Computers, Materials & Continua
DOI:10.32604/cmc.2020.013249
Article

Tech Science Press

Artificial Intelligence-Based Semantic Segmentation of Ocular Regions for Biometrics and Healthcare Applications

Rizwan Ali Naqvi[1], Dildar Hussain[2] and Woong-Kee Loh[3,*]

[1]Department of Unmanned Vehicle Engineering, Sejong University, Seoul, 05006, Korea
[2]School of Computational Science, Korea Institute for Advanced Study (KIAS), Seoul, 02455, Korea
[3]Department of Software, Gachon University, Seongnam, 13120, Korea
[*]Corresponding Author: Woong-Kee Loh. Email: wkloh2@gachon.ac.kr
Received: 30 July 2020; Accepted: 11 September 2020

Abstract: Multiple ocular region segmentation plays an important role in different applications such as biometrics, liveness detection, healthcare, and gaze estimation. Typically, segmentation techniques focus on a single region of the eye at a time. Despite the number of obvious advantages, very limited research has focused on multiple regions of the eye. Similarly, accurate segmentation of multiple eye regions is necessary in challenging scenarios involving blur, ghost effects low resolution, off-angles, and unusual glints. Currently, the available segmentation methods cannot address these constraints. In this paper, to address the accurate segmentation of multiple eye regions in unconstrainted scenarios, a lightweight outer residual encoder-decoder network suitable for various sensor images is proposed. The proposed method can determine the true boundaries of the eye regions from inferior-quality images using the high-frequency information flow from the outer residual encoder-decoder deep convolutional neural network (called ORED-Net). Moreover, the proposed ORED-Net model does not improve the performance based on the complexity, number of parameters or network depth. The proposed network is considerably lighter than previous state-of-the-art models. Comprehensive experiments were performed, and optimal performance was achieved using SBVPI and UBIRIS.v2 datasets containing images of the eye region. The simulation results obtained using the proposed ORED-Net, with the mean intersection over union score (mIoU) of 89.25 and 85.12 on the challenging SBVPI and UBIRIS.v2 datasets, respectively.

Keywords: Semantic segmentation; ocular regions; biometric for healthcare; sensors; deep learning

1 Introduction

In the last few decades, researchers have made significant contributions to biometrics, liveness detection, and gaze estimation systems that rely on traits such as the iris, sclera, pupil, or other periocular regions [1]. Interest in these traits is increasing each day because of the considerable importance of ocular region applications and their significant market potential. Biometric technology has become a vital part of our

daily life, as unlike conventional methods, these approaches do not require an individual to memorize or carry any information, such as pins, passwords, or IDs [2]. Iris segmentation has drawn significant attention from the research community owing to the rich and unique textures of the iris, such as: rings, crypts, furrows, freckles, and ridges [3]. While ocular traits other than the iris are less frequently studied, researchers have been actively investigating other periocular regions, such as the sclera and retina, to collect identity cues that might be useful for stand-alone recognition systems or to supplement the information generally used for iris recognition [4].

A majority of previous research studies on eye region segmentation were restricted to a single ocular region at a time, for e.g., focusing only on the iris, pupil, sclera, or retina. In multi-class segmentation, more than one eye region is segmented from the given input image using a single segmentation network. Inexplicably, very few researchers have developed multi-class segmentation techniques for the eye regions, despite several advantages in different applications. Namely, under challenging conditions, the segmentation performance can be maintained or sometimes be even enhanced when using multiple region segmentation, as the targeted region can provide useful contextual information about other neighboring regions [5]. For example, boundary of iris region can provide useful information about the boundary of the sclera and pupil region. Similarly, the eyelash area presents a constraint for the sclera region [6]. Another potential advantage is that multi-biometric systems can be introduced without cost and computation overheads, which can work efficiently work for the segmentation of multiple target classes using a single approach [7].

In this work, we attempt to address the research gaps in the segmentation of multiple eye regions using a single network, as shown in Fig. 1. The proposed network can segment the input eye image into four main classes corresponding to the iris, sclera, pupil, and background region using a single model. Over the last few years, deep learning convolutional neural network (CNN) models witnessed rapid development, to be an influential method for image processing tasks. CNNs have outperformed conventional methods in a wide range of applications, such as in medical and satellite image analysis. The proposed method is based on deep learning models for semantic segmentation in images, specifically on convolutional encoder-decoder networks. This design approach is based on the recently introduced SegNet architecture [8]. ORED-Net was developed based on the outer residual encoder-decoder network. The proposed network achieves a higher accuracy with reduced network depth and fewer number of parameters and layers by implementing only non-identity outer residual paths from the encoder to the decoder.

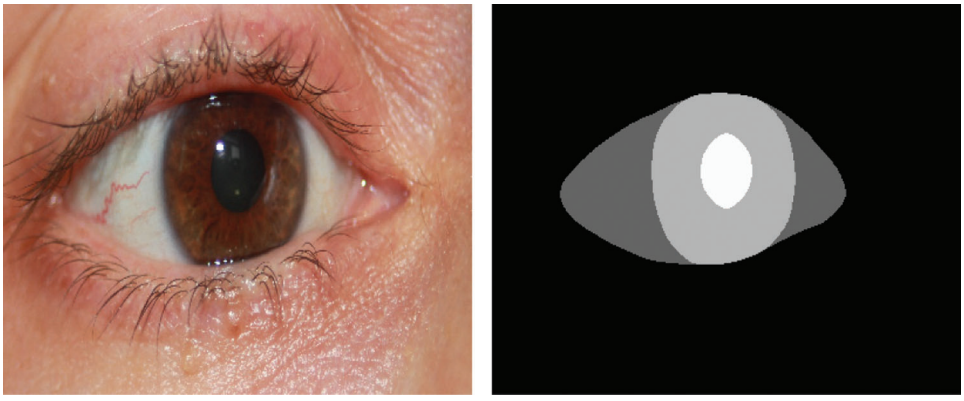


Figure 1: Sample images of multi-class eye segmentation, illustrating the input eye image (left), ground truth image (right). Each color shade represents a different eye region

ORED-Net is novel in the following four ways:

- ORED-Net is a semantic segmentation network without a preprocessing overhead and does not employ conventional image processing schemes.
- ORED-Net is a standalone network for the multi-class segmentation of ocular regions.
- ORED-Net uses residual skip connections from the encoder to the decoder to reduce information loss, which allows the flow of high-frequency information through the model, thus achieving higher accuracy with a few layers.
- The performance of the proposed ORED-Net model was tested on public datasets collected under various environments.

In this study, the results obtained with the SBVPI [9] and UBIRIS.v2 [10] datasets for the iris, sclera, pupil, and background classes are reported. In addition, the proposed model is compared with state-of-the-art techniques from the literature. The results demonstrate that the proposed method is the most suitable technique for ocular segmentation, which can be incorporated in recognition procedures.

The rest of the paper is structured as follows. In Section 2, a brief overview of related literature is provided. In Section 3, the proposed approach and working procedure are described. The results of the evaluation and analysis are discussed in Section 4. Finally, conclusion and future work are presented in Section 5.

2 Literature Review

Very few studies have focused on multi-class eye segmentation, particularly for segmenting multiple eye regions from the given images using a single segmentation model. Recently, Rot et al. [7] reported the segmentation of multi-class eye regions based on the well-known convolutional encoder-decoder network SegNet. They studied the segmentation of multiple eye regions such as the iris, sclera, pupil, eyelashes, medial canthus, and periocular region. This study required post-processing through thresholding strategy on probability maps and an atrous CNN with the conditional random field detailed in Luo et al. [11]. The results were extracted using the Multi-Angle Sclera Database (MASD). Naqvi et al. proposed the Ocular-Net CNN for the segmentation of multiple eye regions, including the iris and sclera. This network consists of non-identity residual paths in a lighter version of both the encoder and decoder. Residual shortcut connections were employed with increasing network depth to enhance the performance of the model [12]. In addition, the iris and sclera were evaluated on different databases. Hassan et al. proposed the SIP-SegNet CNN for joint semantic segmentation of the iris, sclera, and pupil. A denoising CNN (DnCNN) was used to denoise the original image. In SIP-SegNet, after denoising with DnCNN, reflection removal and image enhancement were performed based on contrast limited adaptive histogram equalization (CLAHE). Then, the periocular information was extracted using adaptive thresholding, and this information was suppressed using the fuzzy filtering technique. Finally, a densely connected fully convolutional encoder-decoder network was used for semantic segmentation of multiple eye regions [13]. Various metrics were used to evaluate the proposed method that was tested on the CASIA sub-datasets.

The Eye Segmentation challenge for the segmentation of key eye regions was organized by Facebook Research with the purpose to developing a generalized model with the condition of least complexity in terms of the model parameters. Experiments were conducted on the OpenEDS dataset which is a large-scale dataset of eye images captured by a head-mounted display with two synchronized eye facing cameras [14]. To address the challenge concerning the semantic segmentation of eye regions, Kansal et al. [15] proposed Eyenet, Attention-based Convolutional Encoder-Decoder Network for accurate segmentation of four different eye regions, namely the iris, sclera, pupil and background. Eyenet is based on non-identity mapping based residual connections in both the encoder and decoder. Two types of attention units and

multiscale supervision were proposed to obtain accurate and sharp boundary eye regions. Eye segmentation using a lightweight model was demonstrated by Huynh et al. Their approach involved the conversion of the input image to grayscale, segmentation of the eye regions with a deep network model, and removal of the incorrect areas using heuristic filters. A heuristic filter was used to reduce the false positive in the output of the model [16].

Tab. 1 presents a comparison of the proposed method with other methods for the multi-class segmentation of ocular regions along with their strengths and weaknesses.

Table 1: Comparison of the proposed method with other multi-class segmentation methods

Methods	Strengths	Weaknesses
Deep multi-class eye segmentation based on the SegNet architecture [7]	—A single model is used for the segmentation of multiple eye regions	—A major part of the training data is artificially created. —Considerable post-processing is involved.
Lighter residual encoder-decoder network, Ocular-Net [12]	—Residual connectivity between adjacent convolutional layers is involved	—The method is trained separately for each region —Only one ocular region is addressed at a time
Joint semantic segmentation of eye regions, SIP-SegNet [13]	—DnCNN is used for denoising the original images	—Considerable preprocessing of the original image is involved. —Periocular region suppression is required.
Encoder-decoder structure based on A depthwise convolution operation [16]	—Can be run on any hardware for real-time implementation with low computational cost	—Post-processing is performed via heuristic filtering —The method was trained and tested only on the OpenEDS dataset.
Outer residual encoder-decoder network, termed as ORED-Net (**Proposed Method**)	—Information loss is reduced by using outer residual skip paths from the encoder to the decoder. —The training time is also reduced because of the outer residual paths.	—Rigorous training is required.

3 Proposed Method for Eye Region Segmentation

3.1 Overview of the Proposed Model

The flowchart of the proposed ORED-Net for semantic segmentation of multiple eye regions is shown in Fig. 2. The proposed network is a fully convolutional network based on non-identity residual connections from the encoder network to the decoder network. The input image is fed into the convolutional network without an initial preprocessing overhead. An encoder and decoder are incorporated in the proposed ORED-Net for multi-class segmentation of the full input eye images. The functionality of the encoder is to downsample the given input image until it can be represented in terms of very small features, whereas the decoder performs the reverse operation. The decoder upsamples the image back to its original dimensions using the small features produced by the encoder. In addition to the reverse process of

downsampling, the decoder plays another very important role of predicting multiple classes, namely the iris, sclera, pupil, and background. The prediction task is performed using the Softmax loss function and a pixel classification layer. The class of each pixel in the image is predicted by the pixel classification layer, and the designated label is assigned.

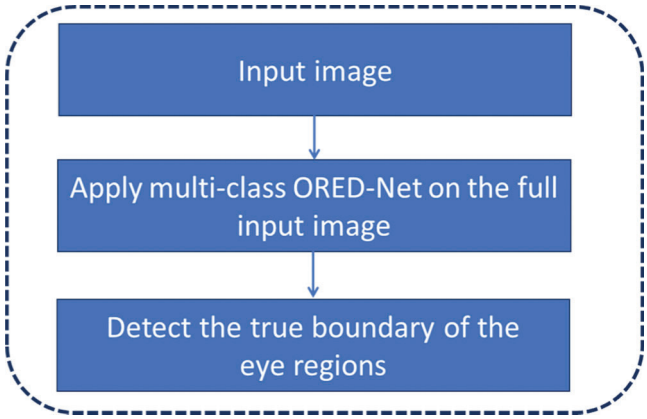


Figure 2: Flowchart of the proposed method for eye regions segmentation

3.2 Segmentation of Multiple Eye Regions Using ORED-Net

The image in typical encoder-decoder networks is downsampled and represented by very small features, which basically degrades the high-frequency contextual information. This results in the vanishing gradient problem for the classification of image pixels as the image is broken down into 7 × 7 sized patches [17]. The vanishing gradient problem was addressed by introducing identity and non-identity mapping residual blocks. When a residual block is introduced in a CNN, the accuracy achieved is higher than that of simple CNNs such as VGGNet [18]. Typically, residual building blocks (RBBs) are based on identity and non-identity mapping. In identity mapping, the features are directly provided for element-wise addition to perform the residual operation. In contrast, in the case of non-identity mapping, a 1 × 1 convolution is performed in each RBB before the features are subject to the element-wise addition. Identity mapping is not considered in the proposed network. Instead, non-identity mapping is performed by a 1 × 1 convolution layer through outer residual paths from the encoder to the decoder, as shown in Fig. 3.

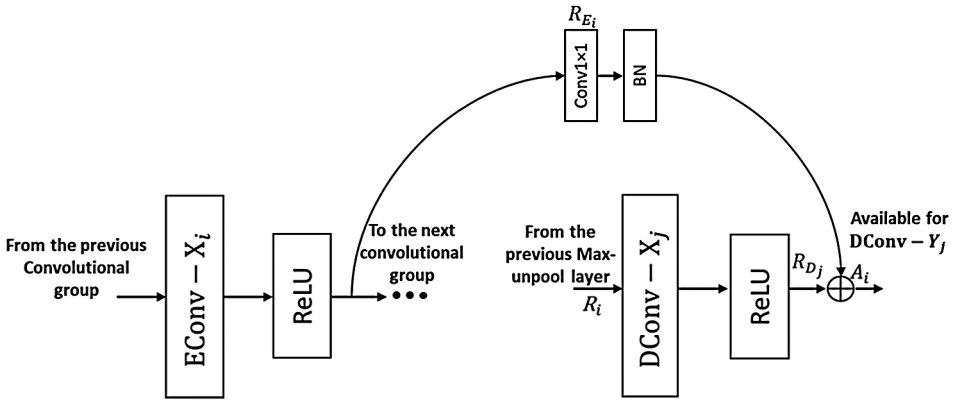


Figure 3: Residual building block (RBB) used in the proposed method

The proposed ORED-Net is executed via different developmental stages to perform the multi-class segmentation task with good accuracy, as compared with the basic encoder-decoder networks. In the first stage, a well-known network for segmentation i.e., SegNet-Basic is employed [8]. SegNet-Basic consists of 13 convolutional layers in both the encoder and decoder parts. This network is reduced to its simplest possible form by removing 5 convolutional layers from both the encoder and decoder parts of the network. Hence, the proposed network has only 8 convolutional layers in the encoder and decoder parts. In addition, each group in the encoder and decoder architectures consists of two convolutional layers, resulting in a lightweight encoder and decoder convolutional network. ORED-Net ensures the empowerment of the high-frequency features. In the next preparation stage of the proposed ORED-Network, non-identity residual connections are introduced from the layers on the encoder side to the corresponding layers on the decoder side through the outer residual paths, as schematically shown in Fig. 4. Hence, the residual connectivity introduced in ORED-Net is different from that of the original ResNet [19] and previously proposed residual-based networks such as Sclera-Net [4]. Tab. 2 highlights the main differences between the proposed network and previously reported networks such as ResNet [19] and Sclera-Net [4].

Table 2: Key architectural differences between ORED-Net and other residual based methods

ResNet [19]	Sclera-Net [4]	ORED-Net
ResNet uses a large number of identity mapping and a small number of non-identity mapping residual connections.	Convolutional layers in the encoder and decoder have identity and non-identity based residual connectivity.	Convolutional layers in the encoder and decoder do not have internal residual connectivity.
ResNet uses the skip path connection only between adjacent layers.	There are no outer skip path connections from the encoder to the decoder.	The outer skip path connections from the encoder to the decoder are non-identity residual connections.
Different variants of ResNet such as ResNet-50/101/152 have a 1×1 convolutional layer in each block.	There are 6 identity and 8 non-identity residual connections in the overall encoder-decoder network.	There are 4 non-identity residual paths from the encoder to the decoder.
Different variants of ResNet, such as ResNet-18/34/50/101, are based on post activation as a ReLU is used after the elementwise addition.	In the overall network, a ReLU is used after the elementwise addition. Hence, Sclera-Net uses post activation.	On the decoder side, a ReLU is used before the elementwise addition. Hence, ORED-Net uses pre-activation.
At the end of all the convolutional layers, average pooling is involved.	Residual connections are introduced immediately after max pooling and unpooling in the encoder and decoder networks, respectively.	The max-pooling layer is used in all the convolutional blocks to provide index information to the decoder

The overall structure of ORED-Net is shown in Fig. 4. Here, four non-identity outer residual paths (Outer-Residual-Path-1, ORED-P-1 to Outer-Residual-Path-4, ORED-P-4) from the encoder to the decoder are illustrated. The group containing a convolutional layer of size 3 × 3 and batch normalization layers is represented as Conv + BN, the activation layer, i.e., rectified linear unit is represented as ReLU,

the combination of a convolution layer of size 1 × 1 and batch normalization layers is represented as 1 × 1 Conv + BN, the max pooling layer is represented as Max-pool, and the reverse of the max pooling layer, i.e., the max unpooling layer, is represented as Max-unpool. There are four convolutional groups in the encoder, with each group consisting of two convolutional layers before each Max-pool, i.e., E-Conv-X and E-Conv-Y. Similarly, in the decoder, there are four convolutional groups, with each decoder group also consisting of two convolutional layers after each Max-unpool layer, i.e., D-Conv-X and D-Conv-Y. Therefore, the 1st convolutional layer of the i-th encoder of the convolutional group is represented as E-Conv-Xi, and the 2nd convolutional layer of the i-th encoder of the convolutional group is represented as E-Conv-Yi. Similarly, the 1st convolutional layer of the j-th decoder of the convolutional group is represented as D-Conv-Xj, and the 2nd convolutional layer of the j-th decoder of the convolutional group is represented as D-Conv-Yj. Here, the values of i and j are in the range of 1–4. The 1st encoder-decoder convolutional groups located at the extreme left and right sides of the network are connected through ORED-Path-1. Similarly, the 2nd convolutional groups located 2nd from the left and right sides of the convolutional group are connected through ORED-Path-2, as shown in Fig. 4.

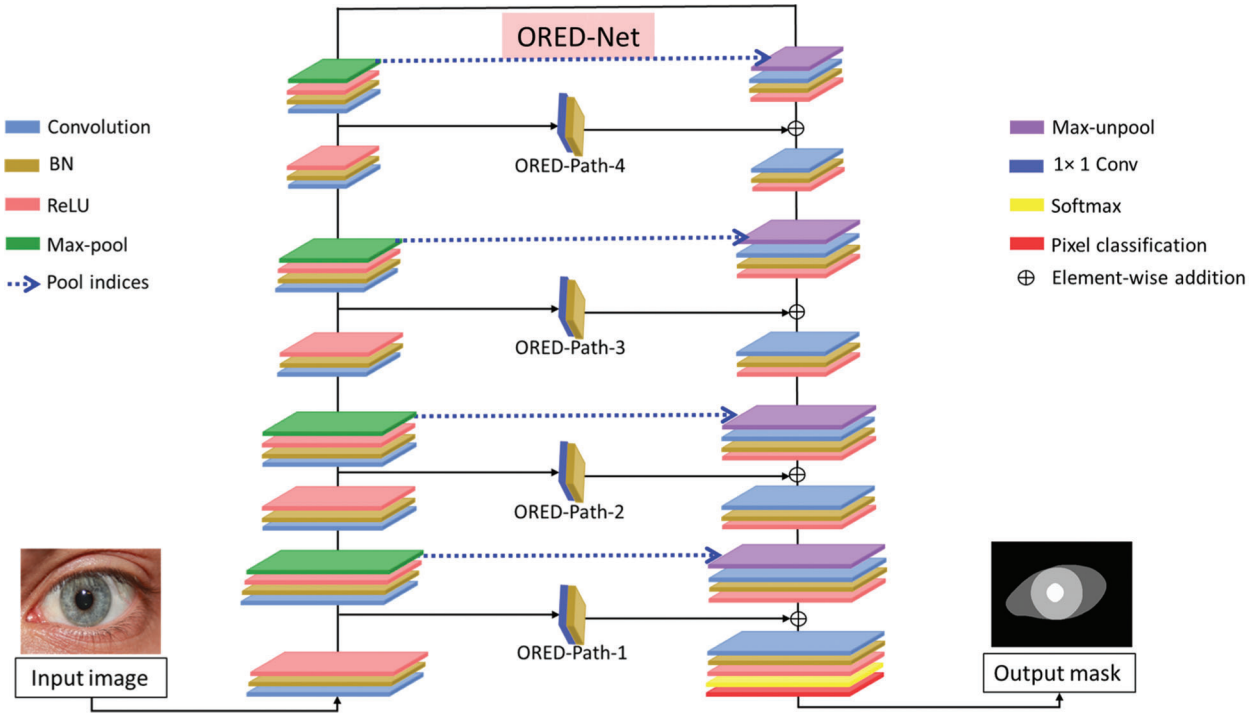


Figure 4: Deep learning-based eye region segmentation system with light-residual encoder and decoder network

Based on Fig. 4, it can be observed that at the decoder part, the 2nd convolutional layer in the 1st convolutional group receives the element-to-element addition of the residual features R_{E_1} and R_{D_1}, wherein these features are from the 1st convolutional layer in the encoder convolutional-group-1 (E-Conv-X1) after the ReLU and the 1st convolution layer in the 1st decoder convolutional group-1 (D-Conv-X1) after the ReLU, respectively, via ORED-Path-1. This can be described by the following equation:

$$A_1 = R_{E_1} + R_{D_1} \tag{1}$$

Here, A_1 is the residual feature with element-to-element addition input to D-Conv-Y1 through ORED-Path-1. Typically, the outer residual block shown in Fig. 3 can be represented by the following equation:

$$A_i = R_{E_i} + R_{D_j} \tag{2}$$

where A_i is the sum of the features presented to D-Conv-Yj by the outer residual connection, R_{E_i} represents the residual features from the 1st convolutional layer of the i-th convolutional group (E-Conv-Xi) after the ReLU at the encoder part, and R_{D_j} represents the residual features obtained from the 1st convolutional layer of the j-th convolutional group (D-Conv-Xj) after the ReLU on the decoder side. Furthermore, the values of i and j are between 1 and 4. Thus, to enhance the ability of the network for robust segmentation, each of the four outer residual paths (ORED-P-1 to ORED-P-4) provides the residual features R_{E_i} from each of the convolutional groups from the encoder side to the decoder side. This direct path of the spatial edge information from the encoder side empowers the residual features of the decoder side, i.e., the R_{D_j} features.

3.2.1 ORED-Net Encoder

It can be seen from Fig. 4 that the encoder consists of 4 convolutional groups, with each group containing two convolutional layers along with the batch normalization and ReLU activation layers. The core and exclusive characteristic of the ORED-Net encoder is that the spatial information is input to the subsequent decoder group by the residual paths. These outer residual paths originate after each ReLU layer on the encoder side. Due to the outer residual connections, better results can be achieved with a lighter network compared with other networks used for a similar purpose. In ORED-Net, the important features are downsampled through the Max-pool layers, which also provide pooling indices to the decoder side. The pooling indices contain the index information and feature map size, which are required on the decoder side.

The encoder structure of ORED-Net is presented in Tab. 3. It can be observed that there are 4 outer residual encoder-decoder paths that connect the encoder with the decoder through the non-identity residual connection shown in Fig. 4. These outer residual encoder-decoder non-identity residual connections achieve feature empowerment through the spatial information of the preceding layers. The outer residual encoder-decoder connections originate after the ReLU activation layer on the encoder side and end next to the ReLU activation layer on the decoder side. The proposed network uses pre-activation because summation is performed after each ReLU layer on the decoder side. In every convolutional group on the encoder and decoder sides, an equal number of convolutional layers are present i.e., two convolutional layers, which makes ORED-Net a balanced network.

Table 3: The ORED-Net encoder based on outer residual encoder decoder paths

Group	Size/Name	No. of filters	Output (w × h × ch)
EC-G-1	3 × 3 × 3/E-Conv-1_1††	64	224 × 224 × 64
	To decoder 1 × 1 × 64/ORED-P-1†	64	
	3 × 3 × 64/E-Conv-1_2††	64	
Pool-1	2 × 2/Pool-1	–	112 × 112 × 64
EC-G-2	3 × 3 × 64/E-Conv-2_1††	128	112 × 112 × 128

Table 3 (continued).

Group	Size/Name	No. of filters	Output (w × h × ch)
	To decoder 1 × 1 × 128/ORED-P-2†	128	
	3 × 3 × 128/E-Conv-2_2†	128	
Pool-2	2 × 2/Pool-2	–	64 × 64 × 128
EC-G-3	3 × 3 × 128/E-Conv-3_1††	256	64 × 64 × 256
	To decoder 1 × 1 × 256/ORED-P-3†	256	
	3 × 3 × 256/E-Conv-3_2†	256	
Pool-3	2 × 2/Pool-3	–	32 × 32 × 256
EC-G-4	3 × 3 × 256/E-Conv-4_1††	512	32 × 32 × 512
	To decoder 1 × 1 × 512/ORED-P-4†	512	
	3 × 3 × 512/E-Conv-4_2††	512	
Pool-4	Pool-4/2 × 2	–	16 × 16 × 512

Tab. 3 presents the ORED-Net encoder with outer residual paths based on an image with size 224 × 224 × 3. Here, E-Conv, ORED-P, and Pool represents the encoder convolution layers, outer residual encoder-decoder paths, and pooling layers, respectively. The convolutional layers in the encoder, represented by the symbol "††", include both the ReLU activation and batch normalization (BN) layers, while the convolution layers represented by "†" include only the BN layer. The outer residual encoder-decoder skip paths, denoted by ORED-P-1 to ORED-P-4, start from the encoder and carry edge information to the decoder. As the proposed model includes pre-activation, the ReLU activation layer is used prior to the element-wise addition.

3.2.2 ORED-Net Decoder

The architecture of the ORED-Net decoder shown in Fig. 4 is such that it mirrors the encoder and performs a similar convolutional operation as that performed by the encoder. The pooling layers of the encoder provide the size information and indices to the decoder, which are used to maintain the size of the feature map. In addition, the decoder features are upsampled to ensure that the size of the network output is the same as that of the input image. Furthermore, the outer residual paths input the features to the ORED-Net decoder. All the 4 outer residual encoder-decoder paths, i.e., ORED-P-1 to ORED-P-4, originate from the encoder side and end on the decoder side. Element-to-element addition between the ORED-P and previous convolution is performed in the addition layers (Add-4 to Add-1), resulting in features that are useful to the convolutional layers in the next group, as shown in Fig. 4. In this work, as four classes namely the iris, sclera, pupil, and background, are evaluated for segmentation task, the decoder produces four masks corresponding to these classes, i.e., the number of filters for the last convolutional layer in the decoder. The pixel classification and Softmax layers facilitate the pixel-wise prediction of the network. To implement post activation in the decoder, the outer residual path is terminated immediately after each ReLU activation layer. For each class, the output of ORED-Net is a mask, which outputs "0" for the BG class, "100" for the sclera class, "180" for the iris class, "250" for the pupil class.

4 Results and Discussion

In this work, two-fold cross-validation was performed for training and testing the proposed model. To this end, two subsets were created from the available images by randomly dividing the collected database. From the images of 55 participants, two subsets were created, where the data from 28 participants were used for training and that from 27 participants were used for testing. To avoid overfitting issues, data augmentation of the training data was performed. To train and test ORED-Net, a desktop computer with an Intel® Core™ (Santa Clara, CA, USA) i7-8700 CPU @3.20 GHz, 16 GB memory, and an NVIDIA GeForce RTX 2060 Super (2176 CUDA cores and 8 GB GDDR6 memory) graphics card were employed. The above-mentioned experiments were conducted using MATLAB R2019b.

4.1 Training of ORED-Net

ORED-Net is based on outer residual paths from the encoder to the decoder for transferring spatial information from the encoder side to the decoder side. Therefore, high frequency information travels through the convolutional network that empowers training of this information without a preprocessing overhead. To train ORED-Net, original images without any enhancement or preprocessing were employed, and a classical stochastic gradient descent (SGD) method was used as an optimizer. SGD minimizes the difference between the actual and predicted outputs. During network training, the proposed model executed the entire dataset 25 times, i.e., 25 epochs, and a mini-batch size of 5 was selected for the ORED-Net design owing to its low memory requirement. The mini-batch size was determined by the size of the database. Once training was performed with the entire dataset, one epoch was counted, as shown in Eqs. (3) and (4).

$$u_{i+1} := mu_i - x\eta v_i - \eta\left\langle \frac{\partial S_i(v)}{\partial v}\middle| v_i > T_i \right. \tag{3}$$

$$v_{i+1} := v_i + u_{i+1} \tag{4}$$

In Eqs. (3) and (4), u_i is the momentum variable, v_i is the learnt weight at the i^{th} iteration, m is the momentum, η is the learning rate and x is the weight decay. The average over the i^{th} batch T_i of the derivative of the object with respect to v, evaluated at v_i, is given by $\left\langle \frac{\partial S_i(v)}{\partial v}\middle| v_i > T_i \right.$. Using the SGD method, the optimal training parameters m, η, and x defined in Eqs. 3 and 4 were set to 0.9, 0.001, and 0.0005, respectively.

The ORED-Net model converges very quickly because of the outer residual connections from the encoder to the decoder. Therefore, the ORED-Net model was only trained for 25 epochs. The mini-batch size was kept to 5 images during 25 epochs of training with shuffling after each epoch. Here, the training loss was calculated based on the image pixels in the mini-batch using the cross-entropy loss reported [8]. The loss calculation was based on the cross-entropy loss over all the pixels accessible in the candidate mini-batch based on the iris, sclera, pupil, or background classes. Moreover, the network convergence and accuracy were affected due to a higher difference between the number of pixels in different classes and bias of the network towards learning the dominant class, as described in Arsalan et al. [20]. During class training, the imbalance among the classes can be removed by assigning an inverse frequency weighting approach, as defined in Eqs. (5) and (6).

$$Freq. = \frac{Pixels(i)}{Total\ Pixels} \tag{5}$$

$$Class\ Weights = \frac{1}{Freq.} \tag{6}$$

Here, *Pixels(i)* is the total number of pixels belonging to class ð in the training data. In this study, ð = 4 represents the four classes namely the iris, sclera, pupil, and background.

4.2 Testing of ORED-Net

4.2.1 Evaluation Metrics

To validate and compare ORED-Net with previous models, the average segmentation error (Err_{avg}), mean Intersection over Union (mIoU), Precision (P), Recall (R), and F1-score (F) were adopted as evaluation protocols.

$$Err_{avg} = \frac{1}{MxNxT}\left[\sum_{k=1}^{T}\sum_{i,j\in(M,N)} G(i,j) \oplus O(i,j)\right] \tag{7}$$

Here, T represents the total number of images with a $M \times N$ spatial resolution. $G(i, j)$ and O(i, j) are the pixels of the mask or ground truth and the predicted labels, respectively.

$$mIoU = \frac{1}{N_c}\left[\sum_{i=1}^{N_c}\left(\frac{N_{xx}(i)}{N_{xx}(i) + N_{xy}(i) + N_{yx}(i)}\right)\right] \tag{8}$$

$$P = \frac{N_{xx}}{N_{xx} + N_{xy}} \tag{9}$$

$$R = \frac{N_{xx}}{N_{xx} + N_{yx}} \tag{10}$$

$$F = \frac{2RP}{R + P} \tag{11}$$

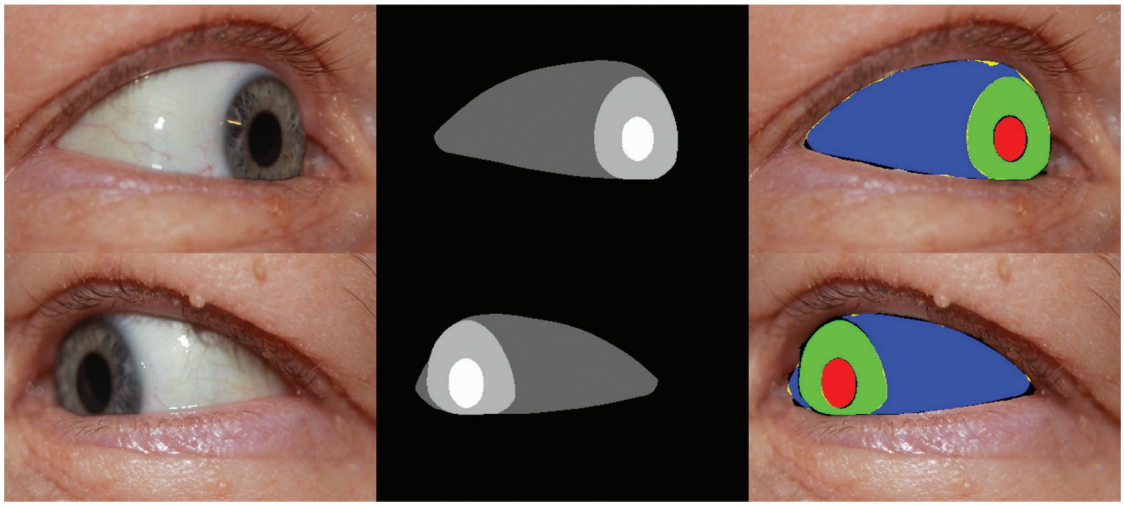
Here N_c represents the total number of classes, and N_{xx} is defined as the true positive where the number of pixels predicted as x also belong to class x. Similarly, the other terms are defined as the true negatives N_{yy}, false positives N_{xy}, and false negatives N_{yx}.

4.2.2 Eye Regions Segmentation Results Obtained with ORED-Net

In Figs. 5 and 6, the correct and incorrect results of multi-class eye region segmentation of eye images obtained with ORED-Net for the SBVPI dataset are illustrated. These pictorial representations follow the convention of *FP* (shown in black for each class), *FN* (shown in yellow for each class), and *TP* (shown in green, blue, and red for the iris, sclera, and pupil classes respectively).

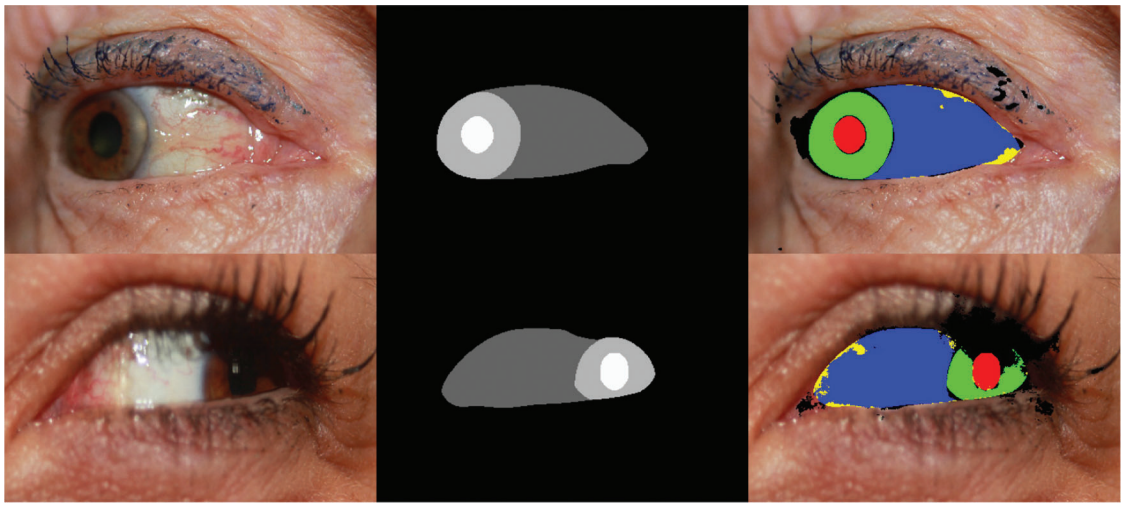
4.2.3 Comparison of ORED-Net with Other Methods

The segmentation performance of ORED-Net was compared with previous methods in terms of the Err_{avg}, mIoU, P, R, and F described in Section 4.4.1. Tab. 4 presents a comparison of the segmentation performance of existing methods with that achieved by ORED-Net for the SBVPI dataset. The results demonstrate the superior performance of ORED-Net for eye region segmentation compared with the current methods, based on the values of Err_{avg}, mIoU, P, R, and F. Comparisons are presented for the iris, sclera, pupil, and background regions with the current state-of-the-art methods in Tab. 4. Additionally, the results of mIoU, P, R, and F in Tab. 4 are presented through bar graphs in Fig. 7.



(a) (b) (c)

Figure 5: Examples of good eye region segmentation by ORED-Net for the SBVPI dataset: (a) Original image, (b) Ground-truth mask, and (c) Predicted mask result obtained with ORED-Net



(a) (b) (c)

Figure 6: Examples of bad eye region segmentation by ORED-Net for the SBVPI dataset: (a) Original image, (b) Ground-truth mask, and (c) Predicted mask result obtained with ORED-Net

4.2.4 Eye Region Segmentation with Other Open Datasets Using ORED-Net

To evaluate the segmentation performance of ORED-Net under different image acquisition conditions, experiments with another publicly available datasets for eye region segmentation, i.e., UBIRIS.v2 dataset, were included in this study [10]. In previous studies, masks for the iris and sclera were provided for only 300 images [21]. The ground truth images of the iris and sclera were merged and the ground truths for the pupil were designed to evaluate the proposed ORED-Net model on the iris, sclera and pupil using the UBIRIS.v2 dataset. Of the 300 images considered in the UBIRIS.v2 dataset, 50% of the images (150) were used for training, while the remaining 50% (150) were used for testing with two-fold

cross-validation. To train ORED-Net with the UBIRIS.v2 dataset, similar data augmentation as that used for the SBVPI dataset was employed.

Table 4: Comparison of the proposed method with existing methods for the SBVPI dataset (unit: %)

Evaluation Metrics	Classes	SegNet [8]			ScleraNet [4]			ORED-Net		
		Fold 1	Fold 2	Average	Fold 1	Fold 2	Average	Fold 1	Fold 2	Average
Err_{avg}	**Background**	3.34	1.84	2.59	3.15	1.57	2.36	2.16	1.40	1.78
	Iris	1.54	0.89	1.22	1.90	0.68	1.29	1.12	0.62	0.87
	Sclera	2.69	1.79	2.24	1.93	1.51	1.72	1.67	1.34	1.51
	Pupil	0.19	0.20	0.20	0.31	0.18	0.25	0.33	0.14	0.24
	All classes	1.94	1.18	**1.56**	1.82	0.99	**1.40**	1.32	0.88	**1.10**
mIoU	**Background**	95.84	97.67	96.76	96.07	98.12	97.10	97.30	98.24	97.77
	Iris	82.99	86.15	84.57	82.59	88.62	85.61	86.80	89.65	88.23
	Sclera	81.05	86.44	83.75	85.37	88.58	86.98	87.39	89.49	88.44
	Pupil	79.89	79.92	79.91	79.9	84.48	82.19	78.74	86.35	82.55
	All classes	84.94	87.55	**86.24**	85.98	89.95	**87.97**	87.56	90.93	**89.25**
P	**Background**	99.72	99.79	99.76	99.73	99.78	99.76	99.70	99.75	99.73
	Iris	85.27	89.85	87.56	85.62	92.43	89.03	90.97	93.52	92.25
	Sclera	83.53	88.52	86.03	88.35	90.49	89.42	89.95	91.52	90.74
	Pupil	92.83	85.16	89.00	80.15	88.57	84.36	79.25	87.87	83.56
	All classes	90.34	90.83	**90.58**	88.46	92.82	**90.64**	89.97	93.17	**91.57**
R	**Background**	96.09	97.88	96.99	96.31	98.23	97.27	97.59	98.48	98.04
	Iris	96.90	95.44	96.17	95.85	95.24	95.55	94.93	95.28	95.11
	Sclera	96.20	97.34	96.77	96.04	97.68	96.86	96.81	97.61	97.21
	Pupil	85.51	94.01	89.76	99.66	95.49	97.58	99.19	98.24	98.72
	All classes	93.68	96.17	**94.92**	96.97	96.66	**96.81**	97.13	97.40	**97.27**
F	**Background**	97.82	98.80	98.31	97.92	98.99	98.46	98.59	99.11	98.85
	Iris	90.05	92.13	91.09	89.19	93.58	91.39	92.39	94.25	93.32
	Sclera	89.05	92.66	90.86	91.88	93.88	92.88	93.03	94.41	93.72
	Pupil	88.27	87.55	87.91	88.62	90.79	89.71	88.08	92.05	90.07
	All classes	91.30	92.79	**92.04**	91.90	94.31	**93.11**	93.02	94.96	**93.99**

In Figs. 8 and 9, the correct and incorrect results of multi-class eye region segmentation of eye images obtained with ORED-Net for the UBIRIS.v2 dataset are illustrated. This pictorial representation follows the convention of *FP* (shown in black for each class), *FN* (shown in yellow for each class), and *TP* (shown in green, blue, and red for the iris, sclera, and pupil classes, respectively). As ORED-Net is powered by outer residual paths, there are no significant errors in the segmentation of multiple eye region from a challenging dataset like UBIRIS.v2.

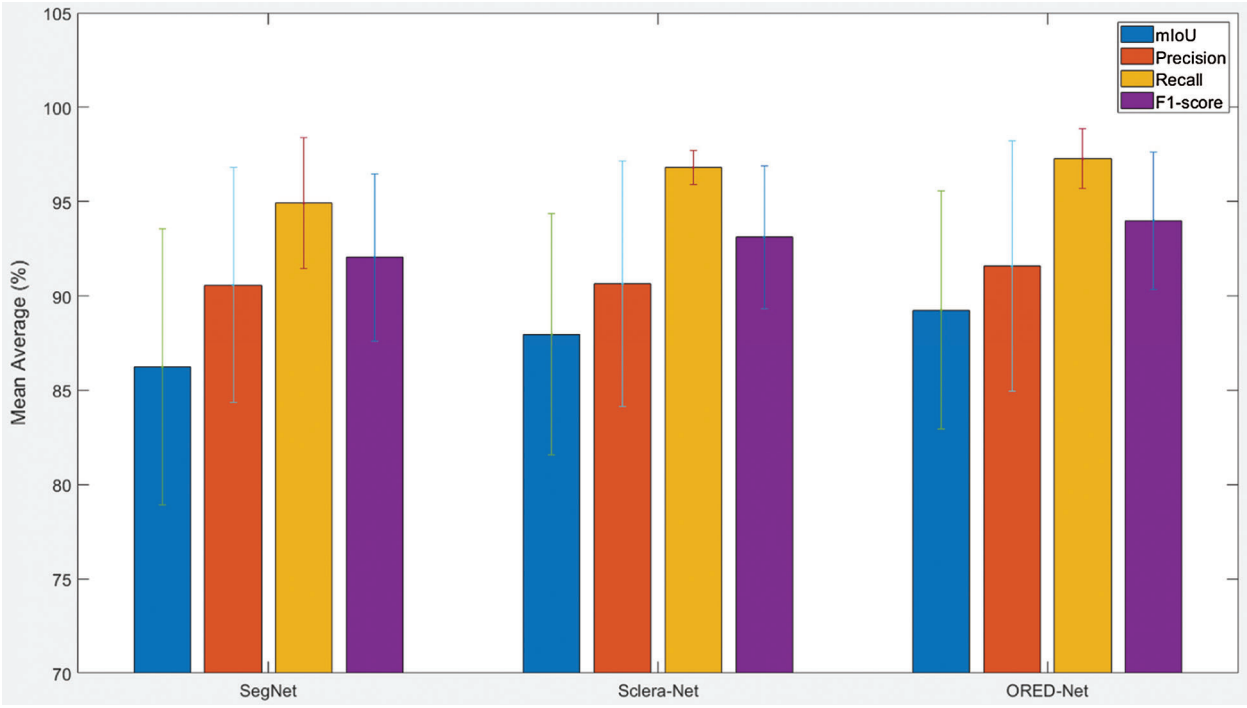
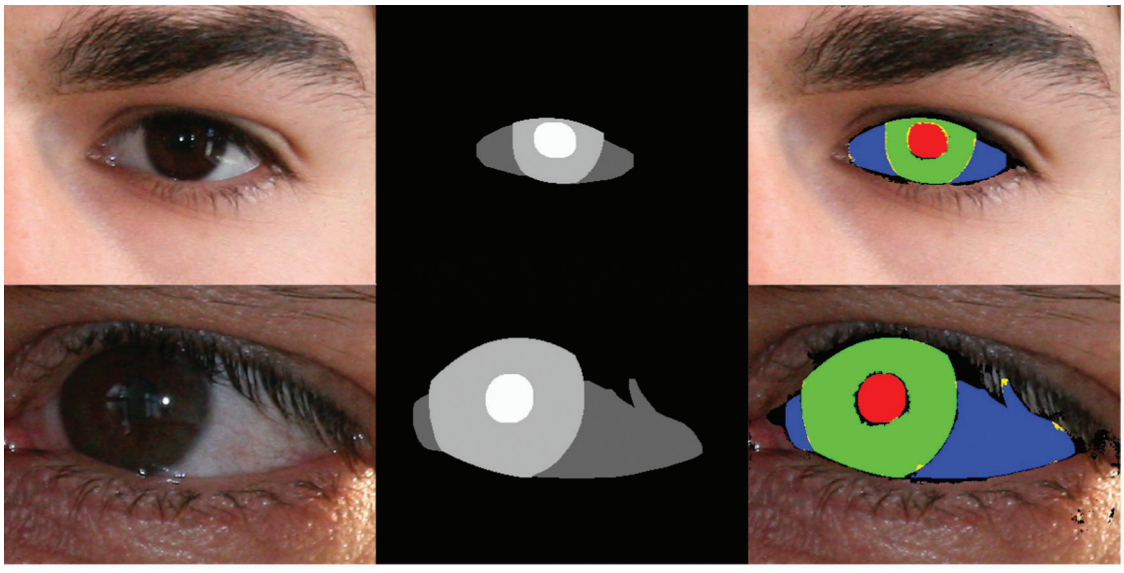
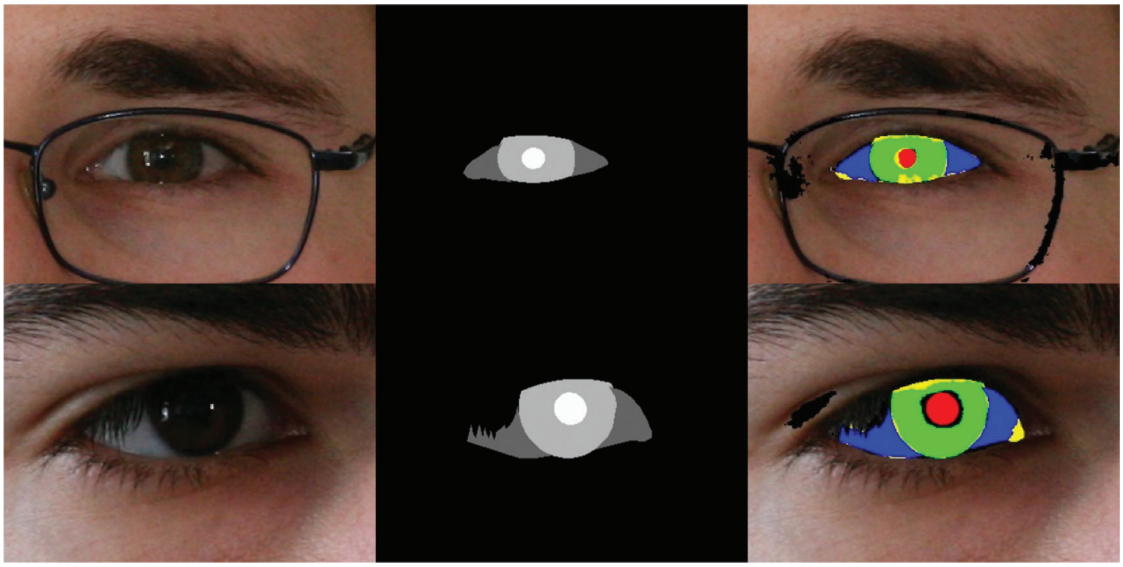


Figure 7: Mean and standard deviation of the proposed method and existing alternatives in terms of mean intersection over union, precision, recall and F1-score based on SBVPI database



(a) (b) (c)

Figure 8: Examples of good eye region segmentation by ORED-Net for the UBIRIS.v2 dataset: (a) Original image, (b) Ground-truth mask, and (c) Predicted mask result obtained with ORED-Net



(a) (b) (c)

Figure 9: Examples of bad eye region segmentation by ORED-Net for the UBIRIS.v2 dataset: (a) Original image, (b) Ground-truth mask, and (c) Predicted mask result obtained with ORED-Net

Tab. 5 presents a comparison of the segmentation performance of existing methods with that of ORED-Net for the UBIRIS.v2 dataset. Additionally, the results of mIoU, P, R, and F in Tab. 5 are presented through bar graphs in Fig. 10.

Based on the results presented in Tabs. 4 and 5 (Figs. 7 and 10), it can be concluded that the performance of the proposed ORED-Net framework is consistent with that of state-of-the-art algorithms. A noteworthy point is that ORED-Net is a novel method that performs multi-class semantic segmentation of different eye regions such as iris, sclera, and pupil simultaneously, unlike all other existing algorithms that only address one or two eye region at a time. In addition, the performance of the ORED-Net model was evaluated on different publicly available datasets for comparisons with other methods, as shown in Tabs. 4 and 5 (Figs. 7 and 10).

Table 5: Comparison of the proposed ORED-Net method with existing methods for the UBIRIS.v2 dataset (unit: %)

Evaluation Metrics	Classes	SegNet [8]			ScleraNet [4]			ORED-Net		
		Fold 1	Fold 2	Average	Fold 1	Fold 2	Average	Fold 1	Fold 2	Average
Err_{avg}	**Background**	2.73	1.28	2.01	2.47	1.47	1.97	2.36	1.30	1.83
	Iris	1.38	0.69	1.04	1.36	0.87	1.12	1.61	0.77	1.19
	Sclera	2.19	1.03	1.61	1.42	1.10	1.26	1.16	0.92	1.04
	Pupil	0.42	0.21	0.32	0.30	0.24	0.27	0.27	0.18	0.23
	All classes	1.68	0.80	**1.24**	1.39	0.92	**1.15**	1.35	0.79	**1.07**
mIoU	**Background**	96.79	98.46	97.63	97.03	98.23	97.63	97.29	98.44	97.87
	Iris	78.43	89.02	83.73	77.99	87.14	82.57	79.98	88.42	84.20
	Sclera	64.01	81.06	72.54	73.43	79.76	76.60	77.54	82.54	80.04

(Continued)

Table 5 (continued).

Evaluation Metrics	Classes	SegNet [8]			ScleraNet [4]			ORED-Net		
		Fold 1	Fold 2	Average	Fold 1	Fold 2	Average	Fold 1	Fold 2	Average
	Pupil	63.45	78.62	71.04	71.69	78.29	74.99	74.92	81.86	78.39
	All classes	75.67	86.79	**81.23**	80.04	85.86	**82.95**	82.43	87.82	**85.12**
P	**Background**	99.65	99.88	99.77	99.59	99.89	99.74	97.99	99.86	98.93
	Iris	87.02	92.92	89.97	84.59	91.64	88.12	87.20	92.21	89.71
	Sclera	66.63	83.03	74.83	76.27	81.49	78.88	79.98	84.40	82.19
	Pupil	68.43	81.89	75.16	73.76	82.09	77.93	77.90	85.42	81.66
	All classes	80.43	89.43	**84.93**	83.55	88.78	**86.17**	85.77	90.47	**88.12**
R	**Background**	97.12	98.56	97.84	97.42	98.34	97.88	99.22	98.58	98.90
	Iris	88.60	95.48	92.04	90.95	94.74	92.85	89.14	95.55	92.35
	Sclera	94.36	97.24	95.80	95.07	97.48	96.28	95.90	97.45	96.68
	Pupil	92.43	95.81	94.12	96.12	95.41	95.77	92.54	95.50	94.02
	All classes	93.13	96.77	**94.95**	94.89	96.49	**95.69**	94.20	96.77	**95.49**
F	**Background**	98.36	99.22	98.79	98.49	99.10	98.80	98.60	99.21	98.91
	Iris	87.97	94.12	91.05	87.01	92.99	90.00	87.33	93.52	90.43
	Sclera	77.45	89.47	83.46	84.08	88.59	86.34	86.77	90.32	88.55
	Pupil	76.39	87.61	82.00	82.77	87.44	85.11	84.23	89.46	86.85
	All classes	85.04	92.61	**88.82**	88.09	92.03	**90.06**	89.23	93.13	**91.18**

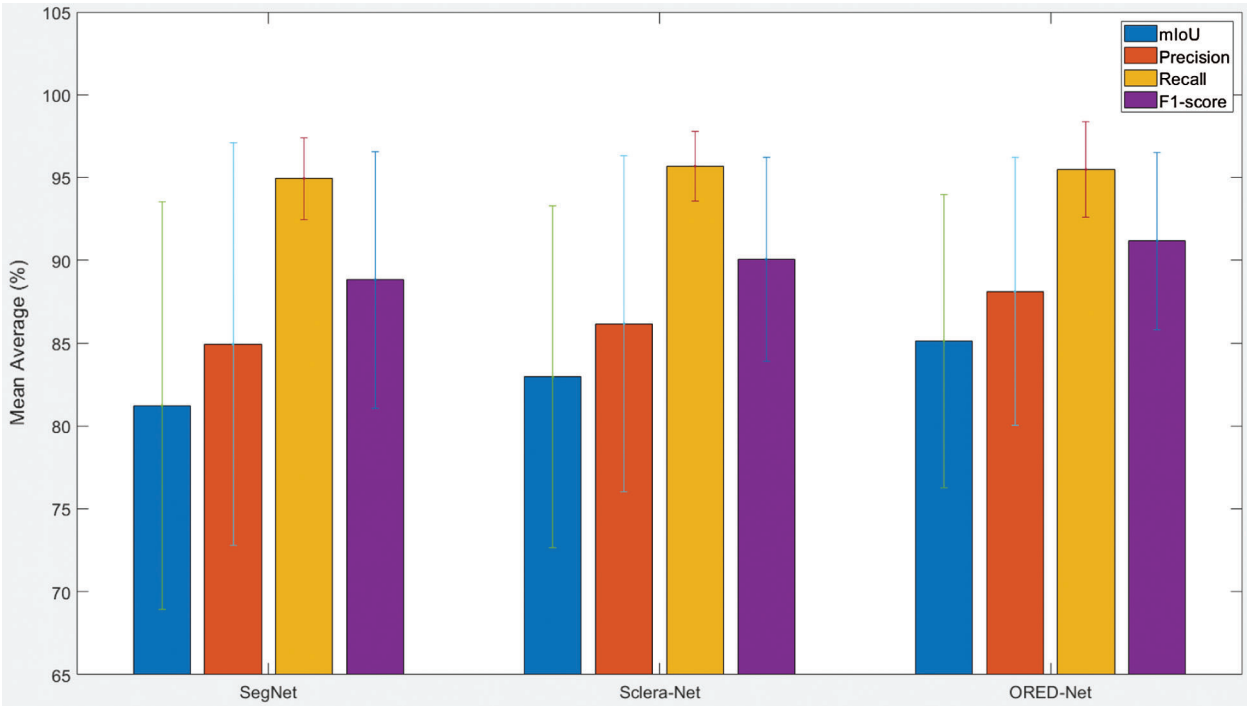


Figure 10: Mean and standard deviation of the proposed method and existing alternatives in terms of mean intersection over union, precision, recall and F1-score based on UBIRIS.v2 database

5 Conclusions

In this paper, a novel multi-class semantic segmentation network called ORED-Net was proposed for the segmentation of eye regions such as the iris, sclera, pupil, and background. ORED-Net is based on the concept of outer residual connections for transferring spatial edge information directly from the initial layers of the encoder to the decoder layers. This framework enhances the performance of the network in the case of bad quality images. ORED-Net has fewer layers, which reduces the number parameters along with the computation time. The most notable aspects of the proposed ORED-Network are that it achieves a high accuracy with a lighter network and converges in considerably fewer number of epochs with direct flow of edge information, resulting in faster training. In ORED-Net, the original image is used for both training and testing, as no extra overhead is required in the form of preprocessing. ORED-Net is the first network of its kind that simultaneously segments three important eye regions, namely iris, sclera, and pupil, without any preprocessing overhead. The robustness and effectiveness of the proposed method were tested on various publicly available databases for eye region segmentation, including the SBVPI and UBIRIS.v2 datasets. In future studies, this work will be extended to a robust multimodal biometric identification system based on multiple eye regions.

Funding Statement: This work was supported by the National Research Foundation of Korea (NRF, www.nrf.re.kr) grant funded by the Korean government (MSIT, www.msit.go.kr) (No. 2018R1A2B6009188) (received by W. K. Loh).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Z. Zhao and A. Kumar, "Accurate periocular recognition under less constrained environment using semantics assisted convolutional neural network," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1017–1030, 2017.

[2] A. S. Al-Waisy, R. Qahwaji, S. Ipson and S. Al-Fahdawi, "A robust face recognition system based on curvelet and fractal dimension transforms," in *Proc. CIT/IUCC/DASC/PICOM*, Liverpool, UK, pp. 548–555, 2015.

[3] R. Hentati, M. Hentati and M. Abid, "Development a new algorithm for iris biometric recognition," *International Journal of Computer and Communication Engineering*, vol. 1, no. 3, pp. 283–286, 2012.

[4] R. A. Naqvi and W. K. Loh, "Sclera-Net: Accurate sclera segmentation in various sensor images based on residual encoder and decoder network," *IEEE Access*, vol. 7, pp. 98208–98227, 2019.

[5] N. Susitha and R. Subban, "Reliable pupil detection and iris segmentation algorithm based on SPS," *Cognitive Systems Research*, vol. 57, pp. 78–84, 2019.

[6] P. Rot, M. Vitek, K. Grm, Z. Emersic, P. Peer *et al.,* "Deep sclera segmentation and recognition, " in *Handbook of Vascular Biometrics*. Chapter no. 13, vol. 79. Cham, Switzerland: Springer, pp. 395–432, 2020. [Online]. Available: https://www.springer.com/gp/book/9783030277307.

[7] P. Rot, Z. Emersic, V. Struc and P. Peer, "Deep multi-class eye segmentation for ocular biometrics," in *Proc. IWOBI*, San Carlos, Costa Rica, pp. 1–8, 2018.

[8] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[9] SBVPI Dataset, 2020. [Online]. Available: http://sclera.fri.uni-lj.si/database.html.

[10] H. Proenca, S. Filipe, R. Santos, J. Oliveira and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010.

[11] B. Luo, J. Shen, Y. Wang and M. Pantic, "The iBUG eye segmentation dataset," in *Proc. ICCSW*, Dagstuhl, Germany, pp. 1–9, 2018.

[12] R. A. Naqvi, S. W. Lee and W. K. Loh, "Ocular-Net: Lite-residual encoder decoder network for accurate ocular regions segmentation in various sensor images," in *Proc. BigComp,*, Busan, Korea, pp. 121–124, 2020.

[13] B. Hassan, R. Ahmed, T. Hassan and N. Werghi, "SIP-SegNet: A deep convolutional encoder-decoder network for joint semantic segmentation and extraction of sclera, iris and pupil based on periocular region suppression," *arXiv preprint arXiv:2003.00825*, 2020.

[14] C. Palmero, A. Sharma, K. Behrendt, K. Krishnakumar, O. V. Komogortsev *et al.,* "OpenEDS2020: open eyes dataset," *arXiv preprint arXiv:2005.03876*, 2020.

[15] P. Kansal and S. Devanathan, "EyeNet: Attention based convolutional encoder-decoder network for eye region segmentation," in *Proc. ICCVW*, Seoul, Korea, pp. 3688–3693, 2019.

[16] V. T. Huynh, S. H. Kim, G. S. Lee and H. J. Yang, "Eye semantic segmentation with a lightweight model," in *Proc. ICCVW*, Seoul, Korea, pp. 3694–3697, 2019.

[17] F. Yu, V. Koltun and T. Funkhouser, "Dilated residual networks," in *Proc. CVPR*, *Honolulu, HI, USA*, pp. 636–644, 2017.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, San Diego, CA, USA, pp. 1–14, 2015.

[19] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, pp. 770–778, 2016.

[20] M. Arsalan, R. A. Naqvi, D. S. Kim, P. H. Nguyen, M. Owais *et al.,* "IrisDenseNet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near infrared light camera sensors," *Sensors*, vol. 18, no. 5, pp. 1–30, 2018.

[21] C. S. Bezerra, R. Laroca, D. R. Lucio, E. Severo, L. F. Oliveira *et al.,* "Robust iris segmentation based on fully convolutional networks and generative adversarial networks," in *Proc. SIBGRAPI*, Parana, Brazil, pp. 281–288, 2018.