International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Semantic Segmentation of Ocular Regions Using Artificial Intelligence for Biometric and Medical Uses

Dr. Nirgish Kumar[1], Komal[2], Dr. C.S. Raghuvamshi[3]

[1,2,3]Computer Science & Engineering Deportment, Rama University Knapur

Abstract:
Segmenting multiple ocular regions is crucial for a variety of applications, including gaze estimation, liveness detection, biometrics, and healthcare. Segmentation methods usually concentrate on one area of the eye at a time. Very little research has been done on some areas of the eye, despite the many clear benefits. Similarly, in difficult situations with blur, ghost effects, low resolution, off-angles, and odd glints, precise segmentation of several eye regions is required. These limitations cannot currently be addressed by the segmentation techniques that are available. This research proposes a lightweight outer residual encoder-decoder network that can be used with a variety of sensor images in order to accurately segment multiple eye regions in unconstrained circumstances. The suggested technique uses the high-frequency information flow from the outer residual encoder-decoder deep convolutional neural network (named ORED-Net) to identify the actual borders of the eye regions from low-quality photos. Furthermore, the performance of the suggested ORED-Net model is not enhanced by network depth, complexity, or parameter count. The suggested network weighs significantly less than earlier cutting-edge models. Extensive studies were conducted utilizing the SBVPI and UBIRIS.v2 datasets, which contain images of the eye Semantic Segmentation of Ocular Regions Using Artificial Intelligence for Biometric and Medical Uses The Tech Science Pressgion, and ideal performance was attained. The mean intersection over union score (mIoU) of the simulation results produced by the suggested ORED Net on the difficult SBVPI and UBIRIS.v2 datasets was 89.77 and 87.27, respectively.

Keywords: Deep Learning, Sensors, Biometrics, Medical Uses, Semantic Segmentation, and Ocular Regions.

Introduction:
Researchers have been actively studying other periocular regions, such as the sclera and retina, to gather identity cues that might be helpful for stand-alone recognition systems or to supplement the information typically used for iris recognition, even though ocular traits other than the iris are less commonly studied [1]. Periocular regions, such as the sclera and retina, to gather identity cues that might be helpful for stand-alone recognition systems or to supplement the information typically used for iris recognition, even though ocular traits other than the iris are less Co-research on biometrics, liveness detection, and gaze estimating systems that depend on characteristics of the iris, sclera, pupil, or other periocular regions has advanced significantly in the past few decades [2]. Due to the substantial market potential and the importance of ocular area applications, interest in these features is growing daily. Nowadays,

biometric technology is an essential component of our daily life, as these techniques, in contrast to traditional methods, do not require a person to carry or memorize any information, including IDs, passwords, or pins [3]. The varied and distinctive textures of the iris, including rings, crypts, furrows, freckles, and ridges, have attracted a lot of interest from the research community in iris segmentation [4]. One ocular region at a time, such as the iris, pupil, sclera, or retina, was the only focus of the majority of earlier research studies on eye region segmentation. In multi-class segmentation, a single segmentation network is used to segment multiple eye regions from the provided input image. Despite several benefits in various applications, surprisingly few researchers have created multi-class segmentation approaches for the ocular regions. Specifically, employing multiple region segmentation can preserve or even improve segmentation performance under difficult circumstances since the targeted region can offer helpful contextual information about other nearby regions [5]. An additional possible benefit is the ability to implement multi-biometric systems without incurring expenses or computational overheads, which can effectively segment numerous target classes using a single method [6]. For instance, the iris region's border can reveal important details regarding the sclera and pupil regions' boundaries. In a similar vein, the sclera region is constrained by the eyelash area [7].

In this study, as illustrated in Fig. 1, we try to fill the research gaps in the segmentation of several eye areas utilizing a single network. With a single model, the suggested network may divide the input eye image into four primary classes that represent the iris, sclera, pupil, and background area. Deep learning convolutional neural network (CNN) models have advanced quickly in recent years and are now a popular technique for image processing jobs. CNNs have proven to be more effective than traditional techniques in a variety of fields, including satellite image processing and medicine. The suggested approach is predicated on convolutional encoder-decoder networks, which are deep learning models for semantic segmentation in images. The recently unveiled SegNet architecture serves as the foundation for this design methodology [8]. The outer residual encoder-decoder network served as the foundation for the development of ORED-Net. By using solely non-identity outer residual pathways from the encoder to the decoder, the suggested network achieves a higher accuracy with a lower network depth and fewer parameters and layers.

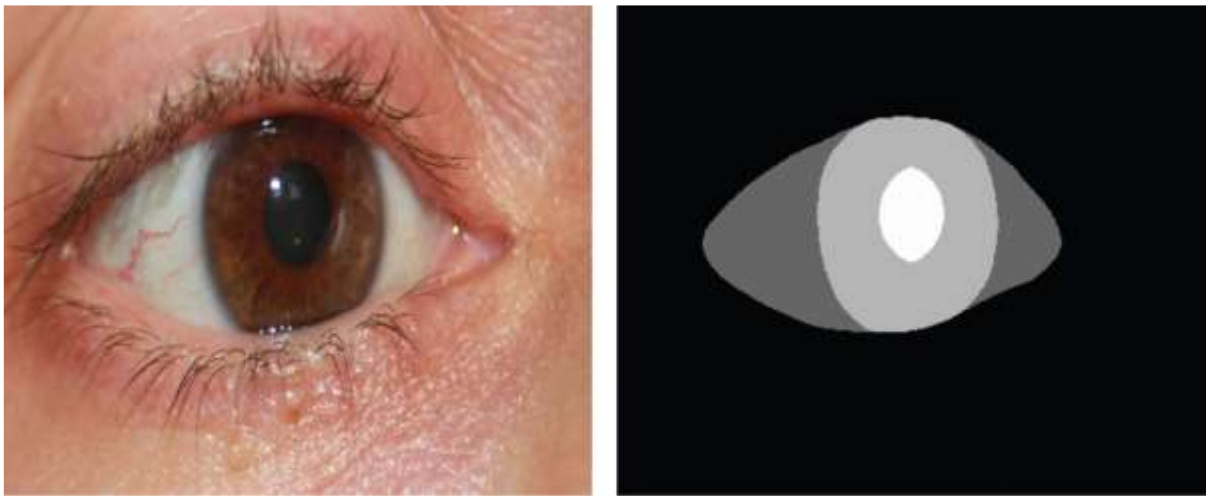


Figure 1: Multi-class eye segmentation sample images are shown in the input eye image on the left and the ground truth image on the right.

ORED-Net is unique in the four ways listed below:

- ORED-Net is a semantic segmentation network that does not use traditional image processing techniques and has no preprocessing overhead.
- ORED-Net is a stand-alone network designed for multi-class ocular area segmentation.
- ORED-Net reduces information loss by using residual skip connections between the encoder and the decoder. This enables high-frequency information to pass through the model, increasing accuracy with a few layers.
- The suggested ORED-Net model's performance was evaluated using publicly available datasets gathered in a variety of settings.

Results for the iris, sclera, pupil, and backdrop classes are presented in this study using the SBVPI [9] and UBIRIS.v2 [10] datasets. Furthermore, the suggested model is contrasted with cutting-edge methods found in the literature. The outcomes show that the suggested approach is the best one for ocular segmentation, which can be used in recognition processes. This is how the remainder of the paper is organized. A synopsis of relevant material is given in Section 2. The suggested methodology and process are explained in Section 3. Section 4 discusses the review and analysis's findings. Section 5 concludes with a presentation of future projects.

Review of Literature :

Few 2 Review of Literature Few studies have concentrated on multi-class eye segmentation, especially when it comes to using a single segmentation model to segment different eye regions from the provided images. Segmenting multi-class eye areas using the popular convolutional encoder-decoder network SegNet was recently reported by Rot et al. [7]. The iris, sclera, pupil, eyelashes, medial canthus, and periocular region were among the various eye regions they examined for segmentation. This work necessitated post-processing using an atrous CNN with the conditional random field described in Luo et al. [11] and a thresholding method on probability maps. The Multi-Angle Sclera Database (MASD) was used to retrieve the data. The Ocular-Net CNN was proposed by Naqvi et al. to segment the iris and sclera, among other eye regions. Non-identity residual paths in a lighter encoder and decoder version make up this network. To improve the model's performance, residual shortcut connections were used as the network depth increased [12]. Furthermore, separate databases were used to assess the iris and sclera. The SIP-SegNet CNN was proposed by Hassan et al. for integrated semantic segmentation of the pupil, iris, and sclera. The original image was denoised using a denoising CNN (DnCNN). Following DnCNN denoising, SIP-SegNet used contrast limited adaptive histogram equalization (CLAHE) to remove reflections and improve images. After that, adaptive thresholding was used to retrieve the periocular information, and fuzzy filtering was used to suppress this information. Lastly, several eye areas were semantically segmented using a densely linked fully convolutional encoder-decoder network [13]. The suggested approach was evaluated using a variety of metrics and tested on the CASIA sub-datasets.

The Eye Segmentation challenge to segment important eye regions with the goal of creating a generalized model with the least amount of complexity in terms of model parameters. The OpenEDS dataset, a sizable collection of eye photos taken by a head-mounted display equipped with two synchronized eye-facing cameras, was used for the experiments [14]. The iris, sclera, pupil, and backdrop are the four distinct eye regions that Kansal et al. [15] suggested Eyenet, an Attention-based Convolutional Encoder-Decoder Network, to accurately segment in order to address the problem of semantic segmentation of eye regions. Eyenet relies on residual connections in the encoder and decoder that are not reliant on identity mapping. There are two kinds of attention units and It was suggested to

use multiscale supervision in order to provide precise and distinct boundary eye regions. Huynh et al. demonstrated eye segmentation using a lightweight model. In their method, the input image was converted to grayscale, the eye regions were segmented using a deep network model, and the erroneous portions were eliminated using heuristic filters. To lower the false positive in the model's output, a heuristic filter was applied [16].

A comparison of the suggested approach with alternative techniques for the multi-class segmentation of ocular regions, together with an analysis of their advantages and disadvantages, is shown in Tab. 1. Table 1: Evaluation of the suggested approach in relation to alternative multi-class segmentation techniques

Table:1 A comparison between the suggested approach and alternative multi-class segmentation techniques Methods:

Methods	Strengths	Weaknesses
Seg Net architecture-based deep multi-class eye segmentation [7]	Several eye areas are segmented using a single model	*A significant portion of the training data is synthetic
Ocular-Net, a lighter residual encoder-decoder network [12]	There is residual connection between neighboring convolutional layers	*There is some post-processing involved *The technique is trained independently for every area
Eye region joint semantic segmentation, SIP Seg Net [13]	The original images are denoised using Dn CNN	*Only one area of the eye is treated at a time *The original image undergoes thoughtful preparation.
A depthwise convolution operation serves as the foundation for the encoder-decoder structure [16].	It has a minimal computational cost and can be implemented in real time on any system	*Suppression of the periocular area is necessary * Heuristic filtering is used for post-processing
The The ORED-Net (Proposed Method) is an outer residual encoder-decoder network ORED-Net (Proposed Method) is an outer residual encoder-decoder network	Outer residual skip pathways from the encoder to the decoder minimize information loss. Additionally, the outer residual pathways shorten the training time	* Only the Open EDS dataset was used to train and test the technique. * Thorough training is necessary

.

3: Suggested Approach for Segmenting Eye Regions

3.1 Summary of the Suggested Model

Fig. 2 displays the flowchart of the suggested ORED-Net for semantic segmentation of various eye areas. Based on non-identity residual connections between the encoder and decoder networks, the suggested network is a fully convolutional network.

The image input is entered into the convolutional network receives the input image without any upfront preprocessing overhead. The suggested ORED-Net for multi-class segmentation of the complete input eye pictures includes both an encoder and a decoder. While the decoder does the opposite, the encoder's job is to down sample the input image until it can be represented in terms of very small features. With the help of the encoder's tiny characteristics, the decoder upsamples the image to its original size. Along with the opposite procedure of The decoder's ability to predict many classes—the iris, sclera, pupil, and background—is another crucial function of downsampling. A pixel classification layer and the Softmax loss function are used to complete the prediction task. The pixel classification layer predicts the class of each pixel in the image and assigns the appropriate label.

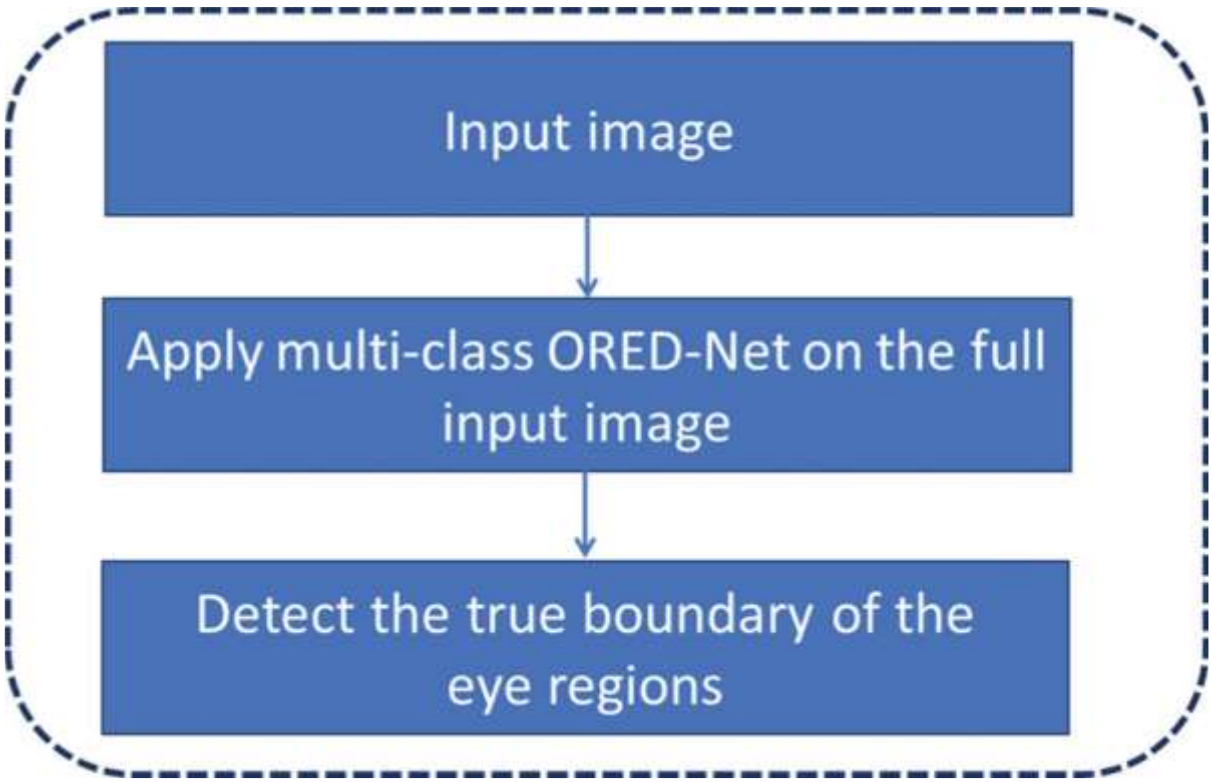


Figure 2: Flowchart of the suggested technique for segmenting the ocular areas

3.2 Using ORED-Net to Segment Several Eye Regions.

The high-frequency contextual information is essentially degraded in standard encoder-decoder networks since the image is down sampled and represented by relatively small features. When the image is divided into patches of size 7 × 7, this leads to the vanishing gradient problem for the categorization of image pixels [17]. Identity and non-identity mapping residual blocks were introduced in order to solve the vanishing gradient problem. A CNN's accuracy is higher than that of basic CNNs like VGG Net when a residual block is included [18]. Residual building blocks (RBBs) are generally

The foundation of (RBBs) is identity and non-identity mapping. In identity mapping, the residual operation is carried out by directly providing the features for element-wise addition. In contrast, each RBB undergoes a 1 × 1 convolution prior to the features being added element-wise in the case of non-identity mapping. The suggested network does not take identity mapping into account. Rather, as illustrated in Fig. 3, a 1 x 1 convolution layer uses outer residual routes from the encoder to the decoder to carry out non-identity mapping.

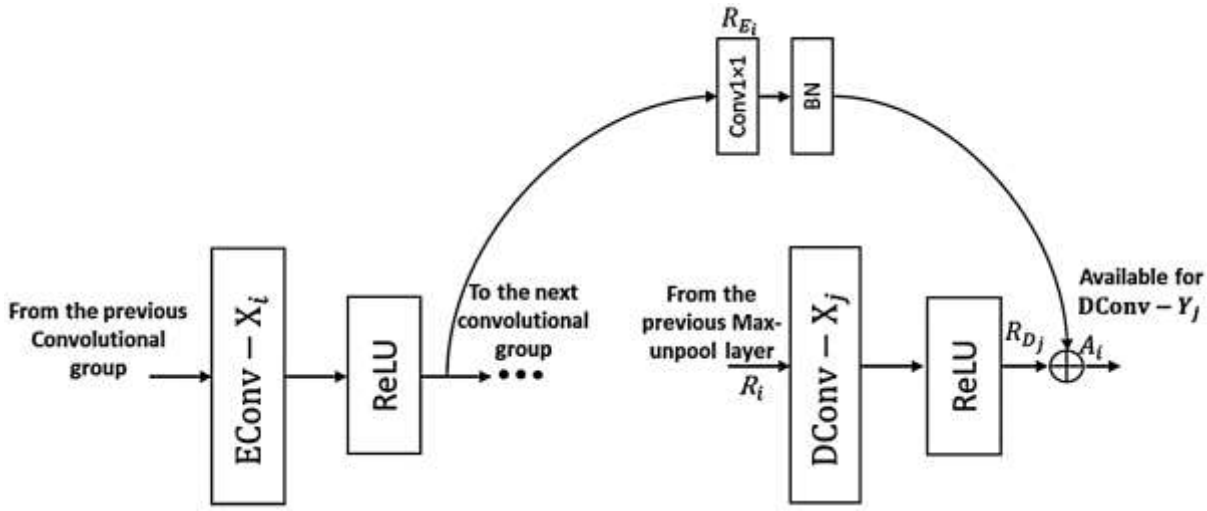


Figure 3: The suggested method's utilization of residual building blocks (RBBs)

In contrast to simple encoder-decoder networks, the suggested ORED-Net is implemented through various developmental stages to accomplish the multi-class segmentation task with high accuracy. Seg Net-Basic, a popular segmentation network, is used in the first stage [8]. In both the encoder and decoder sections, Seg Net-Basic has thirteen convolutional layers. Five convolutional layers are eliminated from the encoder and decoder portions of the network to simplify it to its most basic form. Because of this, the encoder and decoder portions of the suggested network only contain eight convolutional layers. A lightweight encoder and decoder convolutional network is produced by the addition of two convolutional layers to each group in the encoder and decoder architectures. The high-frequency features are empowered thanks to ORED-Net. Non-identity residual connections are added from the encoder side layers to the matching layers on the decoder side in the following stage of the proposed ORED-Network's preparation. The encoder side layers to the appropriate decoder side layers via the outer residual routes, as illustrated schematically in Figure 4. As a result, ORED-Net's residual connectivity differs from that of the original Res Net [19] and other residual-based networks that have been proposed, like Sclera-Net [4]. The primary distinctions between the suggested network and previously published networks like Res Net [19] and Sclera-Net [4]

Tab. 2. Table 2: ORED-Net's primary architectural distinctions from other residual-based techniques

ResNet [19]	Sclera-Net [4]	ORED-Net
ResNet employs a limited number of residual connections that are not identity mapping and	The encoder and decoder's convolutional layers include residual connectivity depending	The encoder and decoder's convolutional layers lack internal residual connectivity.

a high number of identity mapping connections	on identity and non-identity.	
ResNet only connects nearby layers via the skip path connection.	The encoder and decoder do not have any external skip path connections.	Non-identity residual connections make up the outer skip path connections between the encoder and the decoder
Each block of several ResNet variations, including ResNet-50/101/152, contains a 1x1 convolutional layer	The encoder-decoder network as a whole consists of 6 identity and 8 non-identity residual connections.	From the encoder to the decoder, there are four non-identity residual routes.
Since a ReLU is utilized following the elementwise addition, several ResNet variations, including ResNet-18/34/50/101, are predicated on post-activation.	The elementwise addition is followed by a ReLU in the overall network..	The elementwise addition is preceded on the decoder side by a ReLU. ORED-Net thus makes use of pre-activation.
Average pooling is used at the end of each convolutional layer	Sclera-Net thus makes use of post-activation. In the encoder and decoder networks, residual connections are added right after max pooling and unpooling, respectively.	All of the convolutional blocks employ the max-pooling layer to give the decoder index information.

Fig. 4 depicts ORED-Net's general architecture. Four non-identity outer residual paths from the encoder to the decoder are shown here: Outer-Residual-Path-1, ORED-P-1 to Outer-Residual-Path-4, ORED-P-4. Conv + BN represents the group that contains a 3 x 3 convolutional layer and batch normalization layers, while ReLU represents the activation layer, or rectified linear unit. The max pooling layer is represented as Max-pool, the max unpooling layer is represented as Max-unpool, and the combination of a convolution layer of size 1 × 1 and batch normalization layers is represented as 1 × 1 Conv + BN. Each of the encoder's four convolutional groups—E-Conv-X and E-Conv-Y—consists of two convolutional layers prior to each Max-pool. Similar to this, the decoder has four convolutional groups, each of which has two convolutional layers (D-Conv-X and D-Conv-Y) following each Max-unpool layer. Consequently, E-Conv-Xi represents the first convolutional layer of the i-th encoder of the convolutional group, and the second convolutional layer D-Conv-Yj is the representation of the second convolutional layer of the j-th decoder of the convolutional group. In this case, i and j have values between 1 and 4. Through ORED-Path-1, the first encoder-decoder convolutional groups at the network's far left and right are connected. Similarly, as illustrated in Fig. 4

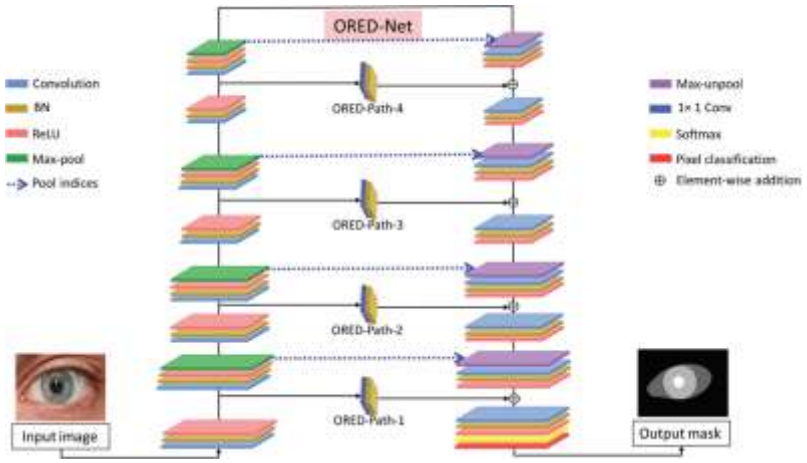


Figure 4: Deep learning-based eye region segmentation system with light-residual encoder and decoder

4, the second convolutional groups, which are positioned second from the left and right sides of the convolutional group, are connected via ORED-Path-2. Eye region segmentation system using deep learning and a light-residual encoder and decoder network According to Fig. 4, the residual features RE1 and RD1 are added element-to-element to the second convolutional layer in the first convolutional group at the decoder part. These features come from the first convolutional layer in the encoder convolutional-group-1 (E-Conv-X1) after the ReLU and the first convolution layer in the first decoder convolutional group-1 (D-Conv-X1) after the ReLU, respectively, via ORED-Path-1. The following equation can be used to describe this:

$$A1 = R_{E1} + R_{D1} \quad (1)$$

Here, A1 is the residual feature that is fed into D-Conv-Y1 via ORED-Path-1 with element-to-element addition. Generally, the following equation can be used to represent the outer residual block in Fig. 3:

$$Ai = R_{Ei} + R_{Dj} \quad (2)$$

where RDj is the residual feature obtained from the first convolutional layer of the i-th convolutional group (E-Conv-Xi) following the ReLU at the encoder section, REi is the residual feature from the first convolutional layer of the i-th convolutional group (E-Conv-Xi), and Ai is the total of the features supplied to D-Conv-Yj by the outer residual connection.

the residual features from the j-th convolutional group's first convolutional layer (D-Conv-Xj) following the decoder side's ReLU. Additionally, i and j have values between 1 and 4. The residual characteristics REi from each of the convolutional groups are thus provided from the encoder side to the decoder side by each of the four outer residual routes (ORED-P-1 to ORED-P-4) in order to improve the network's capacity for robust segmentation. The residual features of the decoder side, or RDj features, are strengthened by this direct transmission of the spatial edge information from the encoder side.

3.2.1 The ORED-Net Encoder

Figure 4 illustrates that the encoder is made up of four convolutional groups, each of which has two convolutional layers in addition to the batch normalization and ReLU activation layers. The fundamental and unique feature of the ORED-Net encoder is that the residual pathways provide the spatial information to the next decoder group. On the encoder side, these outside residual routes start after every ReLU layer. When compared to other networks used for the same purpose, a lighter network can produce superior outcomes because of the exterior residual connections. Through the Max-pool layers of

ORED-Net, the key features are downsampled and give the decoder side pooling indices. The feature map size and index information, which are necessary on the decoder side, are contained in the pooling indices. Tab. 3 displays the ORED-Net encoder structure. Through the non-identity residual link seen in Figure 4, four outside residual encoder-decoder paths are visible, connecting the encoder and the decoder. Using the spatial information of the previous layers, these outer residual encoder-decoder non-identity residual connections enhance features. On the encoder side, the outer residual encoder-decoder connections begin after the ReLU activation layer, and on the decoder side, they terminate adjacent to the ReLU activation layer. Since summing is carried out on the decoder side after every ReLU layer, the suggested network makes use of pre-activation. ORED-Net is a balanced network since each convolutional group on the encoder and decoder sides has an equal number of convolutional layers, or two convolutional layers.

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†† To decoder 1 × 1 × 128/ORED-P-2† 3 × 3 × 128/E-Conv-2_2†	128	112 × 112 × 128
Pool-2	2 × 2/Pool-2	-	64 × 64 × 128
EC-G-3	3 × 3 × 128/E-Conv-3_1†† To decoder 1 × 1 × 256/ORED-P-3† 3 × 3 × 256/E-Conv-3_2†	256	64 × 64 × 256
Pool-3	2 × 2/Pool-3		64 × 64 × 128
EC-G-4	3 × 3 × 256/E-Conv-4_1†† To decoder 1 × 1 × 512/ORED-P-4† 3 × 3 × 512/E-Conv-4_2††	512	32 × 32 × 512
Pool-4	Pool-4/2 × 2	-	16 × 16 × 512

In this work, two-fold cross-validation was performed The ORED-Net encoder with outer residual routes based on a 224 × 224 × 3 picture is shown in Tab. 3. The encoder convolution layers, outer residual encoder-decoder routes, and pooling layers are denoted by E-Conv, ORED-P, and Pool, respectively. The encoder's convolutional layers, denoted by the symbol "††," comprise both the batch normalization (BN) and ReLU activation layers, whereas the convolution layers denoted by "†" only contain the BN layer. ORED-P-1 to ORED-P-4, which stand for outer residual encoder-decoder skip paths, begin at the

encoder and send edge information to the decoder. The ReLU activation layer is employed before the element-wise addition since the suggested model incorporates pre-activation.

3.2.2 Decoder for ORED-Net

The ORED-Net decoder's architecture, as depicted in Figure 4, is designed to mirror the encoder and carry out a convolutional operation that is comparable to the encoder's. The size of the feature map is maintained by the decoder using the size information and indices that are supplied by the encoder's pooling layers. To guarantee that the network output's size matches the input image's, the decoder features are also upsampled. Additionally, the ORED-Net decoder receives the features from the outer residual routes. ORED-P-1 through ORED-P-4, the four outer residual encoder-decoder routes, all start on the encoder side and terminate on the decoder side. As illustrated in Fig. 4, the addition layers (Add-4 to Add-1) execute element-to-element addition between the ORED-P and prior convolution, producing features that are helpful to the convolutional layers in the following group. As four classes—the iris, sclera, pupil, and background—are assessed for the segmentation job in this study, the decoder generates four masks that correspond to these classes, or the number of filters for the decoder's final convolutional layer. The network's pixel-wise prediction is made easier by the Softmax and pixel categorization layers. The outer residual path is immediately ended following each ReLU activation layer in order to implement post activation in the decoder. The ORED-Net output is a mask for each class, producing "0" for the BG class, "100" for the sclera class, and "180" for their respective classes. "250" for the student class and "250" for the iris class.

4 Findings and Conversation

The suggested model was trained and tested in this work using two-fold cross-validation. In order to achieve this, the gathered database was randomly divided into two subsets from the available photos. Two subsets were generated from the 55 participants' photos; 28 of the participants' data were used for training, and the 27 participants' data were utilized for testing. The training data was augmented in order to prevent overfitting problems. A desktop computer equipped with an Intel® CoreTM (Santa Clara, CA, USA) i7-8700 CPU @3.20 GHz, 16 GB of RAM, and an NVIDIA GeForce RTX 2060 Super graphics card (2176 CUDA cores and 8 GB GDDR6 memory) was used to train and test ORED-Net. The investigations described above were carried out with MATLAB R2019b.

4.1 ORED-Net Training

In order to convey spatial information from the encoder side to the decoder side, ORED-Net relies on outer residual pathways from the encoder to the decoder. Consequently, high frequency data passes through the convolutional network, enabling training of this data without the need for preprocessing. Original photos without any preprocessing or enhancement were used to train ORED-Net, and a traditional stochastic gradient descent (SGD) technique was utilized as an optimizer. SGD reduces the discrepancy between expected and actual results. A mini-batch size of five was chosen for the ORED-Net design because of its minimal memory, and the suggested model ran the full dataset 25 times, or 25 epochs, during network training.

ORED-Net design because it requires little memory. The size of the database dictated the size of the mini-batch. As indicated by Eqs. (3) and (4), one epoch was counted after training with the complete dataset.

$$Ui+1:=mui- xnv_i\text{-}n_i \{ \frac{\partial Si\,(V)}{\partial V} \} v_i > T_i \qquad (3)$$

$$V_i+1 := v_i + u_i+\mathbf{1} \qquad (4)$$

In Eqs. (3) and (4), m is the momentum, g is the learning rate, x is the weight decay, vi is the learned weight at the I[th] iteration, and ui is the momentum variable. $\{\frac{\partial Si\,(V)}{\partial V}\}v_i > T_i$ provides the average over the ith batch Ti of the object's derivative with regard to v, assessed at v. Making use of the SGD Eqs. 3 and 4's ideal training parameters, m, g, and x, were determined to be 0.9, 0.001, and 0.0005, respectively, using the @v technique. Because of the outside residual connections between the encoder and the decoder, the ORED-Net model converges fairly quickly. Consequently, only 25 epochs were used to train *the* ORED-Net model. Throughout the 25 training epochs, the mini-batch size was limited to 5 photos, with shuffling occurring after each epoch. Here, the cross-entropy loss published [8] was used to compute the training loss based on the picture pixels in the mini-batch. The cross-entropy loss over all the pixels available in the candidate mini-batch according to the iris, sclera, pupil, or background classes served as the basis for the loss computation. Furthermore, a greater disparity in pixel counts had an impact on network accuracy and convergence.

According to Arsalan et al. [20], a greater disparity in the quantity of pixels in various classes and the network's bias towards learning the dominant class had an impact on the network's accuracy and convergence. By using an inverse frequency weighting strategy, as specified in Eqs. (5) and (6), the imbalance between the classes can be eliminated during class training.

$$\text{Freq} = \frac{\text{Pixels(i)}}{\text{Total Pixels}} \quad (5)$$

$$\text{Classes Weights} = \frac{1}{\text{Freq}} \quad (6)$$

Weights for Classes = 1 Freq. (6) The total number of pixels in the training data that belong to class ð is denoted by Pixels(i). The four groups of this study—iris, sclera, pupil, and background—are represented by ð = 4.

4.2 ORED-Net Testing

4.2.1 Measures of Evaluation

The average segmentation error (Erravg), mean intersection over union (mIoU), precision (P), recall (R), and F1-score (F) were chosen as assessment methods in order to verify and contrast ORED-Net with earlier models.

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

$$\text{mI}_o\text{U} = \frac{1}{Nc}\left[\sum_{\{i=1\}}^{Nc}\left(\frac{Nxx(i)}{Nxx(i)+Nxy(i)+Nyx(i)}\right)\right] \quad (8)$$

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

$$P = \frac{Nxx}{Nxx+Nxy} \quad (9)$$

$$R = \frac{Nxx}{Nxx+Nyx} \quad (10)$$

$$F=\frac{2RP}{R+P} \quad (11)$$

Nxx is the true positive, where the number of pixels projected as x also belong to class x, and Nc is the total number of classes. The other terms are also characterized as false positives (Nxy), false negatives (Nyx), and true negatives (Nyy).

4.2.2 ORED-Net Segmentation Outcomes for Eye Regions

The multi-class eye area segmentation of eye pictures acquired using ORED-Net for the SBVPI dataset is shown in Figs. 5 and 6 with the correct and wrong findings. These visual representations adhere to the tradition that each class is represented by FP (black), FN (yellow), and TP (green, blue, and red for

According to Arsalan et al. [20], a greater disparity in the quantity of pixels in various classes and the network's bias towards learning the dominant class had an impact on the network's accuracy and convergence. By using an inverse frequency weighting strategy, as specified in Eqs. (5) and (6), the imbalance between the classes can be eliminated during class training. Pixels(i) = Freq. Total Pixels (5) Weights for Classes = 1 Freq. (6) The total number of pixels in the training data that belong to class ð is denoted by Pixels(i). The four groups of this study—iris, sclera, pupil, and background—are represented by ð =

4.2.3 ORED-Net Compared to Other Techniques:

ORED-Net's segmentation performance was evaluated against earlier approaches using the Erravg, mIoU, P, R, and F metrics listed in Section 4.4.1. A comparison of ORED-Net's segmentation performance for the SBVPI dataset with those of other approaches is shown in Tab. 4. Based on the values of Erravg, mIoU, P, R, and F, the results show that ORED-Net performs better than the existing techniques for segmenting the eye region. In Tab. 4, comparisons with the state-of-the-art techniques are shown for the iris, sclera, pupil, and background areas. Additionally, bar graphs in Fig. 7 display the results of mIoU, P, R, and F in Tab. 4.

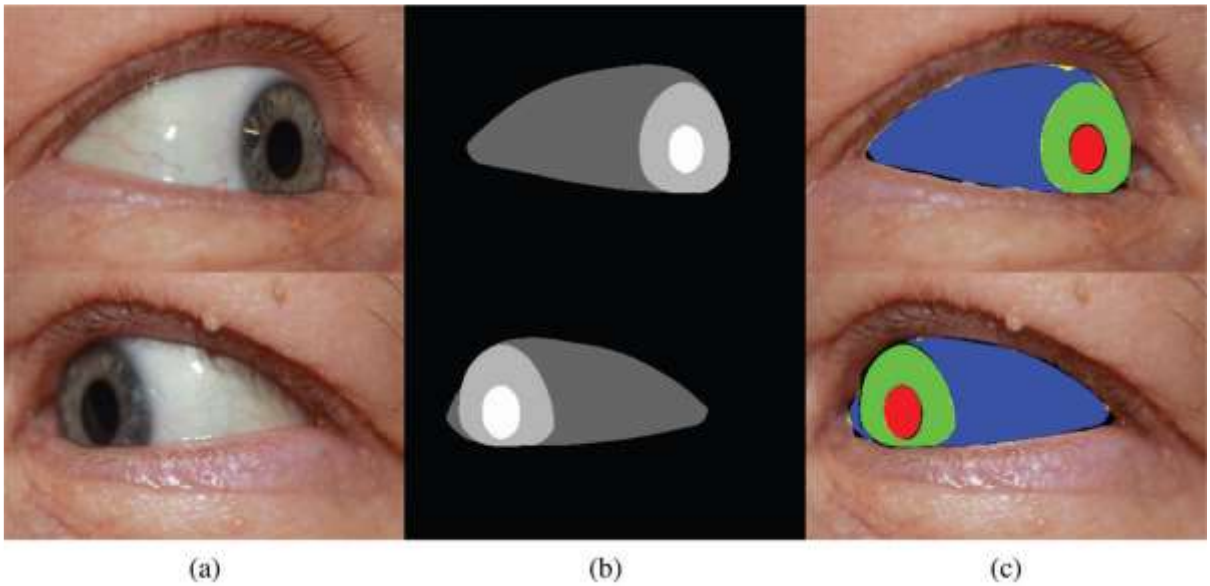
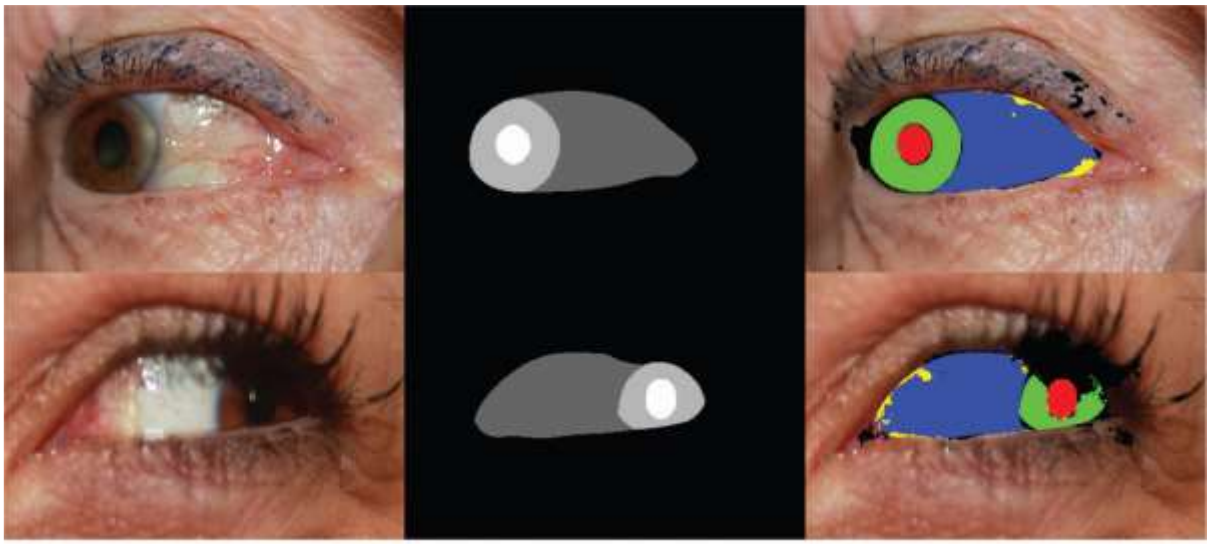


Figure 5: Illustrations of effective ORED-Net eye region segmentation for the SBVPI dataset (a) The original image, (b) the ground-truth mask, and (c) the ORED-Net predicted mask result



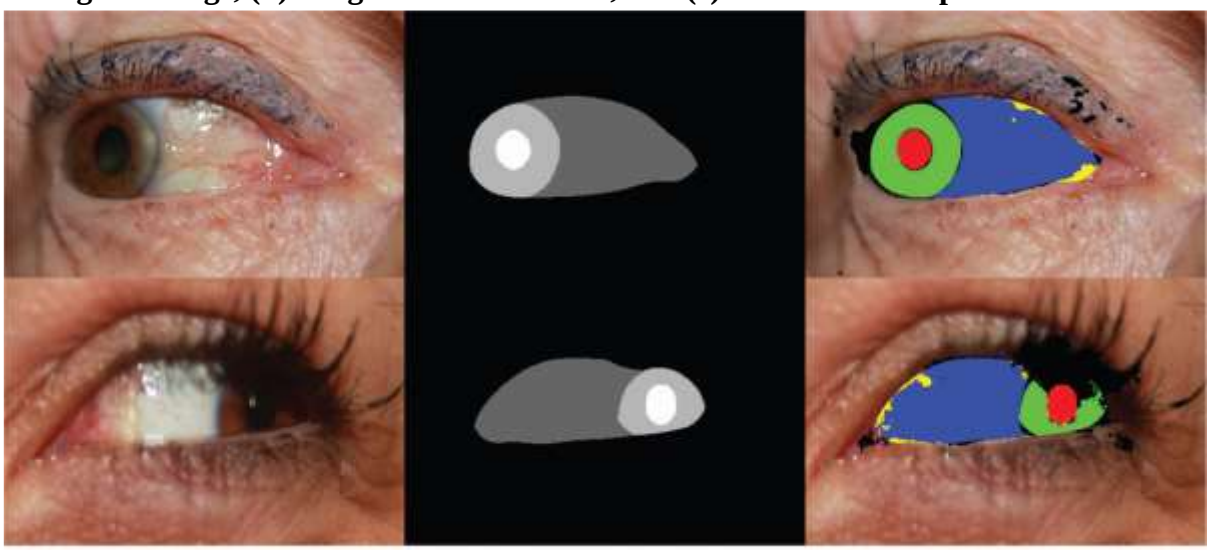
(a) (b) (c)

Figure 6: ORED-Net segmentation examples for the SBVPI dataset's poor eye regions: (a) The original image, (b) the ground-truth mask, and (c) the ORED-Net predicted mask result

4.2.4 ORED-Net-Based Eye Region Segmentation with Additional Open Datasets

This study included trials with another publically accessible dataset for eye area segmentation, namely the UBIRIS.v2 dataset, to assess the segmentation performance of ORED-Net under various picture acquisition settings [10]. Only 300 photos had iris and sclera masks available in earlier research [21]. To test the suggested ORED-Net model on the iris, sclera, and pupil using the UBIRIS.v2 dataset, the ground truth images of the iris and sclera were combined, and the ground truths for the pupil were created. Fifty percent of the 300 photos in the UBIRIS.v2 dataset were used for training, while the other fifty percent (150) were utilized for two-fold cross-validation. The UBIRIS.v2 dataset was utilized to train ORED-Net via data augmentation similar to that used for the SBVPI dataset.

Figure 5: Illustrations of effective ORED-Net eye region segmentation for the SBVPI dataset (a) The original image, (b) the ground-truth mask, and (c) the ORED-Net predicted mask result



(a) (b) (c)

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	87.76	90.98	**89.77**
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

Figure 6: ORED-Net segmentation examples for the SBVPI dataset's poor eye regions: (a) The original image, (b) the ground-truth mask, and (c) the ORED-Net predicted mask result

4.2.4 ORED-Net-Based Eye Region Segmentation with Additional Open Datasets

This study included trials with another publically accessible dataset for eye area segmentation, namely the UBIRIS.v2 dataset, to assess the segmentation performance of ORED-Net under various picture acquisition settings [10]. Only 300 photos had iris and sclera masks available in earlier research [21]. To test the suggested ORED-Net model on the iris, sclera, and pupil using the UBIRIS.v2 dataset, the ground truth images of the iris and sclera were combined, and the ground truths for the pupil were created. Fifty percent of the 300 photos in the UBIRIS.v2 dataset were used for training, while the other fifty percent (150) were utilized for two-fold cross-validation. The UBIRIS.v2 dataset was utilized to train ORED-Net via data augmentation similar to that used for the SBVPI dataset.

Table 4: Evaluation of the suggested approach against current approaches for the SBVPI dataset (unit: %)

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.

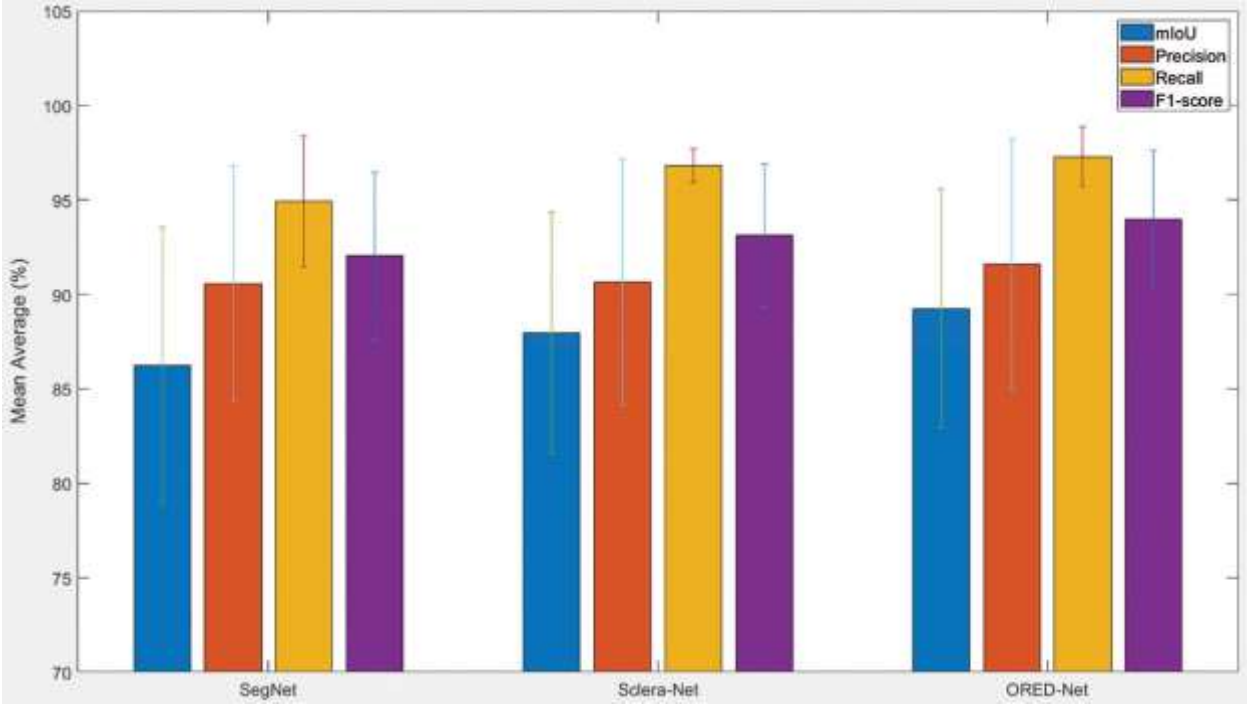


Figure 7: Mean and standard deviation of the proposed method and existing alternatives in terms of mean intersection over union, precision, recall and F1-score based on SBVPI database

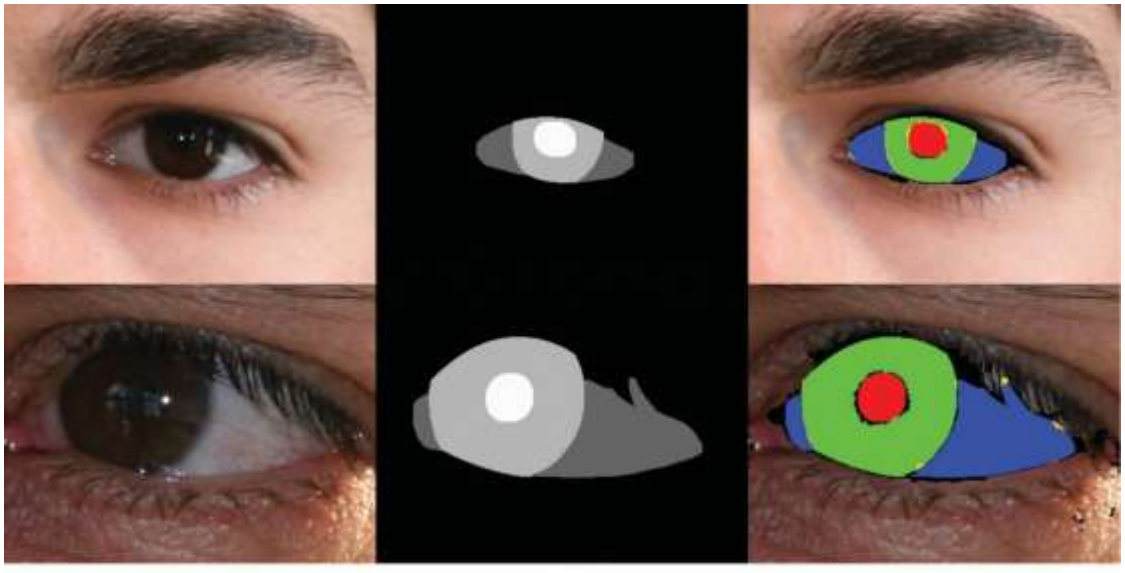


Figure 8: Examples of good eye region segmentation by ORED-Net for the UBIRIS.v2 dataset: (a)

Original image, (b) Ground-truth mask, and (c) Predicted mask result obtained with ORED-Net

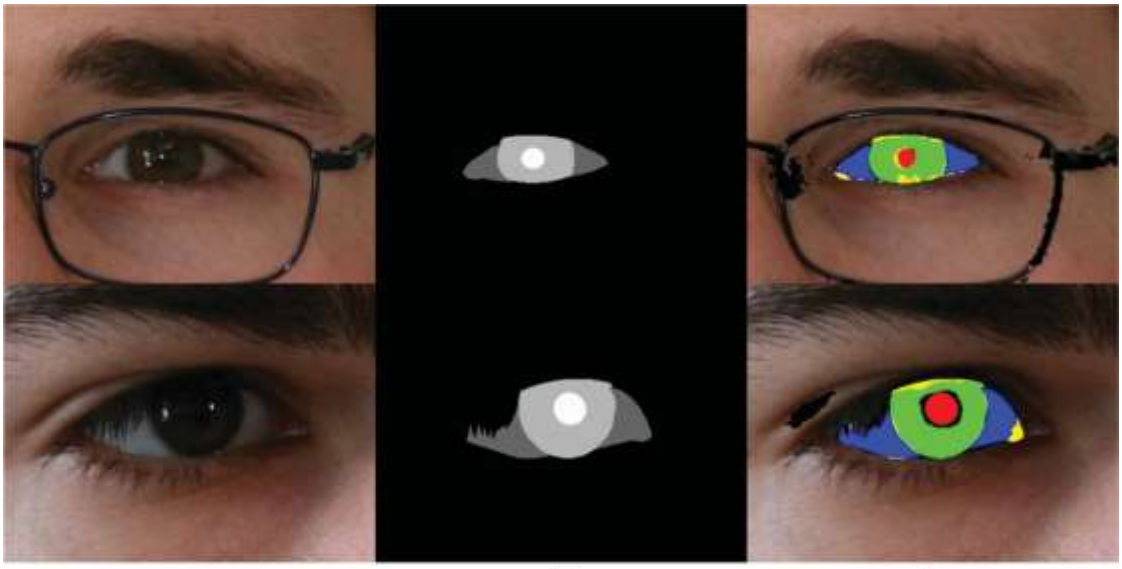


Figure 9: Examples of bad eye region segmentation by ORED-Net for the UBIRIS.v2 dataset: (a) Original image, (b) Ground-truth mask, and (c) Predicted mask result obtained with ORED-Net

For the UBIRIS.v2 dataset, Tab. 5 compares the segmentation performance of ORED-Net with that of other approaches. Additionally, bar graphs in Fig. 10 display the results of mIoU, P, R, and F in Tab. 5. The performance of the suggested ORED-Net framework is comparable to state-of-the-art algorithms, according to the findings shown in
Tabs. 4 and 5 (Figs. 7 and 10). Notably, ORED-Net is a revolutionary approach that, in contrast to all other existing algorithms that only handle one or two eye areas at a time, performs multi-class semantic segmentation of many eye regions, such as the pupil, iris, and sclera, simultaneously. Furthermore, as indicated in Tabs. 4 and 5 (Figs. 7 and 10), the ORED-Net model's performance was assessed on various publically accessible datasets for comparisons with alternative approaches.

Table 5: Evaluation of the suggested ORED-Net approach against current techniques for the UBIRIS.v2 dataset (Unit:%)

Evaluation Metrics		SegNet [8]			Sclera Net [4]			ORED-Net		
	Classes	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
	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	84.10	90.17	**87.27**
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

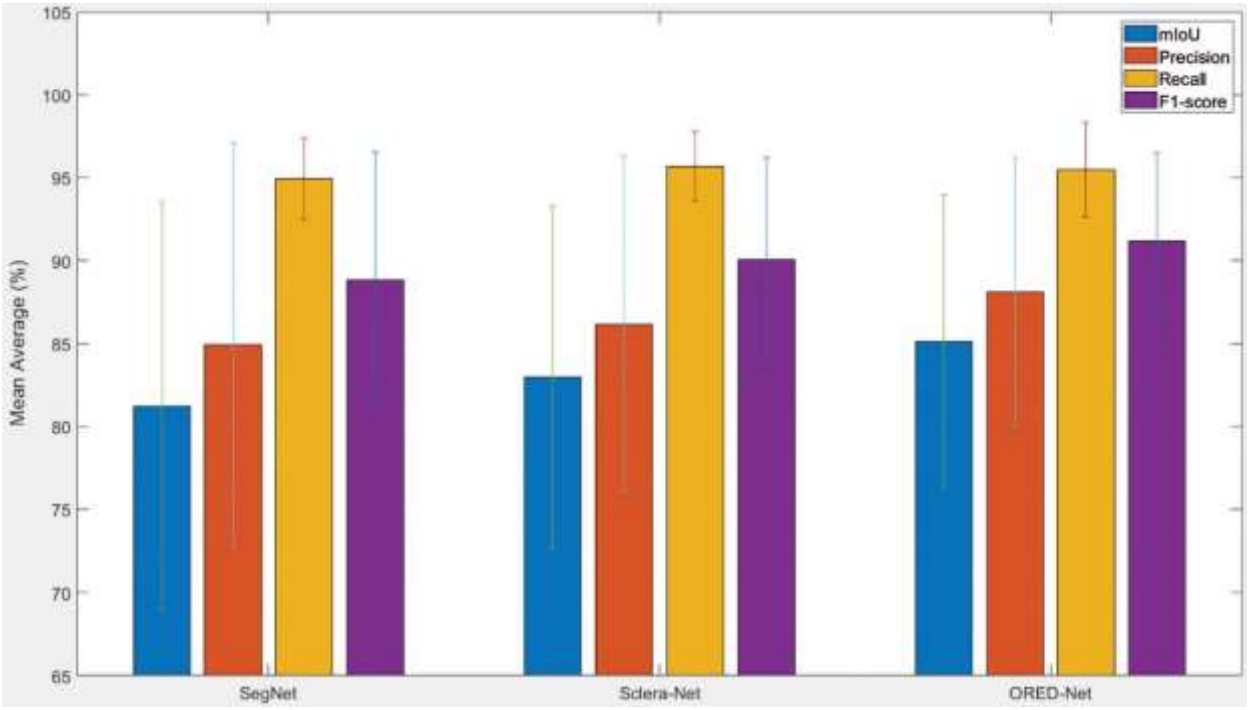


Figure 10: Mean intersection over union, precision, recall, and F1-score based on the UBIRIS.v2 database for the suggested approach and current options Five Takeaways

Conclusions:

This research proposed ORED-Net, a unique multi-class semantic segmentation network for segmenting the iris, sclera, pupil, and backdrop of the eye. The foundation of ORED-Net is the idea of outer residual connections, which allow spatial edge information to be transferred straight from the encoder's initial

layers to the decoder layers. This structure improves the when dealing with low-quality photos, this framework improves the network's performance. Because ORED-Net has fewer layers, it has fewer parameters and takes less time to compute.

The most noteworthy features of the suggested ORED-Network are that it converges in a significantly smaller number of epochs with direct flow of edge information, leading to faster training, and it achieves a high accuracy with a lighter network. Since ORED-Net eliminates the need for additional preprocessing overhead, the original image is used for both training and testing. With no preprocessing overhead, ORED-Net is the first network of its kind to segment the iris, sclera, and pupil—three crucial eye regions—all at once. The SBVPI and UBIRIS.v2 datasets, among other publicly accessible databases for eye area segmentation, were used to assess the suggested method's resilience and efficacy. This technique will be expanded into a reliable multimodal biometric identification system based on several eye regions in subsequent research. Statement of Funding: This work received funding..

References:

1. R. A. Naqvi andW. 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.
2. 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.
3. 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.
4. 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.
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, 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.
7. 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.
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. CMC, 2021, vol.66, no.1 731

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–
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.