epochs with a mini-batch size of 4, initial learning rate of 0.0001, and gradient threshold of 5. The training loss was computed for all pixels as per the desired and undesired classes using cross-entropy loss, and the same loss function was used in the pOSAL network (Wang et al., 2019). As can be seen from the sample images in Figs. 7-10, class pixels belonging to the background are highly dominant compared with other classes (OC/OD), which creates the class imbalance problem. We used median frequency balancing to avoid and minimize the class imbalance problem (Arsalan et al., 2020; Badrinarayanan et al., 2017). As shown in Fig. 11, the training loss finally approaches zero, whereas the training accuracy reaches 100% as the number of epochs increases, which shows that the proposed models are successfully trained with training data.

4.3. Testing of proposed method

The proposed SLS-Net and SLSR-Net were extensively tested on all four datasets for performance evaluation and comparisons. Training with resized images is performed, original images for testing are also resized before evaluation, and the output mask is subsequently resized to the original size for a fair comparison with the ground-truth images of the original size. As medical images are annotated by medical experts, resizing of testing images may eliminate important information from the image. Therefore, training is completed by using resized images, whereas the output mask of our network is resized to its original size using bilinear interpolation. Moreover, resizing using bilinear interpolation does not require learning (Arsalan et al., 2020; Badrinarayanan et al., 2017). In this way, computational and evaluation errors produced by resizing of ground truth images are avoided.

4.3.1. Evaluation metrics

The proposed SLS-Net and SLSR-Net were comprehensively evaluated under standardized benchmarking and evaluation according to the following evaluation protocols. The evaluation criteria include accuracy (Acc), sensitivity (Sen), specificity (Spec), Dice coefficient (DC), and Jaccard index (JI) (Wang et al., 2019; Yu et al., 2019). All these measures are based on true positive (*tp*), true negative (*tn*), false positive (*fp*), and false negative (*fn*) calculations.

$$\text{Acc} = \frac{tp + tn}{tp + fn + fp + tn} \tag{2}$$

Table 9
Comparisons of segmentation accuracies for OC and OD obtained using the proposed SLS-Net and SLSR-Net with those of the state-of-the-art methods on REFUGE dataset. "-" means that there is no accuracy reported.

Methods	OC					OD				
	Acc	Sen	Spec	DC	JI	Acc	Sen	Spec	DC	JI
pOSALseg-T (Wang et al., 2019)	–	–	–	0.882	–	–	–	–	0.960	–
U-Net with VGG16 encoder (Sarhan et al., 2010)	–	–	–	–	–	0.998	0.957	–	0.940	0.890
Mask-RCNN (Almubarak et al., 2020)	–	–	–	0.854	–	–	–	–	0.947	–
ET-Net (Zhang et al., 2019)	–	–	–	0.891	–	–	–	–	0.952	–
U-Net (Ronneberger et al., 2015)	–	–	–	0.854	–	–	–	–	0.930	–
Team masker (Orlando et al., 2020)	–	–	–	0.883	–	–	–	–	0.946	–
Team BUCT (Orlando et al., 2020)	–	–	–	0.872	–	–	–	–	0.952	–
Multi-modal (Hervella et al., 2020)	–	–	–	–	0.790	–	–	–	–	0.922
Conditional GAN (Liu et al., 2019)	–	–	–	–	0.800	–	–	–	–	0.884
SLS-Net (Proposed)	0.999	0.826	0.999	0.880	0.795	0.998	0.986	0.999	0.962	0.928
SLSR-Net (Proposed final)	0.999	0.947	0.999	0.895	0.815	0.998	0.969	0.999	0.965	0.933

$$\text{Sen} = \frac{tp}{tp + fn} \tag{3}$$

$$\text{Spec} = \frac{tn}{tn + fp} \tag{4}$$

$$\text{DC} = \frac{2tp}{2tp + fp + fn} \tag{5}$$

$$\text{JI} = \frac{tp}{tp + fp + fn} \tag{6}$$

4.3.2. Qualitative results of OD and OC segmentation on Drishti-GS dataset (Ablation studies)

As ablation studies, we compared the segmentation accuracies of SLS-Net and SLSR-Net. As summarized in Table 5, SLSR-Net outperforms SLS-Net. This is because SLSR-Net overcomes the feature degradation problem of SLS-Net using residual skip connections. Although Acc and Spec of both models are similar, the other accuracies of SLSR-Net are higher than those of SLS-Net. This is because of residual skip connections that ensure feature empowerment, which consequently improves segmentation performance.

Figs. 12 and 13 show the results of good segmentation for OD and OC, respectively. However, Drishti-GS is a difficult retinal fundus image dataset because its pixel intensity values are not distinctive; thus, there are some cases of bad results, as shown in Fig. 14. As shown in the lower images of Fig. 13, a large *fp* especially occurs in the case of OC segmentation owing to the low distinctive boundary of OC. Nevertheless, SLSR-Net shows higher segmentation accuracies than SLS-Net in all the cases.

In Fig. 15, we compare the graphs of Sen according to the value of 1 – Spec obtained using SLS-Net and SLSR-Net for the segmentation of OD and OC, respectively. If the curve is close to the upper-left corner, it shows a large value of area under the curve (AUC) and high accuracy. As shown in Fig. 15, SLSR-Net exhibits higher segmentation accuracies than SLS-Net in terms of AUC.

4.3.3. Comparisons with the state-of-the-art methods on Drishti-GS dataset

In this section, we compare the segmentation accuracies for OC and OD obtained using the proposed SLS-Net and SLSR-Net with those of the state-of-the-art methods. As summarized in Table 6, the proposed methods exhibit higher accuracies than the state-of-the-art methods including U-Net and SegNet.

4.3.4. Segmentation reliability and performance statistical analysis using t-test on Drishti-GS dataset

Segmentation reliability assessment in image segmentation is one of the essential parts of segmentation evaluation. The differences of reliability and segmentation performance can be statistically analyzed using p-values in *t*-test. We performed *t*-test using proposed SLSR-Net, U-Net (Ronneberger et al., 2015), and SegNet (Badrinarayanan et al., 2017). Considerably small p-values refer to a more reliable segmentation performance (Tanizaki et al., 2020). As shown in Table 7, the p-values by proposed method are sufficiently small compared with famous segmentation architectures. Therefore, *t*-test results confirm the performance difference and reliability of segmentation using proposed SLSR-Net.

4.3.5. Qualitative results of OD and OC segmentation on REFUGE dataset (Ablation studies)

As ablation studies, Table 8 refers to the comparative segmentation accuracies achieved by SLS-Net and SLSR-Net. SLSR-Net outperforms SLS-Net because it empowers the image features through residual connections. This feature empowerment leads to high segmentation accuracies. Although Acc and Spec of both models are same, most of the other accuracies of SLSR-Net are higher than those of SLS-Net. This is because feature degradation in SLS-Net was improved in SLSR-Net by using residual paths. In addition, the optimal combination of SCL layers was confirmed as shown in Table 8 as ablation studies. We found 2 SC layers in combination with BN and ReLU layers exhibiting the better segmentation performance for the proposed network than 4 SC layers in combination with BN and ReLU layers.

Fig. 16 shows the results of good segmentation for OD and OC, and SLSR-Net shows higher segmentation accuracies than SLS-Net in all cases. As shown in Figs. 7–10 and 16, the pixel intensity values and image characteristics for the REFUGE are entirely different compared with other datasets. SLSR-Net still managed an outstanding segmentation performance with a small number of trainable parameters for the REFUGE dataset as well. Fig. 17 shows the results of bad segmentation for OD and OC, and SLSR-Net still shows higher segmentation accuracies than SLS-Net in all cases. Bad segmentation occurs owing to the non-distinctive boundary for OC, and in the case of OD, it occurs due to uneven high pixel intensity values across the OD boundary.

In Fig. 18, we compare the graphs of Sen according to the value of 1 – Spec obtained using SLS-Net and SLSR-Net for the segmentation of OD and OC, respectively. If the curve is close to the upper-left corner, it shows a large value of area under curve (AUC) and high accuracy. As

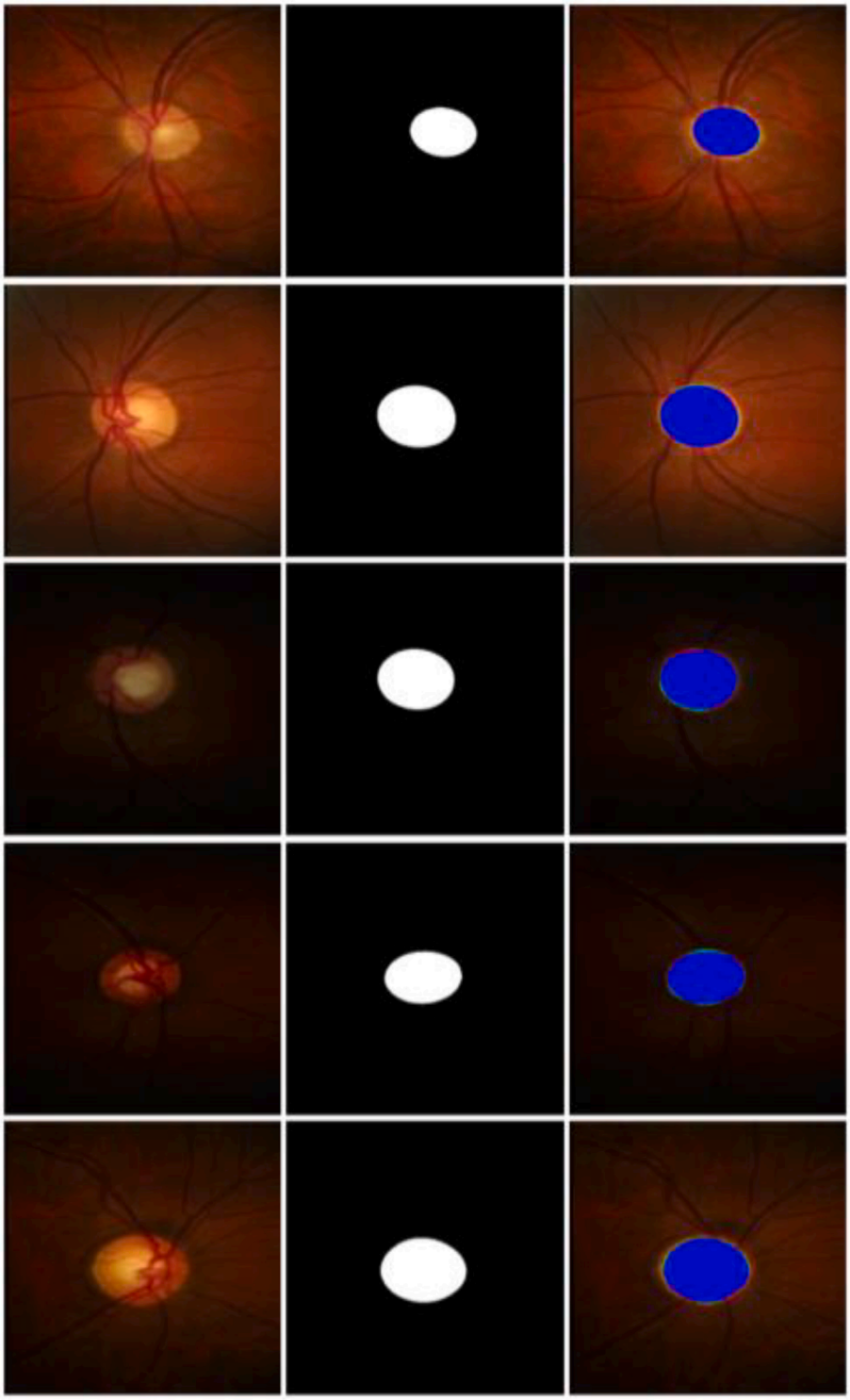


Fig. 19. Good segmentation results for OD with Rim-One-r3 dataset obtained using SLSR-Net. (a) Original image, (b) ground truth image, and (c) segmented images obtained using SLSR-Net (*tp*, *fp*, and *fn* pixels are shown in blue, green, and red, respectively).

shown in Fig. 16, SLSR-Net exhibits higher segmentation accuracies than SLS-Net for the case of OC. Although SLS-Net and SLSR-Net show similar AUC for OD as shown in Fig. 18 (a), all the other metrics such as DC and JI of SLSR-Net are higher than those of SLS-Net as shown in Table 8.

4.3.6. Comparisons with the state-of-the-art methods on REFUGE dataset

In this section, we compare the segmentation accuracies for OC and OD obtained using the proposed SLS-Net and SLSR-Net with those of the state-of-the-art methods. As summarized in Table 9, the proposed methods exhibit higher accuracies than the state-of-the-art methods.

4.3.7. Qualitative results of OD and OC segmentation using SLSR-Net on Rim-One-r3 dataset

SLSR-Net was further evaluated on challenging dataset Rim-One-r3

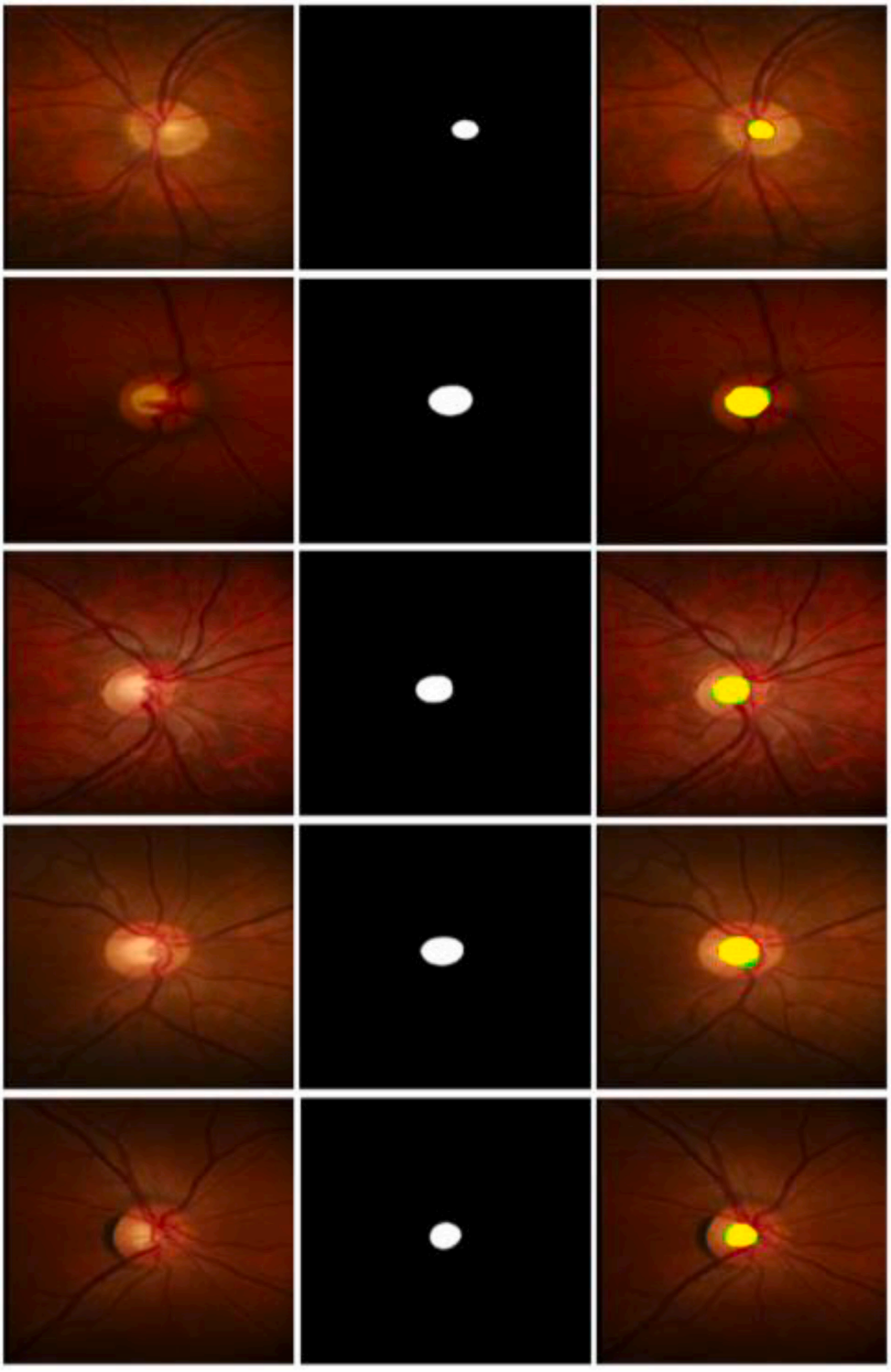


Fig. 20. Good segmentation results for OC with Rim-One-r3 dataset, obtained using SLSR-Net. (a) Original images, (b) ground truth images, and (c) segmented images (*tp*, *fp,* and *fn* pixels are shown in yellow, green, and red, respectively).

(Fumero et al., 2011) for OD and OC segmentation. OC segmentation in the Rim-One-r3 dataset is also challenging due to its indistinctive boundary. SLSR-Net still manages to exhibit good segmentation performance on both OD and OC using its efficient feature empowering architecture. Good qualitative segmentation results of OD and OC are shown in Figs. 19 and 20, respectively. As shown in Fig. 21, bad segmentation qualitative results in Rim-One-r3 are because of extensive color intensity variations and the non-distinctive boundary of OC. However, SLSR-Net still outperforms the other segmentation methods with leading computational efficiency.

4.3.8. Comparisons with the state-of-the-art methods on Rim-One-r3 dataset

In this section, we compare the segmentation accuracies for OC and

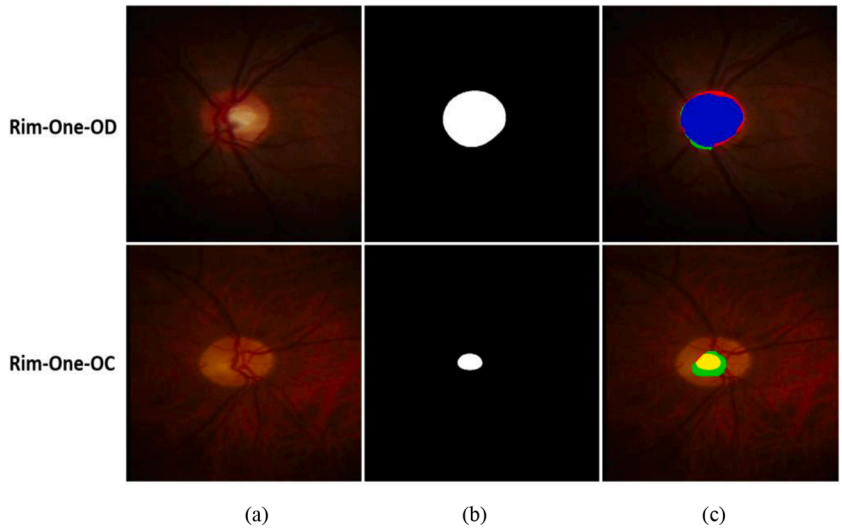


Fig. 21. Bad segmentation results for OD and OC with Rim-One-r3 dataset obtained using SLSR-Net (a) Original images, (b) ground truth images, and (c) segmented images (for OD and OC, *tp* pixels are presented in blue and yellow, respectively, whereas *fp* pixels are shown in green, and *fn* pixels are shown in red).

Table 10
Comparisons of segmentation accuracies for OC and OD were obtained using the proposed SLSR-Net with those of state-of-the-art methods on the Rim-One-r3 dataset. "-" means that there is no accuracy reported.

Methods	OC					OD				
	Acc	Sen	Spec	DC	JI	Acc	Sen	Spec	DC	JI
pOSALseg-T (Wang et al., 2019)	–	–	–	0.856	–	–	–	–	0.968	–
DRIU(Maninis et al., October 2016)	–	–	–	–	–	–	–	–	0.955	–
Modified U-Net (Sevastopolsky, 2017)	–	–	–	–	–	–	–	–	0.950	–
Segmentation using GAN (Son et al., 2019)	–	–	–	–	–	–	–	–	0.955	–
Entropy sampling (Zilly et al., 2017)	–	–	–	0.824	–	–	–	–	0.942	–
SLSR-Net (Proposed)	0.996	0.879	0.998	0.858	0.757	0.997	0.968	0.999	0.969	0.940

OD obtained using the proposed SLSR-Net with those of state-of-the-art methods. Quantitative results given in Table 10 show the promising segmentation accuracies by SLSR-Net compared with other methods for OD and OC.

4.3.9. Qualitative results of OD and OC segmentation using SLSR-Net on Drions-DB dataset

To further check the effectiveness of the proposed method, we evaluated SLSR-Net on Drions-DB. SLSR-Net showed excellent segmentation performance and outperformed the state-of-the-art methods on Drions-DB as well. Good segmentation qualitative results are shown in Fig. 22 whereas bad segmentation results are shown in Fig. 23. Bad segmentation results are because of the variations around OD boundaries which make the OD boundary segmentation difficult. However, SLSR-Net is capable of variations handling and manages to deliver outperforming results.

4.3.10. Comparisons with the state-of-the-art methods on Drions-DB

We evaluated our proposed final network on Drions-DB. SLSR-Net achieved state-of-the-art segmentation performance using its effective network design. SLSR-Net comparison of segmentation accuracies with state-of-the-art methods for Drions-DB is given in Table 11, which confirms that proposed SLSR-Net exhibits better segmentation performance for OD than the state-of-the-art methods, even better than that by human annotator.

4.3.11. The significances of SCL module and comparative computational efficiency

The maximum depth of CNN usually bears the smallest feature map size with the largest number of channels, and it consumes most of the trainable parameters. In our case, SCL is employed at a maximum depth of the network where feature map size is the smallest with a maximum number of channels (512). In SCL, two SC layers in combination with BN and ReLU layers are employed. As presented in Table 3, SCL saves approximately 4.7 million trainable parameters and boosts the computational efficiency of the network. Small feature map size can lead to information loss especially for the minor features of the images. Unlike conventional networks in maximum depth, SCL does not employ any operation of reducing feature map size (pooling) and maintains the feature map size of $62 \times 62 \times 512$ (Table 4), which can help to retain the minor features of OC and OD. Hence, SCL unit enables the network to deliver optimum segmentation performance with enhanced computational efficiency, which are shown in Table 12. Moreover, the optimal combination of SCL layers was confirmed as shown in Table 8 as ablation studies. We found 2 SC layers in combination with BN and ReLU layers exhibiting the better segmentation performance for the proposed network than 4 SC layers in combination with BN and ReLU layers.

The computational efficiency of a network is one of the critical problems directly related to the memory requirements and robustness of the network. Even with the higher accuracies of segmentation obtained using our model, in comparison with those of the state-of-the-art models, the complete framework of SLSR-Net uses approximately 4.67 M trainable parameters, which confirms the superior computational strength of

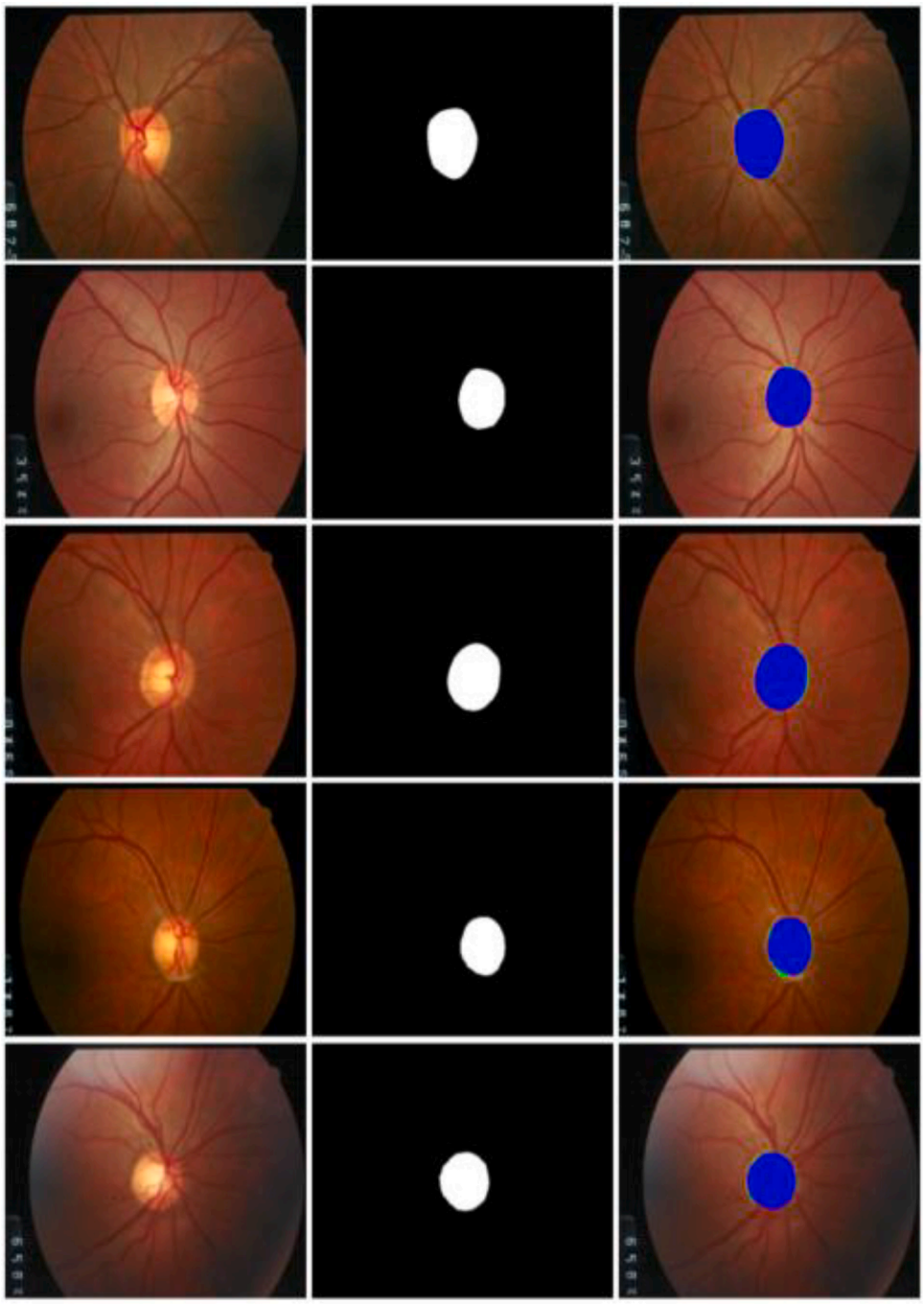


Fig. 22. Good segmentation results for OD with Drions-DB were obtained using SLSR-Net. (a) Original image, (b) ground truth image, and (c) segmented images obtained using SLSR-Net (*tp*, *fp*, and *fn* pixels are shown in blue, green, and red, respectively).

the network over the state-of-the-art models, as summarized in Table 12. The network size of the SLSR-Net model is only 16.8 MB.

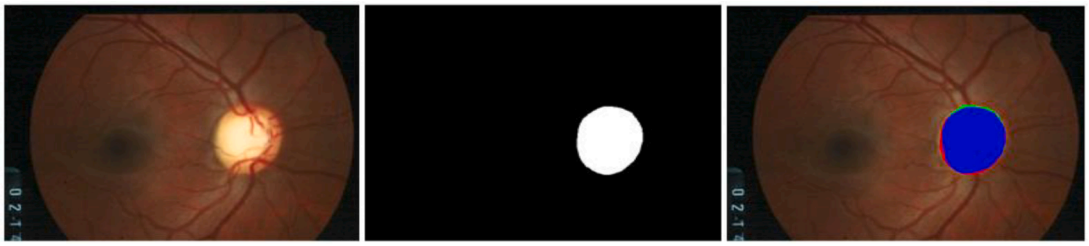
4.3.12. Assisting in automatic glaucoma and other diseases screening

Based on the segmentation results of OD and OC, vertical CDR can be measured by taking the ratio of OC vertical diameter (VD_{cup}) to OD vertical diameter (VD_{disc}), as expressed by below Equation.

$$CDR(vertical) = \frac{VD_{cup}}{VD_{disc}} \tag{7}$$

A high vertical *CDR* value indicates a high probability of occurrence of glaucoma, whereas a low vertical *CDR* represents a low probability of occurrence of glaucoma (Fu et al., 2018a). As shown in Fig. 24, the CDR of ground-truth (CDR_g) annotated by an expert is given as 0.85, and the CDR predicted using the proposed method (CDR_p) is 0.86, which is very close to CDR_g. This confirms the high segmentation performance of our network. In addition, OD and OC segmentation can also be used in aiding the diagnosis of some other diseases such as Alzheimer's disease (AD) and poor cognitive functioning in postmenopausal women (Malik et al., 2020; Vajaranant et al., 2019). Patients with AD have a large cup area and CDR with a thin rim area (Vajaranant et al., 2019).

Automatic glaucoma detection is not limited to segmentation only.



(a) (b) (c)

Fig. 23. Bad segmentation results for OD with Drions-DB obtained using SLSR-Net (a) Original image, (b) ground truth image, and (c) segmented image (*tp, fp, and fn* pixels are presented in blue, green, and red, respectively).

Table 11
Comparisons of segmentation accuracies for OD were obtained using the proposed SLSR-Net with those of state-of-the-art methods on Drions-DB. "-" means that there is no accuracy reported.

Method	Acc	Sen	Spec	DC	JI
DRIU (Maninis et al., October 2016)	–	–	–	0.967	–
U-Net (Ronneberger et al., 2015)	–	–	–	0.968	–
RetinaGAN (Son et al., 2019)	–	–	–	0.968	–
*Human annotator	–	–	–	0.963	–
SLSR-Net (Proposed)	0.997	0.951	0.999	0.969	0.941

* Human annotator result is taken from (Son et al., 2019).

Table 12
Comparisons of the number of trainable parameters and accuracies of the proposed model and state-of-the-art methods with respect to REFUGE database.

Methods	Parameters	DC (OC)	DC (OD)	Network size
pOSAL(Xception)* (Wang et al., 2019)	41.3 M	0.885	0.953	–
pOSAL (MobileNetV2)* (Wang et al., 2019)	5.8 M	0.885	0.956	–
U-Net with VGG16 encoder (Sarhan et al., 2010)	16.8 M	–	0.94	–
Mask-RCNN (Almubarak et al., 2020)	127 M	0.854	0.947	–
Team masker (Orlando et al., 2020)	1,224 M	0.883	0.946	–
GL-Net (Jiang et al., 2019)	30.8 M	–	–	–
Proposed SLSR-Net	**4.67 M**	0.895	0.965	16.8 MB

*Xception (Chen et al., 2018) and MobileNetV2 (Sandler et al., 2018) were used as the backbone networks in pOSAL.

Classification is among deep learning-based non-segmentation methods which can be used for glaucoma screening. Therefore, we also performed experiments using Resnet-18 and Resnet-50 on Rim-One-r3 and REFUGE for glaucoma classification. We used the same data splits (for training, validation, and testing) with same image size and training protocols for the fair comparisons of both segmentation and classification experiments in Tables 13 and 14. Glaucoma screening results with both segmentation and classification on Rim-One-r3 and REFUGE are provided in Tables 13 and 14, respectively. Experimental results clearly show that the segmentation-based (proposed SLSR-Net) method outperforms the non-segmentation methods for glaucoma detection. Because segmentation-based glaucoma screening is not limited to only vertical CDR computation, therefore previous researches on non-CDR-based glaucoma screening are also discussed in upcoming subsection.

4.3.13. Aiding the glaucoma screening (Non-CDR approaches).

Segmentation-based accurate glaucoma screening is not limited to CDR computation only. This is fact that segmentation methods widely employ CDR (Especially vertical CDR) assessment in previous researches for glaucoma screening (Orlando et al., 2020; Soh et al., 2020; Wang et al., 2019; Yu et al., 2019; Fu et al., 2018a), and vertical CDR provides valuable insight. However, vertical CDR-based glaucoma screening is not always correct. Therefore, we also present Non-CDR techniques in this subsection, which can be employed in parallel with existing techniques using OD and OC segmentation to aid the glaucoma screening.

Both OC and OD size assessment has a valuable impact on glaucoma detection (Sekhar et al., 2008). Similarly, increased OC area is also considered an important biomarker for glaucoma detection (Thakur and Juneja, 2019). Area-cup-to-disc-ratio (ACDR) is the ratio between the areas of OC and OD (Dasgupta et al., 2021). ACDR computation provides the ratio of corresponding class areas irrespective of OD and OC shapes or patterns. Semantic segmentation provides pixel-wise predictions. Therefore, accurate OD and OC predictions enable the network to provide reliable area predictions for OD and OC, which consequently helps in the accurate diagnosis of glaucoma (Dasgupta et al., 2021). Likewise, the disc damage likelihood scale (DDLS) delivers the likelihood of disc damage which refers to disease detection with its severity analysis (Thakur and Juneja, 2018). The area between OD and OC is known as the neural rim, and the ratio between a minimum width of the neural rim and disc diameter provides DDLS. In segmentation, boundary prediction pixels are presented in segmented images and neural rim area can be conveniently extracted by subtracting OD boundary from OC boundary. Notching is also one of the important methods to differentiate glaucoma and non-glaucoma cases (Healey and Mitchell, 2015). Notching is about the decrease in the width of the neural rim (Healey and Mitchell, 2015), and it can also be assessed by the inferior superior nasal and temporal (ISNT) rule. ISNT rule itself is used for aiding the glaucoma clinical diagnosis (Pathan et al., 2021). ISNT rule states that the width of the neural rim area should be widest to thinnest for inferior, superior, nasal, and temporal regions in order (Pathan et al., 2021). Cases violating this rule are discriminated against as glaucoma suspects. Subsequently, accurate area and size of OD and OC, the width of the neural rim, and notching are associated with segmentation performance. Therefore, accurate segmentation of OD and OC can help in the computation of the aforementioned measures and consequently aid the glaucoma diagnosis.

5. Conclusion

In this research, we proposed deep learning-based novel models, SLS-Net and SLSR-Net, to segment OD and OC for glaucoma screening. SLSR-Net is the final proposed model that maintains a large final feature map size throughout the network to avoid spatial information loss of even minor features. Memory requirements are one of the major limitations of computer-aided diagnosis. An SCL unit in our model minimizes this problem and significantly increases the computational efficiency of the network. External residual connections settle the feature degradation problem by empowering the features. Training and testing of the network is carried out without any preprocessing or postprocessing overhead. We extensively evaluated the proposed model on four

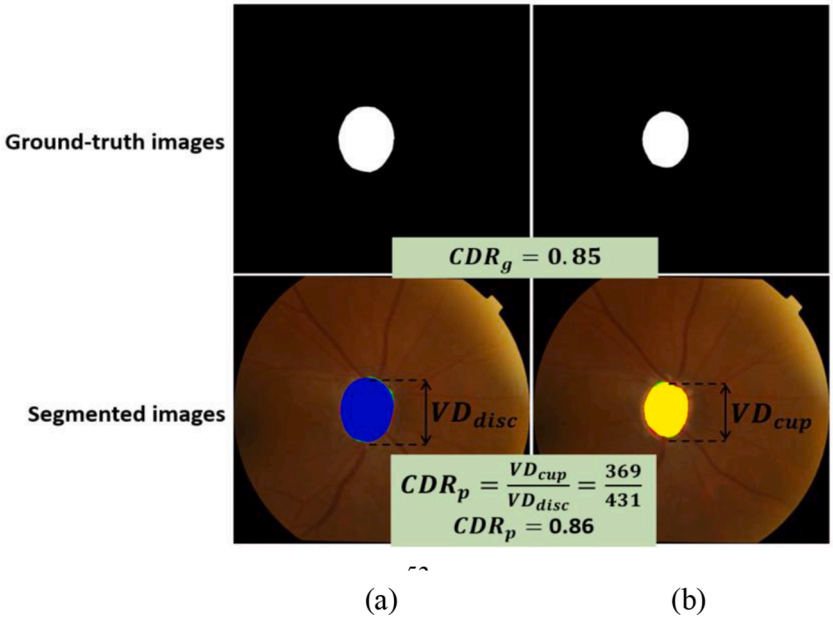


Fig. 24. Detection of glaucoma: (a) OD image and (b) OC image (VD_{disc} represents the vertical diameter of the OD, whereas VD_{cup} represents the vertical diameter of the OC. CDR_p represents the predicted ratio of OC to OD, and CDR_g represents the ground-truth ratio of OC to OD).

Table 13
Comparison of glaucoma detection methods using segmentation (SLSR-Net) and non-segmentation (Resnet-50 and Resent-18) architectures on Rim-One-r3 dataset.

Methods	Acc	Spec	Sen	F1
Classification using Resnet-18	0.85	0.978	0.428	0.596
Classification using Resnet-50	0.783	0.913	0.357	0.513
SLSR-Net (segmentation-based glaucoma classification)	0.90	0.934	0.785	0.853

Table 14
Comparison of glaucoma detection methods using segmentation (SLSR-Net) and non-segmentation (Resnet-50 and Resent-18) architectures on REFUGE dataset.

Methods	Acc	Spec	Sen	F1
Classification using Resnet-18	0.887	0.913	0.65	0.759
Classification using Resnet-50	0.917	0.944	0.675	0.787
SLSR-Net (segmentation-based glaucoma classification)	0.925	0.952	0.675	0.790

publicly available datasets and achieved state-of-the-art performances compared with existing methods. There is a trade-off between segmentation performance and computational efficiency. Therefore, methods that achieve good segmentation performance usually use many parameters, which makes the network computationally expensive. In our network, good results are achieved by using only 4,666,950 trainable parameters, which confirms the outstanding computational efficiency of the network compared to the state-of-the-art methods.

Although the proposed framework achieved state-of-the-art results, some limitations are worth considering for future work. The OC boundary is usually indistinctive, as shown in Figs. 13, 16 and 20; consequently, the segmentation performance of OC falls behind the OD performance. OC segmentation has been challenging for all researchers, and additional preprocessing may be considered for further improving OC segmentation results. In addition, we would further optimize our network for more robust glaucoma screening. We would also utilize this network for other retinal diseases and medical diagnosis application tasks such as diagnosis of brain tumor and brain stroke as well as liver and kidney-stone segmentation.

CRediT authorship contribution statement

Adnan Haider: Methodology, Writing – original draft. **Muhammad Arsalan:** Conceptualization. **Min Beom Lee:** Data curation. **Muhammad Owais:** Formal analysis. **Tahir Mahmood:** Investigation. **Haseeb Sultan:** Software. **Kang Ryoung Park:** Supervision, Writing – review & editing.

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.

Acknowledgments

This research was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) through the Basic Science Research Program (NRF-2021R1F1A1045587), in part by the NRF funded by the MSIT through the Basic Science Research Program (NRF-2022R1F1A1064291), and in part by the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2020-0-01789) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

References

Ahn, J. M., Kim, S., Ahn, K.-S., Cho, S.-H., Lee, K. B., & Kim, U. S. (2018). A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS One, 13*, Article e0207982.

Al-Bander, B., Williams, B. M., Al-Nuaimy, W., Al-Taee, M. A., Pratt, H., & Zheng, Y. (2018). Dense fully convolutional segmentation of the optic disc and cup in color fundus for glaucoma diagnosis. *Symmetry-Basel, 10*, 87.

Almubarak, H., Bazi, Y., & Alajlan, N. (2020). Two-stage mask-RCNN approach for detecting and segmenting the optic nerve head, optic disc, and optic cup in fundus images. *Applied Sciences-Basel, 10*, 3833.

Arsalan, M., Baek, N. R., Owais, M., Mahmood, T., & Park, K. R. (2020). Deep learning-based detection of pigment signs for analysis and diagnosis of retinitis pigmentosa. *Sensors, 20*, 3454.

Arsalan, M., Owais, M., Mahmood, T., Cho, S. W., & Park, K. R. (2019). Aiding the diagnosis of diabetic and hypertensive retinopathy using artificial intelligence-based semantic segmentation. *Journal of Clinical Medicine, 8*, 1446.

Arsalan, M., Owais, M., Mahmood, T., Choi, J., & Park, K. R. (2020). Artificial intelligence-based diagnosis of cardiac and related diseases. *Journal of Clinical Medicine, 9*, 871.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*, 2481–2495.

Baum, J., Chaturvedi, N., Netland, P. A., & Dreyer, E. B. (1995). Assessment of intraocular pressure by palpation. *Ophthalmology, 119*, 650–651.

Carmona, E. J., Rincón, M., García-Feijoó, J., & Martínez-de-la-Casa, J. M. (2008). Identification of the optic nerve head with genetic algorithms. *Artificial Intelligence in Medicine, 43*, 243–259.

Chakravarty, A., & Sivaswamy, J. (2017). Joint optic disc and cup boundary extraction from monocular fundus images. *Computer Methods and Programs in Biomedicine, 147*, 51–61.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *In Proceedings of European Conference on Computer Vision* (pp. 801–818).

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251–1258).

Dasgupta, S., Mukherjee, R., Dutta, K., Sen, A., 2021. Deep Learning based Framework for Automatic Diagnosis of Glaucoma based on analysis of Focal Notching in the Optic Nerve Head. ArXiv211205748 Cs Eess.

Diaz-Pinto, A., Morales, S., Naranjo, V., Köhler, T., Mossi, J. M., & Navea, A. (2019). CNNs for automatic glaucoma assessment using fundus images: An extensive validation. *BioMedical Engineering OnLine, 18*, 29.

Dongguk SLS-Net and SLSR-Net. Available online: https://dm.dongguk.edu/link.html (accessed on June 1, 2020).

Edupuganti, V. G., Chawla, A., & Kale, A. (2018). Automatic optic disk and cup segmentation of fundus images using deep learning. In *In Proceedings of 25th IEEE International Conference on Image Processing* (pp. 2227–2231).

Fu, H., Cheng, J., Xu, Y., Wong, D. W. K., Liu, J., & Cao, X. (2018a). Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transactions on Medical Imaging, 37*, 1597–1605.

Fu, H., Cheng, J., Xu, Y., Zhang, C., Wong, D. W. K., Liu, J., & Cao, X. (2018b). Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Transactions on Medical Imaging, 37*, 2493–2501.

Fumero, F., Alayon, S., Sanchez, J. L., Sigut, J., & Gonzalez-Hernandez, M. (2011). RIM-ONE-R3: An open retinal image database for optic nerve evaluation. In *In Proceedings of 24th International Symposium on Computer-Based Medical Systems* (pp. 1–6).

Gao, F., Yoon, H., Wu, T., & Chu, X. (2020). A feature transfer enabled multi-task deep learning model on medical imaging. *Expert Systems with Applications, 143*, Article 112957.

GeForce GTX 1070. Available online: https://www.nvidia.com/en-gb/geforce/products/10series/geforce-gtx-1070/ (accessed on June 1, 2020).

Haleem, M.S., Han, L., Hemert, J. van, Li, B., Fleming, A., Pasquale, L.R., Song, B.J., 2017. Haleem, M.S., Han, L., Hemert, J. van, Li, B., Fleming, A., Pasquale, L.R., Song, B.J. (2017). A novel adaptive deformable model for automated optic disc and cup segmentation to aid glaucoma diagnosis. *Journal of Medical Systems,* 42, 20.

Healey, P. R., & Mitchell, P. (2015). Presence of an optic disc notch and glaucoma. *Journal of Glaucoma, 24*, 262–266.

Hervella, Á.S., Ramos, L., Rouco, J., Novo, J., Ortega, M. (2020). Multi-modal self-supervised pre-training for joint optic disc and cup segmentation in eye fundus images. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Virtual Barcelona, Spain, 4-8 May (pp. 961–965).

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv*:1704.04861 (2017).

Jiang, Y., Tan, N., & Peng, T. (2019). Optic disc and cup segmentation based on deep convolutional generative adversarial networks. *IEEE Access, 7*, 64483–64493.

Joshi, G. D., Sivaswamy, J., & Krishnadas, S. R. (2011). Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE Transactions on Medical Imaging, 30*, 1192–1205.

Kingma, D. P., & Ba, J. L. (2015). In *Adam: A method for stochastic optimization* (pp. 1–15). San Diego, CA, USA: 7–9 May.

Liu, S., Hong, J., Lu, X., Jia, X., Lin, Z., Zhou, Y., … Zhang, H. (2019). Joint optic disc and cup segmentation using semi-supervised conditional GANs. *Computers in Biology and Medicine, 115*, Article 103485.

Mahmood, T., Arsalan, M., Owais, M., Lee, M. B., & Park, K. R. (2020). Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *Journal of Clinical Medicine, 9*, 749.

Malik, F. H., Batool, F., Rubab, A., Chaudhary, N. A., Khan, K. B., & Qureshi, M. A. (2020). Retinal disorder as a biomarker for detection of human diseases. In *In Proceedings of IEEE 23rd International Conference on Multitopic* (pp. 1–6).

Meier, J., Bock, R., Michelson, G., Nyúl, L. G., & Hornegger, J. (2007). Effects of preprocessing eye fundus images on appearance based glaucoma classification. In *In Proceedings of International Conference on Computer Analysis of Images and Patterns* (pp. 165–172).

Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., & Van Gool, L. (October 2016). *Proceedings of Medical Image Computing and Computer-Assisted Intervention, 17–21*, 140–148.

Mittapalli, P. S., & Kande, G. B. (2016). Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma. *Biomedical Signal Processing and Control, 24*, 34–46.

MATLAB R2020b. Available online: https://www.mathworks.com/products/matlab.html (Accessed on June 1, 2020).

Nguyen, D. T., Lee, M. B., Pham, T. D., Batchuluun, G., Arsalan, M., & Park, K. R. (2020). Enhanced image-based endoscopic pathological site classification using an ensemble of deep learning models. *Sensors, 20*, 5982.

Orlando, J. I., Fu, H., Barbosa Breda, J., van Keer, K., Bathula, D. R., Diaz-Pinto, A., … Bogunović, H. (2020). REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis, 59*, Article 101570.

Orlando, J. I., Prokofyeva, E., del Fresno, M., & Blaschko, M. B. (2017). Convolutional neural network transfer for automated glaucoma identification. In *In Proceedings of 12th International Symposium on Medical Information Processing and Analysis* (p. 101600U).

Owais, M., Arsalan, M., Choi, J., Mahmood, T., & Park, K. R. (2019). Artificial intelligence-based classification of multiple gastrointestinal diseases using endoscopy videos for clinical diagnosis. *Journal of Clinical Medicine, 8*, 986.

Owais, M., Arsalan, M., Choi, J., & Park, K. R. (2019). Effective diagnosis and treatment through content-based medical image retrieval by using artificial intelligence. *Journal of Clinical Medicine, 8*, 462.

Pathan, S., Kumar, P., Pai, R. M., & Bhandary, S. V. (2021). Automated segmentation and classifcation of retinal features for glaucoma diagnosis. *Biomedical Signal Processing and control, 63*, Article 102244.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241).

Roychowdhury, S., Koozekanani, D. D., Kuchinka, S. N., & Parhi, K. K. (2016). Optic disc boundary and vessel origin segmentation of fundus images. *IEEE Journal of Biomedical and Health Informatics, 20*, 1562–1574.

Saeed, E., Szymkowski, M., Saeed, K., & Mariak, Z. (2019). An approach to automatic hard exudate detection in retina color images by a telemedicine system based on the d-eye sensor and image processing algorithms. *Sensors, 19*, 695.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, 18-22 June (pp. 4510–4520).

Sarhan, A., Al-KhazÁly, A., Gorner, A., Swift, A., Rokne, J., Alhajj, R., Crichton, A.: Utilizing transfer learning and a customized loss function for optic disc segmentation from Retinal Images. *arXiv:*2010.00583 (2020).

Sarkar, D., & Das, S. (2017). Automated glaucoma detection of medical image using biogeography based optimization. In *In Proceedings of the 3rd International Conference on Opto-Electronics and Applied Optics* (pp. 381–388).

Sedai, S., Roy, P. K., Mahapatra, D., & Garnavi, R. (2016). Segmentation of optic disc and optic cup in retinal fundus images using shape regression. In *In Proceedings of 38th Annual International Conference of the IEEE Engineering on Medicine and Biology Society* (pp. 3260–3264).

Sekhar, S., Al-Nuaimy, W., & Nandi, A. K. (2008). Automated localisation of optic disk and fovea in retinal fundus images. In *In Proceedings of 16th European Signal Processing Conference* (pp. 1–5).

Septiarini, A., Harjoko, A., Pulungan, R., & Ekantini, R. (2018). Automated detection of retinal nerve fiber layer by texture-based analysis for glaucoma evaluation. *Healthcare Informatics Research, 24*, 335–345.

Sevastopolsky, A. (2017). Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recognition and Image Analysis, 27*, 618–624.

Shankaranarayana, S. M., Ram, K., Mitra, K., & Sivaprakasam, M. (2017). Joint optic disc and cup segmentation using fully convolutional and adversarial networks. In *In Proceedings of International Workshop on Fetal and Infant Image Analysis* (pp. 168–176).

Sivaswamy, J., Krishnadas, S.R., Datt Joshi, G., Jain, M., Syed Tabish, A.U. (2014). Drishti-GS: Retinal image dataset for optic nerve head segmentation, In Proceedings of IEEE 11th International Symposium on Biomedical Imaging, Beijing, China, 29 April-2 May (pp. 53–56).

Soh, Z. D., Chee, M. L., Thakur, S., Tham, Y. C., Tao, Y., Lim, Z. W., … Cheng, C.-Y. (2020). Asian-specific vertical cup-to-disc ratio cut-off for glaucoma screening: An evidence-based recommendation from a multi-ethnic Asian population. *Clinical and Experimental Ophthalmology, 48*, 1210–1218.

Son, J., Park, S. J., & Jung, K.-H. (2019). Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. *Journal of Digital Imaging, 32*, 499–512.

Srivastava, R., Cheng, J., Wong, D.W.K., Liu, J. (2015). Using deep learning for robustness to parapapillary atrophy in optic disc segmentation, In Proceedings of IEEE 12th international symposium on Biomedical Imaging, New York Marriott, NY, USA, 16-19 April (pp. 768–771).

Tan, N.-M., Xu, Y., Goh, W. B., & Liu, J. (2015). Robust multi-scale superpixel classification for optic cup localization. *Computerized Medical Imaging and Graphics., 40*, 182–193.

Tanizaki, K., Hashimoto, N., Inatsu, Y., Hontani, H., & Takeuchi, I. (2020). Computing Valid P-Values for Image Segmentation by Selective Inference. In *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (pp. 9550–9559).

Thakur, N., & Juneja, M. (2019). Optic disc and optic cup segmentation from retinal images using hybrid approach. *Expert System with Applications, 127*, 308–322.

Thakur, N., & Juneja, M. (2018). Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma. *Biomedical Signal Processing and Control, 42*, 162–189.

Tham, Y.-C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., & Cheng, C.-Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *American Journal of Ophthalmology, 121*, 2081–2090.

Thompson, A. C., Jammal, A. A., & Medeiros, F. A. (2020). A review of deep Learning for screening, diagnosis, and detection of glaucoma progression. *Translational vision science & technology, 9*, 42.

Vajaranant, T. S., Hallak, J., Espeland, M. A., Pasquale, L. R., Klein, B. E., Meuer, S. M., … Maki, P. M. (2019). An association between large optic nerve cupping and cognitive function. *American Journal of Ophthalmology, 206*, 40–47.

Wang, S., Yu, L., Yang, X., Fu, C.-W., & Heng, P.-A. (2019). Patch-based output space adversarial learning for joint optic disc and cup segmentation. *IEEE Transactions on Medical Imaging, 38*, 2485–2495.

Xu, Y., Duan, L., Lin, S., Chen, X., Wong, D. W. K., Wong, T. Y., & Liu, J. (2014). Optic cup segmentation for glaucoma detection using low-rank superpixel representation. In *In Proceedings of Medical Image Computing and Computer-Assisted Intervention* (pp. 788–795).

Xu, Y., Lu, S., Li, H., & Li, R. (2019). Mixed maximum loss design for optic disc and optic cup segmentation with deep learning from imbalanced samples. *Sensors, 19*, 4401.

Yu, F., Koltun, V., & Funkhouser, T. (2017). Dilated residual networks. In *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (pp. 472–480).

Yu, S., Xiao, D., Frost, S., & Kanagasingam, Y. (2019). Robust optic disc and cup segmentation with deep learning for glaucoma detection. *Computerized Medical Imaging and Graphics, 74*, 61–71.

Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., Shao, L. (2019). ET-Net: A generic edge-attention guidance network for medical image segmentation. In Proceedings of Medical Image Computing and Computer Assisted Intervention, Shenzhen, China, 10 October (pp. 442–450).

Zhou, W., Wu, C., Chen, D., Yi, Y., & Du, W. (2017). Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method. *IEEE Access, 5*, 2563–2572.

Zilly, J., Buhmann, J. M., & Mahapatra, D. (2017). Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics, 55*, 28–41.