

Received July 10, 2019, accepted July 18, 2019, date of publication July 23, 2019, date of current version August 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930593

Sclera-Net: Accurate Sclera Segmentation in Various Sensor Images Based on Residual Encoder and Decoder Network

RIZWAN ALI NAQVI^{1,2} AND WOONG-KEE LOH¹

¹Department of Software, Gachon University, Seongnam 13120, South Korea

²Faculty of Engineering and Technology, The Superior College, Lahore 55150, Pakistan

Corresponding author: Woong-Kee Loh (wkloh2@gachon.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korean Government (MSIT) under Grant NRF-2018R1A2B6009188.

ABSTRACT Sclera segmentation is revealed to be of noteworthy importance for ocular biometrics. The paramount step for biometric recognition methods is the segmentation of the area of interest, i.e., the sclera in our case. The sclera segmentation process plays a pivotal part in retaining the accuracy of the sclera-based recognition schemes by restraining the errors. However, accurate sclera segmentation in the images from various sensors in a real environment is quite challenging due to the saturated and/or defocused vessel patterns and the vessel structure, which has complex nonlinear deformations due to the multilayered sclera. With the development of deep learning algorithms, studies that are based on the sclera segmentation using convolutional neural networks (CNNs) have achieved promising results for sclera recognition. However, previous CNN-based methods are based on the repeated subsampling stages of convolution strides, or spatial pooling leads to losing much of the finer image structure that significantly decreases overall performance in tasks, such as semantic segmentation. In this paper, we present *Sclera-Net*, a residual encoder and decoder network that exploits identity and non-identity mapping residual skip connections to take benefit of the high-frequency information from the prior layers of both encoder and decoder networks to determine the accurate sclera region as well as other ocular regions. In this way, the finer image structure that was being lost due to repeated subsampling during convolution and pooling can be reutilized using residual skip connections to enhance overall performance. Furthermore, the proposed Sclera-Net does not enhance the performance on the cost of increasing depth, complexity, or the number of parameters. We performed comprehensive experiments and obtained optimum performance not only on sclera datasets but also on the iris datasets. In particular, we achieved an equal error rate and mean F1-score of 0.0093 and 96.2421, respectively, on the challenging SBVPI database, which is the best-reported result to date.

INDEX TERMS Sclera recognition, sclera segmentation, convolutional neural network, semantic segmentation, residual connections, encoder-decoder network, deep learning.

I. INTRODUCTION

In recent years, biometric authentication is being used and incorporated in our daily lives. Different characteristics can be used to verify and/or identify a person, such as biological, behavioral and physical traits. Biometrics is particularly useful because the information cannot be forgotten, changed, stolen, or lost. Hence, it provides an undoubted and reliable connection between the user and the device or application that uses it [1]. Owing to the efforts of researchers, biometric

technology is being utilized in various applications such as person identification in national databases or at the borders. It is playing a key role in behavioral and physiological forms, providing an effective platform for security issues.

Biometric recognition has shown an increased interest in new and unique human traits rather than the typical characteristics of the human body such as fingerprint, iris, face, voice, and retina, etc. [2]. Recognition systems for humans based on blood vessel patterns have been investigated using the palms [3], retina [4]–[6], conjunctival vasculature [7], and sclera [8]. No biometric method is perfect or can be useful uniquely and globally in real environmental conditions [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

To improve resilience to spoofing, provide large population coverage, ensure applicability to various environmental conditions, and attain high recognition accuracy, advance studies on different biometric traits are essential [8]. Among the several biometric methods and techniques, sclera recognition has the advantage of secure biometrics, i.e., the sclera regions are difficult to spoof because they are highly protected portions of the eye. Each individual has a unique structure of the blood vessels of the sclera, and it can be acquired non-intrusively in visible light. An individual's identification can be performed by using their vessel patterns on the sclera because these patterns are unique even for twins [10], these patterns have a high amount of randomness, and the left eye vessel patterns are different from the right eye of the same individual; thus, making them perfect for personal identification. In addition, the patterns remain unchanged throughout an individual's lifetime [11]. Humans among mammals have the uniqueness of extensive exposed sclera, making it feasible for the imaging of the surrounding conjunctival vasculature [12]. This is another important benefit of utilizing the sclera for human biometric authentication. Additionally, the sclera features can be easily fused with iris biometrics. Iris recognition is considered among the most accurate and reliable approaches for personal recognition. The iris images collected in near-infrared reveal complex and rich patterns. However, if the images are acquired in visible light, the iris recognition accuracy is adversely affected. Hence, the fusion of the sclera and iris features makes them more robust for biometrics [13].

Typical sclera recognition systems depend on sclera segmentation, enhancement of the blood vessel, feature extraction, and matching processes. Since sclera segmentation is the initial and basic step in sclera recognition systems, an incorrect segmentation or error will flow through the complete system and affect the overall accuracy. Moreover, incorrect sclera segmentation can reduce the region of the detected blood vessels or introduce new patterns, such as eyelids or eyelashes, which impair the effectiveness of the system. Various-sensor environments [14] such as visible light or near infrared can be an additional challenge for segmentation of ocular regions. Furthermore, a variety of illumination conditions can alter the view of the texture patterns by highlighting and attenuating numerous grey tones. Additionally, a verification system should not consume large computational resources to achieve real-time performance in the representation, extraction, and comparison of the images of texture.

The development of intelligent and expert systems is very helpful for humans in various fields such as recognition, detection, classification, and other challenging application. In automated systems, human level intelligence is imitated by artificial intelligent systems, and for these kinds of applications, deep learning is very famous. Although the recent developments in deep learning approaches shown good results in recognition tasks [15], [16], there exists noticeable limitations as well as room for improvements, when

it comes to tasks like semantic segmentation, in our case sclera segmentation. In order to overcome the challenges related to sclera segmentation and to encourage the creation of new intelligent sclera segmentation system, some competitions were performed [17]–[20]. In these competitions state-of-the-art, results were obtained using deep learning-based methods. SegNet [21] and RefineNet [22] were the winners of the Sclera Segmentation and Eye Recognition Benchmarking Competition (SSERBC) 2017 and Sclera Segmentation Benchmarking Competition (SSBC 2018), respectively. They showed remarkable results for sclera segmentation on provided Multi-Angle Sclera Dataset (MASD) database. However, these deep learning-based methods have clear limitations of information lost due to continuous down-sampling of images and/or increasing the computational complexity, depth and cost in terms of parameters that are required to be addressed. SegNet is an encoder-decoder network that was inspired from VGG-16 [15] network having drawbacks of vanishing gradient and overfitting problems, thereby results in loss of finer image structure. RefineNet has the disadvantage of very deep and complex network that results increase in cost and trainable parameters.

In this study, we emphasize that information lost at multiple stages during continuous down-sampling of the image is important for segmentation tasks. Furthermore, we ensure the efficient reuse of the image features in a manner that it does not increase the computational complexity and cost in terms of parameters. For this purpose, we propose, a deep learning-based sclera network (Sclera-Net) to detect the true boundary of the sclera and sharply acquires the class pixels to ensure correct sclera segmentation. Sclera-Net is based on exploiting residual connections for better flow of information gradient using identity mapping (IM) and non-identity mapping (NIM) residual building blocks (RBBs) [16] from the prior layers in both encoder and decoder networks to determine the accurate sclera region. In Sclera-Net, important information that may be lost during multiple stages of convolution strides or spatial pooling is reutilized from the prior layers through RBBs in a feed-forward fashion. It is useful for strengthening the feature propagation in the subsequent layers.

Sclera-Net is novel in the following four ways:

- It is an end-to-end semantic segmentation network for sclera and other ocular regions.
- It uses residual connectivity with IM and NIM for both encoder and decoder.
- It is a standalone network because the pre-detection of pupil, glint, iris, eyelid, and eyelashes is not required.
- Our Sclera-Net trained models and algorithms are publicly available for fair comparisons [23].

The proposed Sclera-Net achieves new state-of-the-art performance on three open databases of sclera: sclera blood vessels, periocular and iris (SBVPI) [24], [25]; mobile iris challenge evaluation (MICHE-I) [26]; and UBIRIS.v2 datasets [27]. Moreover, we evaluated the performance of the proposed segmentation network for another

very important ocular region, i.e., the iris, and performed experiments with famous iris datasets: noisy iris challenge evaluation (NICE-II) [28] and Chinese Academy of Sciences (CASIA) v4.0 [29] in this study. Sclera-Net showed optimum performance on not only the different sclera datasets but also the iris datasets. In particular, we achieved mean F1-score of 96.24 and 93.49 for challenging SBVPI and MICHE-I (Galaxy S4) datasets, respectively.

The remainder of this paper is organized as follows. In Section 2, we discuss in detail the previous studies related to sclera segmentation. In Section 3, our proposed method and its working methodology are explained. The experimental setup and results are presented in Section 4. Result analysis and related discussions are mentioned in Section 5. Conclusion and discussion on some ideas for the future work are presented in Section 6.

II. LITERATURE SURVEY ON SCLERA SEGMENTATION

Previous studies on sclera segmentation can be broadly categorized into two main classes i.e. handcrafted local features and deep learning features.

Sclera segmentation methods based on handcrafted local features can be further classified into Manual methods, Pixel thresholding methods, Shape contour methods. Manual methods [7], [30] for sclera segmentation were expensive approach for real time applications due to high processing time and mandatory human supervision. Pixel thresholding based methods [31]–[36] are good for certain situations in which illumination changes are not severe. However, the performance of sclera segmentation based on pixel thresholding is unsatisfactory for challenging cases owing to distortion and noise present in the sclera images. Shape contour based methods [37]–[42] are useful in certain cases; however, several challenges remain for sclera segmentation, such as occluded and noisy images need to be discarded manually. The sclera boundary can affect the convergence of the sclera shape contours. Moreover, incorrect sclera segmentation can reduce the region of vessels on the sclera or introduce patterns of eyelids and eyelashes, thereby compromising the effectiveness of the system. However, Convolutional neural network (CNN) using deep learning has been flourishing very fast and has proved as an influential method in tasks related to image processing. CNN outperformed previous conventional methods in a wide range of applications such medical and satellite image analysis [43], [44]. Deep learning features-based methods [45]–[48] have shown the advantages that handcrafted local features-based methods cannot achieve. Thus, in this study, we have focused on the deep learning features-based methods.

The deep learning feature based methods for sclera segmentation are further subdivided into two categories: image patch-based deep learning and full image-based deep learning methods. First, we discuss the former approach. Radu *et al.* [45] proposed a two-stage multi-classifier system architecture trained on randomly collected 60 patches (100 × 100 pixels) each of the sclera and non-sclera regions

from the UBIRIS.v1 database. In the first stage, a normal classifier was used, while in the second stage, a neural network based classifier worked on the probability space produced by the first classifier.

Next, we discuss the sclera segmentation based on full image-based deep learning methods. Rot *et al.* [46] proposed a multiclass segmentation approach for the eye region segmentation based on the encoder-decoder-based neural network model called semantic segmentation network (SegNet) [21]. SegNet includes encoder-decoder pairs, which are used to create feature maps for pixel wise classification of input images with different resolutions. They evaluated the results based on the multi-angle sclera database (MASD) [19]. Lucio *et al.* [47] proposed two approaches for sclera segmentation, i.e., fully convolutional network (FCN) [46] and generative adversarial network (GAN) [50]. The employed FCN without the fully connected layers, i.e., VGG-16 without the last three layers was proposed by Teichmann *et al.* [48].

To promote the development of new segmentation methods for sclera, some competitions were also performed [17]–[20]. In [17], Das *et al.* proposed a benchmark for sclera segmentation where four teams presented their approaches for the defined task. The best results were obtained image patch-based method already explained in [45]. Later, Das *et al.* presented a new benchmark [18], which deals with sclera segmentation as well as recognition. The best segmentation results were achieved based on Fuzzy C Means that considers spatial information and Gaussian kernel function is used for calculating the distance between the data points and the center of the cluster. They achieved segmentation results of 85.21% and 80.21% for precision and recall, respectively. Das *et al.* [19] presented a new sclera segmentation and eye recognition benchmark where seven teams proposed their algorithms for the assigned task. The winner of that competition attained precision and recall rates of 95.34% and 96.65%, respectively. They acquired results using SegNet method [21] in their proposed approach. In a recent competition [20], benchmarking for sclera segmentation was presented by Das *et al.* All submitted algorithms were evaluated based on a cross-sensor scenario, i.e., images were collected using the DSLR and mobile phone cameras. Precision and recall rates of 81.35 and 75.82%, respectively, were attained [20]. These results were obtained using high-resolution semantic segmentation architecture based on the multipath refinement approach called RefineNet [22].

In Table 1, we have prepared a summary of the comparison of the proposed method with previous methods on sclera segmentation along with their strengths and weaknesses.

III. PROPOSED METHOD FOR SCLERA SEGMENTATION

A. OVERVIEW OF SCLERA SEGMENTATION USING SCLERA-NET

Sclera-Net is applied to the full input eye image that has not undergone any preprocessing overhead. Sclera-Net is a

TABLE 1. Comparison of sclera segmentation based on deep learning features.

Category	Methods	Strengths	Weaknesses
Using deep learning features	Image patch-based method [45]	<ul style="list-style-type: none"> - Accuracy is higher for true sclera boundary detection as compared to conventional methods - Two stage multiple classifier system (MCS) architecture is employed 	<ul style="list-style-type: none"> - Post-processing is required along with MCS - Processing time is higher as many iterations are required for accuracy
	Multiclass eye region segmentation based on SegNet [46]	<ul style="list-style-type: none"> - Multiclass eye segmentation is performed - Six eye regions are segmented using encoder-decoder based network, i.e., SegNet 	<ul style="list-style-type: none"> - Method is trained and tested only on one dataset, i.e., MASD dataset - Fine-tuning of previously proposed approach, i.e., SegNet, is employed - Large amount of artificially created data is used in training
	Sclera segmentation based on RefineNet [20,22]	<ul style="list-style-type: none"> - Encoder part is based on ResNet-101 blocks - RefineNet blocks fuse/concatenate high and low resolution features 	<ul style="list-style-type: none"> - Computationally expensive in terms of depth and parameters - Trained only for fixed learning rate i.e. stage 1 [20] - Trained and tested only on MASD dataset [20]
	Full image-based methods	FCN and GAN based sclera segmentation [47]	<ul style="list-style-type: none"> - Promising results for sclera segmentation using deep learning - Manually annotated segmentation mask for subset of the UBIRIS.v2 [27] and MICHE-I [26] databases are provided for fair comparison and objective evaluations
	Sclera-Net (Proposed method)	<ul style="list-style-type: none"> - Accurate detection of the sclera boundary without the extra cost for pre-processing - Better flow of information in network is achieved by IM and NIM RBBS - High accuracy without increasing parameters and depth of network 	<ul style="list-style-type: none"> - Excessive data and time is required for training network - Requires ground truth for training and testing of new models

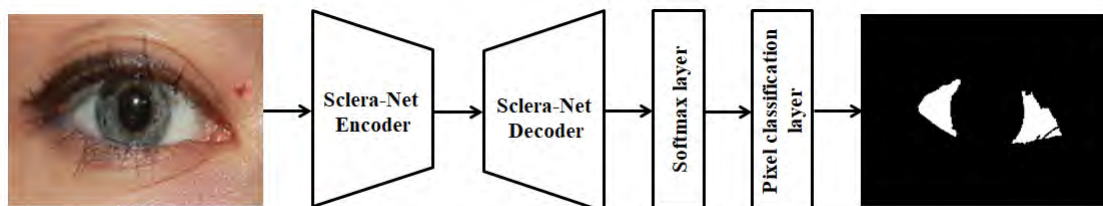


FIGURE 1. Overview of the proposed method for sclera segmentation.

convolutional encoder and decoder network based on the residual connections in the encoder and decoder. The distinct Sclera-Net encoder and decoder are shown in Figure 1 to elaborate the overall process. The encoder compresses the significant or semantic contents of the input image, which can be represented as tiny features. The encoder outputs the descriptive representation of the image, which is then given as input to the decoder. The final segmented output is regenerated by the decoder into the original dimensions of the input image. In the Sclera-Net decoder, the process of upsampling is performed by max-pooling indices and estimation of sclera and non-sclera classes is performed using the softmax loss function in the softmax layer. The pixel classification layer is followed by the softmax layer, which is responsible to predict pixel labels in the given input eye image. Cross-entropy loss function is used in the pixel classification layer.

Sclera-Net is based on SegNet architecture and takes the residual connections concept from the ResNet model, which ensures the accuracy and reliability of CNNs. The SegNet architecture is comprised of two main blocks: encoder and decoder [21]. The encoder compresses significant or semantic contents of the input image and outputs the descriptive representation of the image, which is then given as input to the decoder. The final segmented output is generated by the decoder. The SegNet encoder consists of 13 convolutional layers and 5 pooling layers similar to the VGG-16 network [15], except for the fully connected layers. Hence, SegNet is inspired by the VGG-16 network. On the other hand, the decoder network is the inverted VGG-16 network, also without the fully connected layers. It carries the softmax layer, which generates probability distributions per pixel, which is further classified into target classes. In the training process, the encoder acquires low-resolution semantic feature

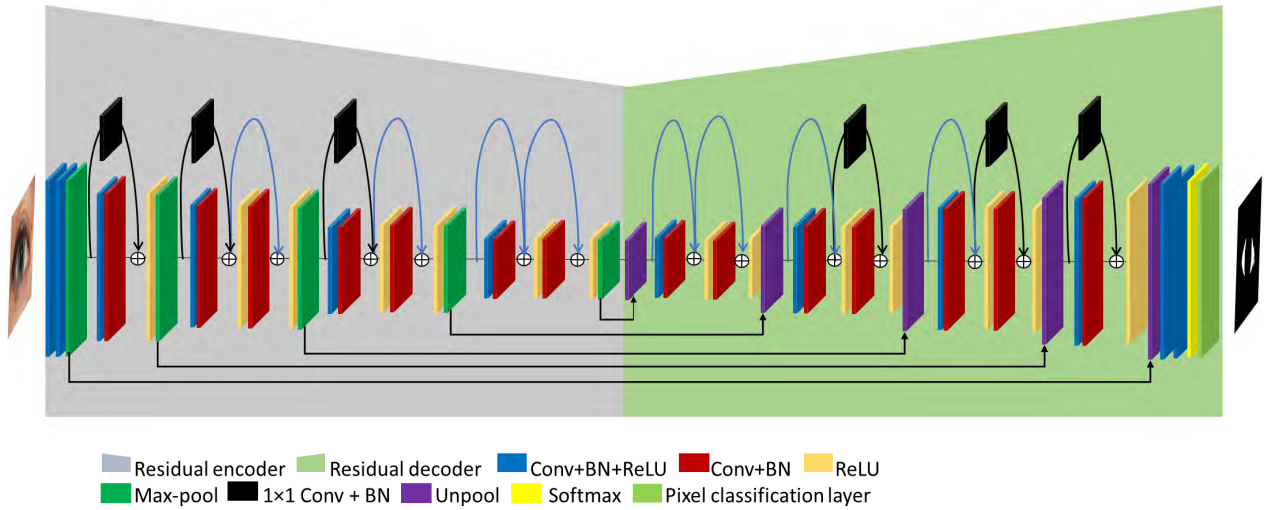


FIGURE 2. Deep learning based sclera segmentation system with residual encoder and decoder networks.

maps, while the decoder learns filters that can create high-resolution masks for segmentation based on the feature maps generated by the encoder.

Furthermore, Sclera-Net inherits the concept of residual connections from ResNet, which has higher accuracy as compared to VGG networks [15], [16]. Additionally, the most important factor for choosing ResNet residual concept is that typical classification networks dramatically downsample the image size to represent the image in the form of features. During this process, high-frequency contextual information is crushed and degraded, known as the vanishing gradient problem [51]. This problem was solved by ResNet through mapping with shortcut connections such as residual connections.

As shown in Figure 2, Sclera-Net is the combination of a residual encoder and decoder based on residual skip connections. The residual skip connections bring a major benefit as they reinforce features by using the high-frequency components from prior layers that are continuously lost as a result of the convolution and up-sampling operations of the encoder and decoder, respectively. The residual skip connections are based on the residual building blocks (RBB) [51], which are categorized into IM and NIM residual connections.

The identity mapping RBB is used if the number of channels in the input features and residual function are the same. It can be explained through Equation (1) and shown in Figure 3 (a).

$$F_{i+1} = F_i + R(F_i, W_i), \tag{1}$$

where F_i represents the input features, F_{i+1} the output features, i is the sequence number of a residual block, W_i is the weight of the corresponding residual block, and R represents the residual function. F_i is the identity feature provided for the element-wise addition to generate the output feature F_{i+1} after the residual operation [51].

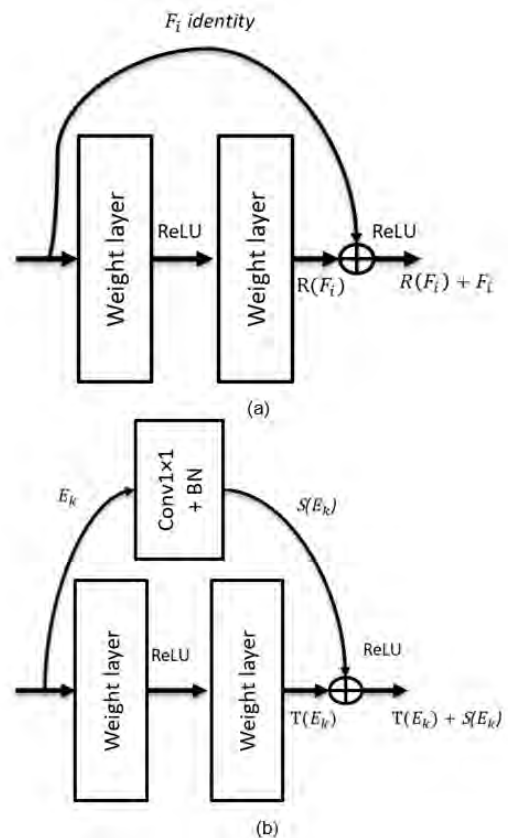


FIGURE 3. Residual block (a) identity mapping (b) non-identity mapping.

Similarly, the non-identity mapping RBB is used when the number of channels in the input feature F_i does not match the number of channels after the residual function R , as we perform the NIM of 1×1 convolutions for the element-wise addition [51]. This is represented in Equation (2) and shown in Figure 3 (b).

$$X_{k+1} = T(E_k) + S(E_k, W_k), \tag{2}$$

where E_k represents the input features, X_{k+1} the output features, k the sequence number of a residual block, W_k the weight of the corresponding residual block, S the residual function, and convolution (Conv 1×1) combined with BN is represented by $T(E_k)$ [52].

In our proposed Sclera-Net, we used both IM and NIM as shown in Figure 2. Hence, spatial information that was lost by the continuous convolution process is retained through the addition of an RBB.

B. SCLERA-NET ENCODER AND DECODER

As explained in previous section, RBBs are more useful than typical series networks because they provide high accuracy by enhancing features.

In the Sclera-Net encoder, 5 Groups comprising of 13 convolution layers of 3×3 filters are used. The first two Groups contain two convolution layers each, while the remaining three Groups contain three convolution layers each. There are seven residual connections in the Sclera-Net encoder, out of which three are NIM and four are IM residual connections. The first Group does not have any residual connections because high-frequency loss is not very significant in the initial Groups. The type of residual connection is based on the transition size between the layers and requirement of matching the channel size of feature map for element-wise addition. If the transition size is same, then IM residual connections are used; however, different size transitions are compensated by NIM residual connections. Hence, Group 2 to Group 4 have three NIM residual connections and Group 3 to Group 5 have four IM residual connections.

In Figure 2, the convolution, batch normalization, rectified linear unit, max-pooling, and unpooling (upsampling) layers are represented as Conv, BN, ReLU, Max-pool, and Unpool, respectively. The NIM residual connection is based on the convolution layer of size 1×1 and BN layer. This size of convolution layer is selected to match channel size for element-wise addition. After summation in each block, the ReLU layers are used for post activation. The pooling indices information of the pooling layers is provided to the decoder part to preserve the indices and input image size. The unpooling layer regenerates the input image in the decoder part based on the indices and size related information.

In Table 10 of the appendix section, we have listed the layer Groups and provided the details of each layer used in the Sclera-Net encoder. The size of the input image depends on the kind of database used for the experiment. Here, we have selected input images of size 224×224 for illustration purposes.

Next, we explain our proposed Sclera-Net structure for the encoder shown in Figure 2 and elaborated in Table 10 of the appendix section. It comprises of 13 convolutional layers and 5 pooling layers. In the first convolutional layer, 64 filters are used with a feature map size of $224 \times 224 \times 64$ and a kernel size (k) of 3×3 . Here, in the feature map, 224×224 is the width (w) and height (h) of the output. They are calculated on

the basis of following equation:

$$output_{h \text{ or } w} = \frac{(input_{h \text{ or } w} - k_{h \text{ or } w} + 2 \times p)}{s} + 1, \quad (3)$$

For example, in Table 10, the input width ($input_w$), kernel width (k_w), padding (p), and stride (s) are 224, 3, 1, and 1, respectively. Therefore, the output width value can be achieved by substituting these values in the above-mentioned equation, i.e., $224 = ((224 - 3 + 1 \times 2)/1 + 1)$.

Following each convolutional layer, batch normalization is performed based on the mean and standard deviation of the data. This reduces the problem of internal covariate shift. Notably, normalization has to be performed independently for each dimension, over ‘mini-batches’, and not in one thread for all dimensions, hence the name ‘batch’ normalization. A rectified linear unit (ReLU) layer is also applied as an activation function following each batch normalization. The function for the ReLU layer is explained in [53], [54]. ReLU is explained through the following simple equation:

$$v = \max(0, u), \quad (4)$$

where u represents the input values and v represents the output values. This function reduces the problem of vanishing gradient [55]. This problem may occur when sigmoid and hyperbolic tangent functions are used in back-propagation for training. ReLU has a faster processing speed than a non-linear activation function. Initially, the feature map is passed through the ReLU layer. Subsequently, the obtained feature map, which is passed through the second convolutional layer, is once again passed through the ReLU layer before it passes through the max pooling layer, as shown in Table 10. Here, the second convolutional layer maintains the feature size of the first convolutional layer, i.e., $224 \times 224 \times 64$; whereas the size of filters, paddings, and strides are 3, 1, and 1, respectively. For maintaining clarity and simplicity in Table 10, BN layer is included in Conv layer.

The Sclera-Net decoder architecture is the mirror image of the encoder as shown in the Table 11 of the appendix section. Features are upsampled by the Sclera-Net decoder by utilizing the pooling indices obtained from the Sclera-Net encoder. For extracting same size images from the encoded features, the images in the decoder are processed through the same number of convolution layers. The features in the decoder are first unpooled, and subsequently the image undergoes the convolution operations. It is in contrast to the encoder, in which the pooling operation is performed after the convolution process. There are two filters in the last convolution layer, which indicate the channel number of the output or the number of classes. In the Sclera-Net decoder part, seven RBBs are used in the inverse to the encoder for upsampling the image. Here, we deal with two classes: sclera and non-sclera, and two outputs or masks can be obtained: sclera and non-sclera pixels. At the end, Sclera-Net includes the classification layer, which classifies each pixel as sclera or non-sclera based on the softmax loss-function.

TABLE 2. Key differences of proposed (Sclera-Net) architecture from SegNet [21] architecture.

SegNet [21]	Proposed (Sclera-Net)
- SegNet is inspired from the VGG16 network [15] that downsample the image rapidly because of the limitations of vanishing gradient and overfitting[16]	- Sclera-Net is a fully residual convolutional network with residual connections in both the encoder and decoder to prevent the network from the vanishing gradient and overfitting problem
- High-frequency contextual information is lost	- High-frequency information is reutilized through residual skip connections
- Feature reuse concept is not utilized in SegNet [21]	- Efficiency is enhanced by virtue of the reuse of the features from the previous layer
- Accuracy is adversely affected by increase in depth and the vanishing gradient issue	- In residual-based networks, the accuracy increases together with the depth by virtue of due to feature empowerment [16]

Residual decoder includes the same number of connections and size of feature maps as that in the residual encoder. The decoder presents an image of the same size as used at the input of the encoder, i.e., 224×224 . The size of image varies based on the kind of database used for segmentation. In the final output layer, a mask is obtained, which comprises of different values of pixels based on the classes, i.e., sclera or non-sclera in this case. Details on each layer while example images of the best, medium, and worst cases traveled through Sclera-Net encoder and decoder is shown in Figures 19, 20, and 21 of the appendix section, respectively.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL DATA AND SETUP

In this study, we used sclera blood vessels, periocular and iris (SBVPI) database, collected for research related to sclera and periocular recognition [24]. This dataset includes 2399 high quality images collected from 55 different individuals. For each individual, 32 images were collected while he/she looked in four different directions: straight, left, right, and up. Most of the previous databases deal with the segmentation of only the iris or pupil in different scenarios. However, these databases cannot be used for the sclera and other ocular region segmentations. MASD version 1 is used for sclera segmentation and eye recognition in the visible spectrum [19]. However, we did not use this database because the information of ground truth was not provided. Without this information, we could not train or test our method. Therefore, we used the SBVPI database [25].

In our experiment, we performed two-fold cross validation for training and testing our proposed model. For this, we randomly divided the database into two subsets. From the images of 55 people, the images of 30 people were used for training by applying augmentation of images described in part B of section 4, and the images of the remaining 25 people were used for the testing purpose without applying augmentation. Data augmentation of the training data is performed to avoid overfitting issues. For the training and testing of Sclera-Net, we used a desktop computer with an Intel®Core™ (Santa Clara, CA, USA) i7-7700 CPU @3.60 GHz, 32 GB memory, and an NVIDIA GeForce GTX 1080 Ti (3584 CUDA cores and 11 GB memory) graphics card. MATLAB 2018b [56] was used for performing the experiments. We performed the training of Sclera-Net by using experimental databases.

Therefore, no fine-tuned or pre-trained networks such as ResNet, DenseNet, Inception Net, or GoogleNet were used.

B. DATA AUGMENTATION

Augmentation of training data is performed to increase the number of training samples to achieve better performance. Specifically for segmentation tasks, accuracy of the task depends on the quantity of training images and their corresponding annotated (ground truth) images. Hence, data augmentation is an artificial way to increase the quantity of training images. Models perform well with more data obtained through data augmentation [57]. In our case, we artificially produced 14 images from each of the 900 training images. Hence, 12,600 images were produced using different augmentation techniques such as image translation (left, right, up, and down), flipping (horizontal), cropping, resizing, etc. Other machine learning and deep learning tasks also used these types of techniques to achieve better accuracy [58].

C. TRAINING OF SCLERA-NET

In this research, the training of Sclera-Net was performed without using pre-trained networks for the segmentation of the ocular eye region, i.e., the sclera, using our designed Sclera-Net model. For this, original images were used without any enhancement or preprocessing, and a stochastic gradient descent (SGD) optimization method was used [59]. This optimization method minimizes the difference between the expected and actual outputs. In SGD, iteration is defined as the number of training samples divided by the mini-batch size, and the number of epochs is set to one. In our experiment, we performed the training for a predefined number of epochs, i.e., 50, and a mini-batch size of 20. In other words, with 50 epochs, the model was exposed to the entire dataset 50 times in the training process. Multiple epochs allow a learning algorithm to run until it converges or its error is sufficiently minimized; hence, we selected 50 epochs. However, the batch size can vary based on the size of the database. As shown in Equations (5) and (6), one epoch is counted when training is performed once with the entire dataset.

$$x_{j+1} := p \cdot x_j - d \cdot \eta \cdot w_j - \eta \cdot \left(\frac{\partial R_j(w)}{\partial w} \Big|_{w_j} \right) D_j, \quad (5)$$

$$w_{j+1} := w_j + x_{j+1}, \quad (6)$$

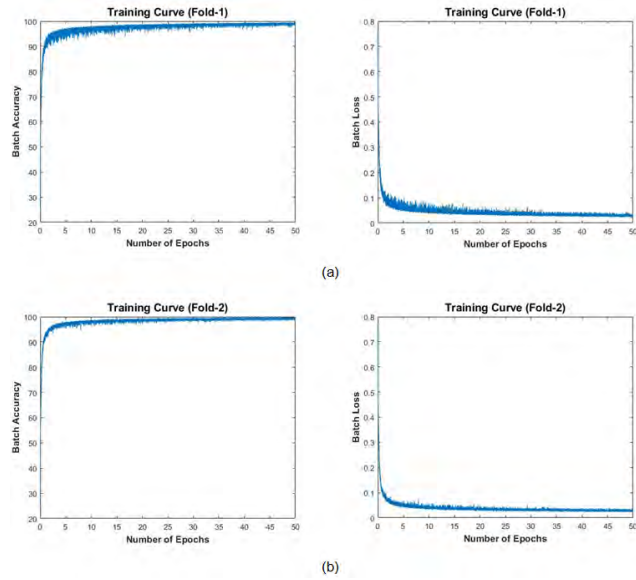


FIGURE 4. The curves for the training accuracy (figures on the left) and training loss (figures on the right) based on the number of epochs for the 2-fold cross validation; (a) First fold (b) Second fold.

Here, w_j is the learnt weight at the j^{th} iteration, x_j is the momentum variable, p is the momentum, d is the decay of weight, and η is the learning rate. $\left\langle \frac{\partial R_j(w)}{\partial w} \right\rangle_{w_j} D_j$ is the average over the j^{th} batch, D_j of the derivative of the object with respect to w evaluated at w_j . In view of the optimal parameters of training using SGD, p , d , and η of Equations 4 and 5 were set as 0.9, 0.0005, and 0.01, respectively.

Training loss is calculated based on all the image pixels present in the mini-batch using cross-entropy loss function [21]. The relationship between training loss and training accuracy for two-fold cross validation, i.e., training accuracy and loss curves for the first and second folds are shown in Figures 4 (a) and (b), respectively. The figures on the left represent the training accuracy curves, while the figures on the right represent the training loss curves. The x-axis represents the number of epochs, while the y-axis represents the training accuracy or training loss of each batch for 50 epochs. The increase or decrease in the loss factor depends on the batch size (20) and learning rate (0.01). Loss decreases gradually when the learning rate is lowered, thereby showing the linearity. However, if the learning rate is high, the loss value decreases sharply. An optimal CNN model cannot be obtained if the learning rate is high because this may lead to the problem of a high loss value, thereby, resulting in a poorly trained model. In this experiment, we used optimal models with loss curves close to 0 (0 %) and training accuracies close to 1 (100 %) as shown in Figure 4.

The ocular regions, such as sclera, pupil, iris, etc., in the eye images databases are usually very small compared to the other regions; hence, the quantity of non-sclera, non-pupil, or non-iris pixels are much larger than that of the sclera, pupil, or iris pixels. Therefore, there can be a big difference of frequencies

in each class while training over a dataset. The frequency difference between each class indicates that the non-ocular regions dominate during training; hence, there should be a balance between each class. For avoiding the under representation of ocular classes during training, frequency balancing is applied. Here, we use median frequency balancing, where weight is allocated to the cross-entropy loss [60]. Weights are determined from the training data sample using the following equations:

$$WC_1 = \frac{Med_{frequency}}{freq_{c1}}, \quad (7)$$

$$WC_2 = \frac{Med_{frequency}}{freq_{c2}}, \quad (8)$$

where WC_1 and WC_2 are the weights of the sclera and non-sclera classes, respectively, $freq_{c1}$ represents the number of sclera pixels in the image, $freq_{c2}$ represents the number of non-sclera pixels in the image, and $Med_{frequency}$ is the median of $freq_{c1}$ and $freq_{c2}$. Based on frequency balancing, a weight value of less than 1 is assigned to a major class, whereas a minor class is assigned a value greater than 1.

D. TESTING OF SCLERA-NET FOR SCLERA SEGMENTATION

To acquire the results of segmentation from the proposed Sclera-Net, an image is given as an input to the trained model, and there are no extra steps, such as preprocessing, involved during the training and testing of the Sclera-Net model. The given input image passes through the proposed Sclera-Net encoder and decoder, and the output is a binary segmentation mask. This segmentation mask is further used by the trained model to evaluate and generate the sclera region segmentation results. The segmentation performance of the proposed Sclera-Net is evaluated using different metrics such as average segmentation error ($Error_{avg}$), intersection over union (IoU), and precision, recall, and F1-score (PRF). First, we discuss the average segmentation error, which is being used by many researchers to evaluate the segmentation performance. Pixel classification accuracy is calculated by the exclusive-OR (XOR) logic between the output image ($O_j(m, n)$) from the trained model and its ground truth mask ($M_j(m, n)$), which is given as

$$P_j = \frac{1}{m \times n} \sum_{m,n} O_j(m, n) \oplus M_j(m, n), \quad (9)$$

where m and n are the width and height, respectively, of the image, and P_j is the pixel classification accuracy of each image. The overall error in segmentation is represented as $Error_{avg}$ and calculated by averaging the classification error (P_j) over all images in the database as shown in Equation 10.

$$Error_{avg} = \frac{1}{T} \sum_j P_j, \quad (10)$$

where T represents the total number of tested images. The value of $Error_{avg}$ always lies between 0 and 1. The error is minimum if the value of $Error_{avg}$ is close to 0, whereas the error will be the largest if value of $Error_{avg}$ is close to 1.

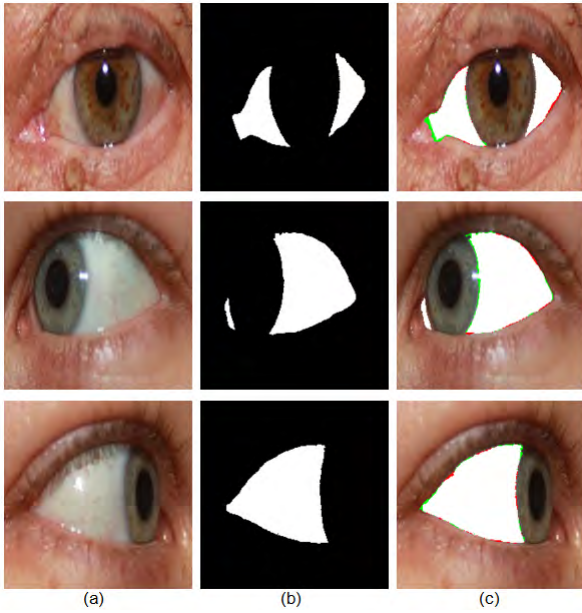


FIGURE 5. Sample images of the correct results of segmentation using Sclera-Net on the SBVPI database [25]. (a) Original image, (b) annotated image, and (c) segmentation result, where the white regions represent the true positive pixels, whereas the green and red regions represent the false positive and false negative pixels, respectively.

TABLE 3. Comparison of Sclera-Net and previous methods based on the SBVPI database [25].

Method	Folds/Average	Average Segmentation Error <i>Error_{avg}</i>
SegNet [21]	Fold 1	0.01699
	Fold 2	0.02182
	Average	0.01941
RefineNet [22]	Fold 1	0.01481
	Fold 2	0.01686
	Average	0.01584
Sclera-Net (Proposed Method)	Fold 1	0.00751
	Fold 2	0.01104
	Average	0.00928

1) RESULTS OBTAINED FOR SCLERA SEGMENTATION FROM THE PROPOSED SCLERA-NET METHOD

Figures 5 and 6 show the correct and incorrect results of sclera segmentation using Sclera-Net on the SBVPI database [25]. For pictorial representation of the results, two types of errors, false positive and false negative are defined that are represented in green and red color, respectively. A false positive error occurs when a non-sclera pixel is misclassified as a sclera pixel by the network, whereas a false negative error occurs when a sclera pixel is misclassified as a non-sclera pixel by the network. Whereas, the sclera pixels that are correctly classified as sclera pixels are known as true positive that are represented in white.

2) COMPARISON OF THE PROPOSED SCLERA-NET METHOD WITH PREVIOUS METHODS

In the next experiment, we compared our proposed Sclera-Net model with the previous state-of-the-art models,

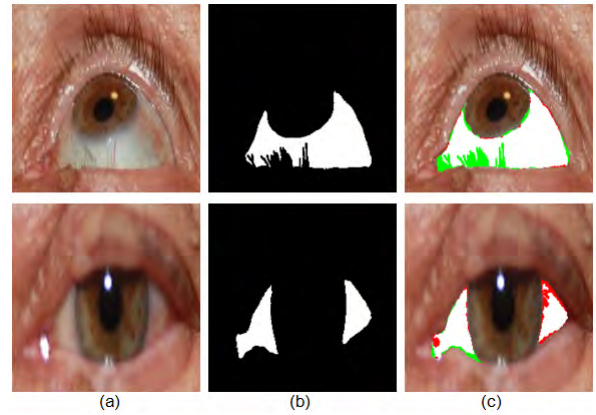


FIGURE 6. Sample images of the incorrect results of segmentation using Sclera-Net on the SBVPI database [25]. (a) Original image, (b) annotated image, and (c) segmentation result, where the white regions represent the true positive pixels, whereas the green and red regions represent the false positive and false negative pixels, respectively.

TABLE 4. Comparison of the Sclera-Net and previous methods using *IoU* based on the SBVPI database [25].

Method	Folds/Average	Intersection over Union <i>IoU</i>
SegNet [21]	Fold 1	0.86375
	Fold 2	0.85963
	Average	0.86169
RefineNet [22]	Fold 1	0.89206
	Fold 2	0.87776
	Average	0.88491
Sclera-Net (Proposed Method)	Fold 1	0.92362
	Fold 2	0.91237
	Average	0.91800

i.e., SegNet [21] and RefineNet [22]. The overall error in segmentation, i.e., *Error_{avg}* is the metrics used to evaluate the performance of Sclera-Net with previous methods. For comparison purposes, we have extracted the results on the SBVPI dataset [25]. We performed a 2-fold cross validation for the training and testing of our proposed model on the 2399 images of the SBVPI database collected from 55 individuals. For this purpose, we randomly divided the database into two subsets i.e., the 1st subset and the 2nd subset. In fold 1, the augmented images of the 1st subset were used for training and the images of 2nd subset were used for testing with no augmentation applied. Similarly, in fold 2, the augmented images of the 2nd subset were used for training and the images of 1st subset with no augmentation were used for testing. Table 3 shows the average segmentation error *Error_{avg}* for two-folds cross validation and their average.

To quantify the overlap in percentage terms between the ground truth image and the predicted output, some of the previous studies have used the *IoU* metric [61], [62], also referred to as the Jaccard index. The *IoU* measures the pixels shared between the predicted output and the ground truth image divided by the total number of pixels present across both. The *IoU* can be determined using the

TABLE 5. Comparison of the Sclera-Net and previous methods based on the SBVPI database [25].

Method	Folds/Average	Precision		Recall		F1-score	
		Mean (μ)	SD (σ)	Mean (μ)	SD (σ)	Mean (μ)	SD (σ)
SegNet [21]	Fold 1	89.59	6.43	96.20	5.16	92.70	3.28
	Fold 2	87.94	5.20	97.53	2.73	92.36	3.31
	Average	88.76	5.81	96.87	3.94	92.53	3.30
RefineNet [22]	Fold 1	93.66	6.07	97.42	5.67	93.37	4.06
	Fold 2	90.56	4.97	96.67	4.75	93.42	3.83
	Average	92.11	5.52	97.05	5.21	93.40	3.95
Sclera-Net (Proposed Method)	Fold 1	95.74	4.13	97.17	1.73	96.48	1.70
	Fold 2	93.06	2.43	99.16	1.46	96.00	1.72
	Average	94.40	3.28	98.17	1.60	96.24	1.71

TABLE 6. Comparison of Sclera-Net and previous methods based on the MICHE-I database [26].

Method	Subset of MICHE-I	Precision		Recall		F1-score	
		Mean (μ)	SD (σ)	Mean (μ)	SD (σ)	Mean (μ)	SD (σ)
SegNet* [21]	Galaxy S4	76.09	23.80	66.50	26.34	67.92	23.87
	Galaxy Tab 2	76.46	18.29	73.77	21.20	72.33	18.26
	Iphone 5	79.40	32.93	31.90	26.05	40.95	29.19
FCN* [49]	Galaxy S4	88.65	10.62	88.24	12.03	88.12	10.56
	Galaxy Tab 2	88.50	12.68	87.86	12.23	87.94	11.59
	Iphone 5	89.32	5.22	87.51	11.61	87.80	8.24
Sclera-Net (Proposed Method)	Galaxy S4	92.24	5.20	95.05	3.52	93.49	3.39
	Galaxy Tab 2	94.57	4.00	92.29	5.61	93.26	3.49
	Iphone 5	88.83	3.50	96.73	5.14	92.66	4.90

*The precision, recall and F1-score results for [21] and [49] are referred from [47].

TABLE 7. Comparison of Sclera-Net and previous methods using the UBIRIS.v2 database [27] based on the PRF-protocol. Greater mean and smaller standard deviation values indicate better performance.

Method	Precision		Recall		F1-score	
	Mean (μ)	SD (σ)	Mean (μ)	SD (σ)	Mean (μ)	SD (σ)
SegNet* [21]	87.52	8.53	72.48	17.15	77.82	13.08
FCN* [49]	88.45	6.98	87.31	6.68	87.48	3.90
Sclera-Net	92.55	4.87	89.18	6.85	90.56	3.62

*The precision, recall and F1-score results for [21] and [49] are referred from [47].

TABLE 8. Comparison of Sclera-Net and previous methods for iris segmentation using the NICE-II database.

Method	Average Segmentation Error $Error_{avg}$
Haindl et al. [65]	0.0124
Zhao et al. [66]	0.0121
Arsalan et al. [67]	0.0082
SegNet* [21]	0.0078
Sclera-Net (Proposed Method)	0.0063

*The average segmentation error for [21] is referred from [68].

following equation:

$$IoU = \frac{TP}{TP + FP + FN}, \quad (11)$$

As we examine the equation, we can see that the IoU is the ratio of the true positive value divided by the sum of true positive, false positive and false negative values. Table 4 shows the results of the IoU for two-folds cross validation and their average.

TABLE 9. Comparison of Sclera-Net and previous methods for iris segmentation using the CASIA v4.0 distance database.

Method	Average Segmentation Error $Error_{avg}$
Tan et al. [69]	0.0081
Zhao et al. [66]	0.0068
Liu et al. [64]	0.0059
SegNet* [21]	0.0044
Sclera-Net (Proposed Method)	0.0033

*The average segmentation error for [21] is referred from [68].

For sclera segmentation, some of the previous researches used other metrics based on PRF , i.e., precision or positive predictive value (PPV), recall or true positive rate (TPR), and $F1$ -score. For evaluating sclera segmentation, this metrics is useful to compare the results obtained using different methods [63]. Strength and weakness of the method can be measured using Equations (12) to (14), where TP , TN , FP , and FN are the numbers of true positive, true negative, false positive, and false negative, respectively.

$$PPV = \frac{TP}{TP + FP}, \quad (12)$$

$$TPR = \frac{TP}{TP + FN}, \quad (13)$$

$$F1\text{-score} = \frac{2 \times PPV \times TPR}{PPV + TPR}, \quad (14)$$

Table 5 and Figure 16 shows the PRF results calculated based on above defined equations. The lowest and highest values of PPV , TPR , and $F1$ -score were 0 and 100 %, respec-

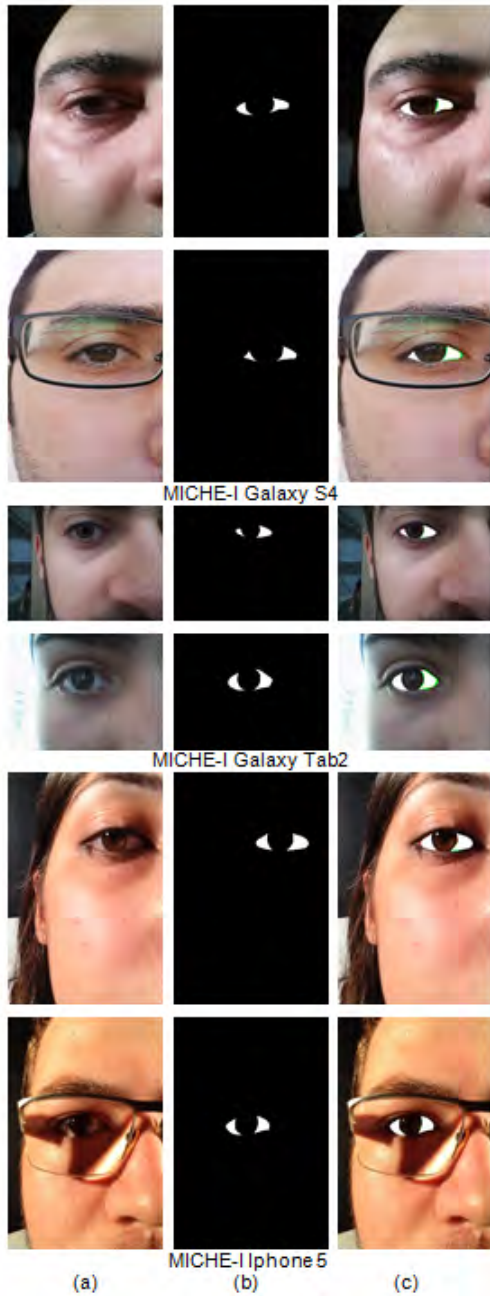


FIGURE 7. Sample images of the correct segmentation results using Sclera-Net on the MICHE-I (MICHE-I Galaxy S4, MICHE-I Galaxy Tab2, MICHE-I Iphone 5) subdatabases [26] while viewing in different directions. (a) Original image, (b) annotated image, and (c) segmentation result, where the white regions represent the true positive pixels, whereas the green and red regions represent the false positive and false negative pixels, respectively.

tively. Results show that our proposed method outperformed other methods in all the defined metrics.

3) SCLERA SEGMENTATION ACCURACIES BY SCLERA-NET BASED ON OPEN DATABASES

For evaluating Sclera-Net in various sensors environment, this research includes experiments with additional open databases: MICHE-I [26] and UBIRIS.v2 [27]. Sclera-Net is

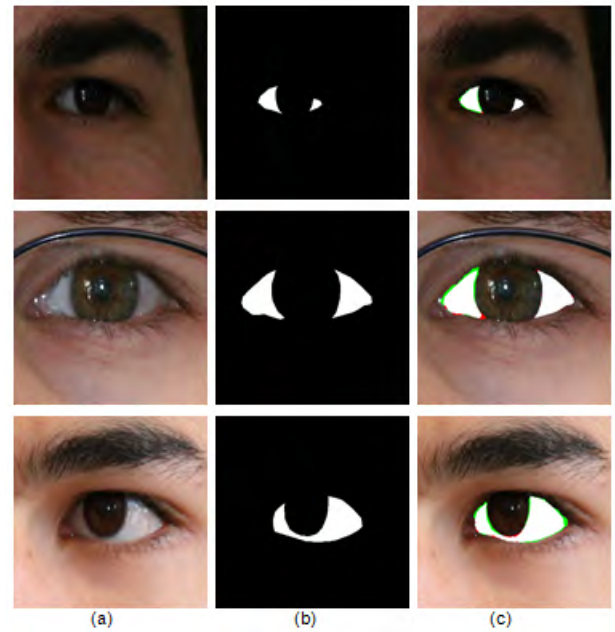


FIGURE 8. Sample images of the correct sclera segmentation results using Sclera-Net on the UBIRIS.v2 database [27]. (a) Original image, (b) annotated image, and (c) segmentation result, where the white regions represent the true positive pixels, whereas the green and red regions represent the false positive and false negative pixels, respectively.

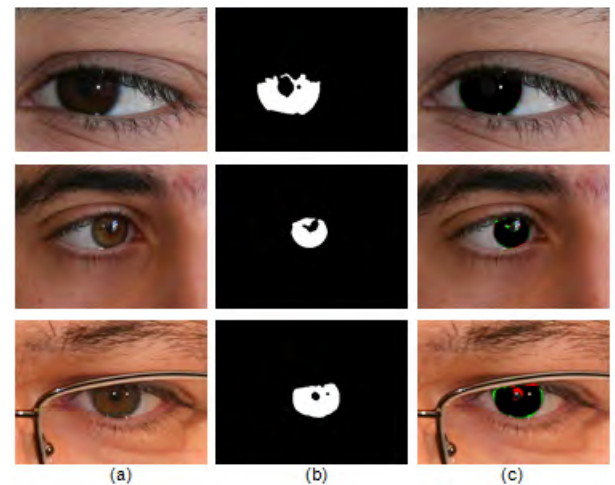


FIGURE 9. Sample images of the iris segmentation results using the proposed Sclera-Net on the NICE-II database. (a) Original image, (b) annotated image, and (c) segmentation result, where the black regions represent the true positive pixels, whereas the green and red regions represent the false positive and false negative pixels, respectively.

separately trained without fine-tuning for the open databases. Additional details related to hardware and capturing strategies used in the databases can be found in their respective references. The ground truth images for MICHE-I and UBIRIS.v2 were obtained from [47]. Two databases comprised of 1,300 manually annotated sclera images, being 1,000 sclera images from the MICHE-I database and 300 sclera images from the UBIRIS.v2.

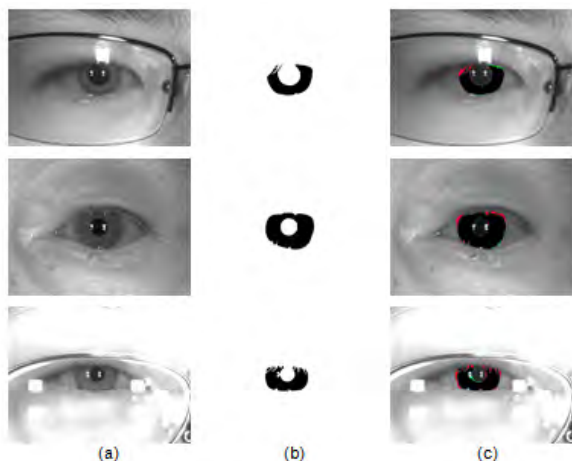


FIGURE 10. Sample images of the iris segmentation results using the proposed Sclera-Net on the CASIA v4.0 distance database. (a) Original image, (b) annotated image, and (c) segmentation result, where the white regions represent the true positive pixels, whereas the green and red regions represent the false positive and false negative pixels, respectively.

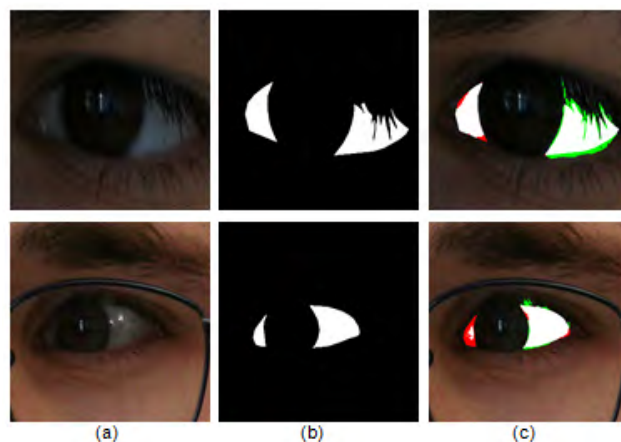


FIGURE 12. Sample images of incorrect segmentation results using Sclera-Net on the UBIRIS.v2 database [27] while looking in different directions. (a) Original image, (b) annotated image, and (c) segmentation result, where the white regions represent the true positive pixels, whereas the green and red regions represent the false positive and false negative pixels, respectively.

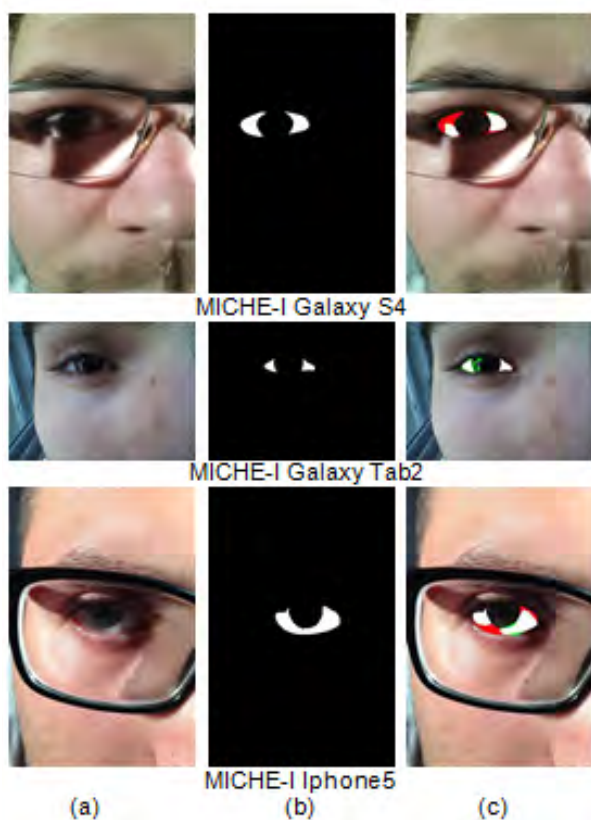


FIGURE 11. Sample images with incorrect segmentation results using Sclera-Net on the MICHE-I sub-databases [26] while looking in different directions. (a) Original image, (b) annotated image, and (c) segmentation result, where the white regions represent the true positive pixels, and the green and red regions represent the false positive and false negative pixels, respectively.

The examples of the correct segmentation results of Sclera-Net on MICHE-I and UBIRIS.v2 are shown in Figures 7 and 8, respectively. Here, the green color pixels

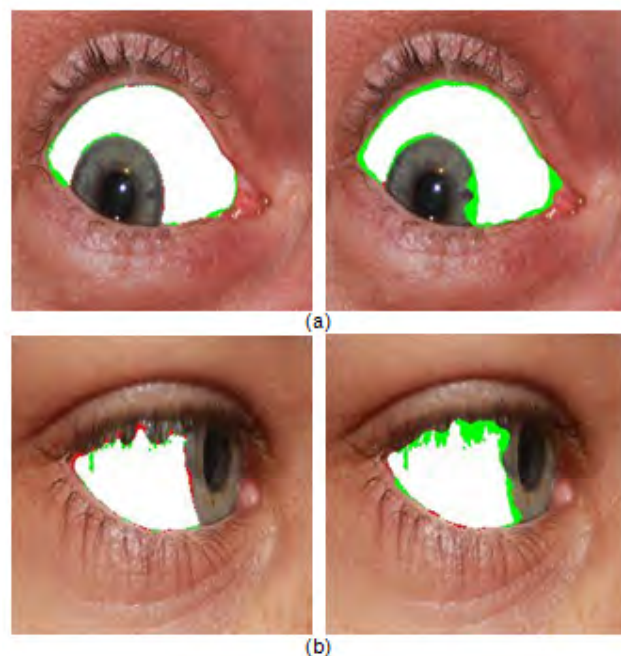


FIGURE 13. Comparison of segmentation results obtained by Sclera-Net (left) and SegNet [21] (right) on images with (a) sclera thin boundary (b) sclera occluded with eye lashes.

represent the false positive cases, red color pixels represent the false negative cases, and white color pixels represent the true positive cases. We obtained some incorrect results of sclera segmentation with our proposed Sclera-Net method. Figures 11 and 12 show the incorrect results for sclera segmentation on MICHE-I and UBIRIS.v2, respectively. The incorrect results are mainly due to ambient light, reflections from skin, or other kinds of noises whose pixels are similar to those of the sclera. Some errors are due to very light-colored pixel values similar to the non-sclera pixel values.



FIGURE 14. (SegNet [21]) the first 64 channel features from Group 4 before the 4th max-pooling (Pool-4).

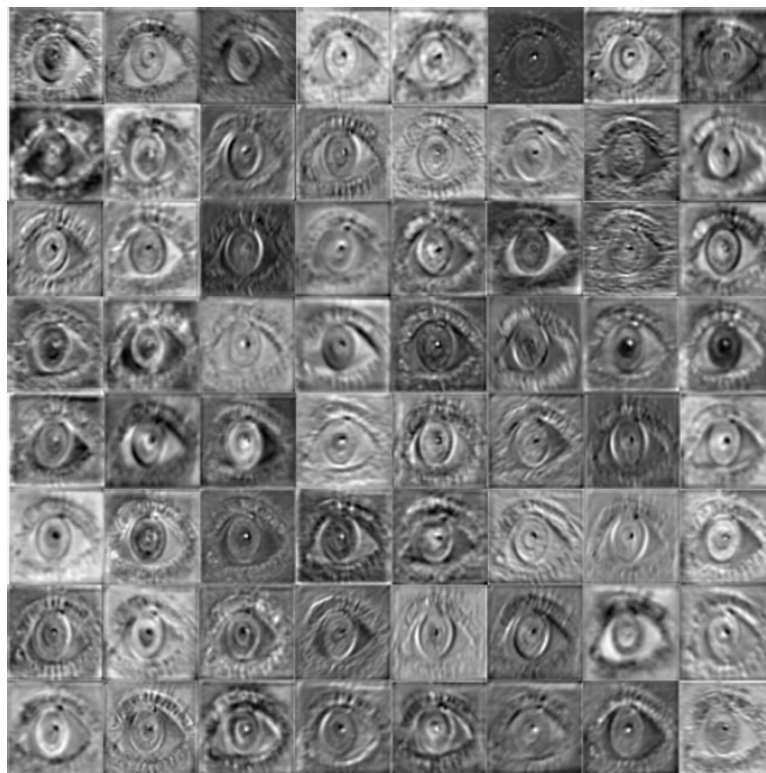


FIGURE 15. (Sclera-Net) the first 64 channel features from Group 4 before the 4th max-pooling (Pool-4).

For a fair comparison of the proposed segmentation method with the existing researches on the MICHE-I and UBIRIS.v2 datasets, the *PRF* protocol is used. The *PRF*

metrics evaluates the segmentation performance based on the ground truth masks. Tables 6 and 7 list the *PRF* values of the segmentation methods based on MICHE-I [26]

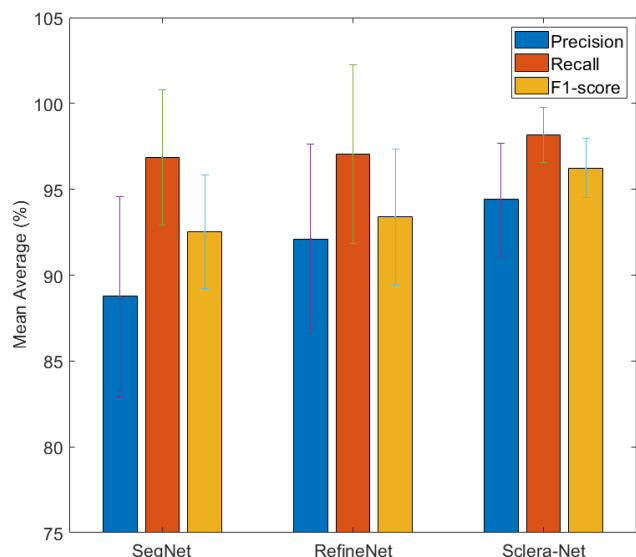


FIGURE 16. Mean and standard deviation of the proposed method and existing alternatives in terms of precision, recall and F1-score based on SBVPI database [25].

and UBIRIS.v2 [27], respectively. Additionally, the results in Tables 6 and 7 are presented through bar graphs in Figures 17 and 18. The results show that Sclera-Net shows optimal segmentation performance as compared to the other methods. It can be seen that Sclera-Net provides higher *F1-score* measures than the previous methods. Hence, our

proposed method has outperformed the other methods in all defined criteria and metrics.

4) SCLERA-NET FOR OTHER EYE REGION SEGMENTATION USING OPEN DATABASES

Sclera-Net not only shows remarkable results for the segmentation of sclera and non-sclera classes, but it can also deal with the segmentation of other eye regions such as iris. Iris segmentation is very useful in different applications of biometric recognition. In this section, we evaluated two well-known iris datasets: noisy iris challenge evaluation (NICE) II [28] and Chinese Academy of Sciences (CASIA) v4.0 distance [29] for iris segmentation. The NICE-II dataset includes 1000 eye images with noisy iris, irregular illumination, motion blurs, partial open eyes, off-angle eyes, occluded iris, etc. The annotated images for this dataset are publicly available for comparison and evaluation purposes. The CASIA v4.0 distance database contains 2567 images from 142 participants. The images are captured at a 3 m distance from the camera. The ground truth for CASIA v4.0 distance is obtained from [64]. Figures 9 and 10 show the segmented results on the NICE II and CASIA v4.0 distance datasets, respectively.

Tables 8 and 9 list the comparative results for iris segmentation using the NICE-II and CASIA v4.0 distance datasets, respectively. The sample images (Figures 9 and 10) and comparative analysis (listed in Tables 8 and 9) show that the proposed Sclera-Net shows less segmentation error rate than the previous methods.

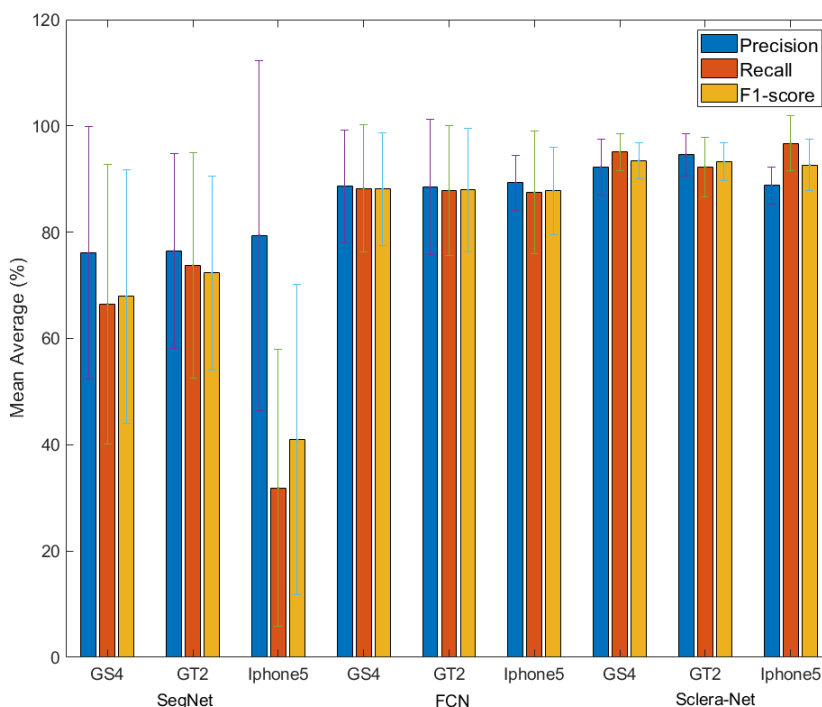


FIGURE 17. Mean and standard deviation of the proposed method and existing alternatives in terms of precision, recall and F1-score based on MICHE-I database [26].

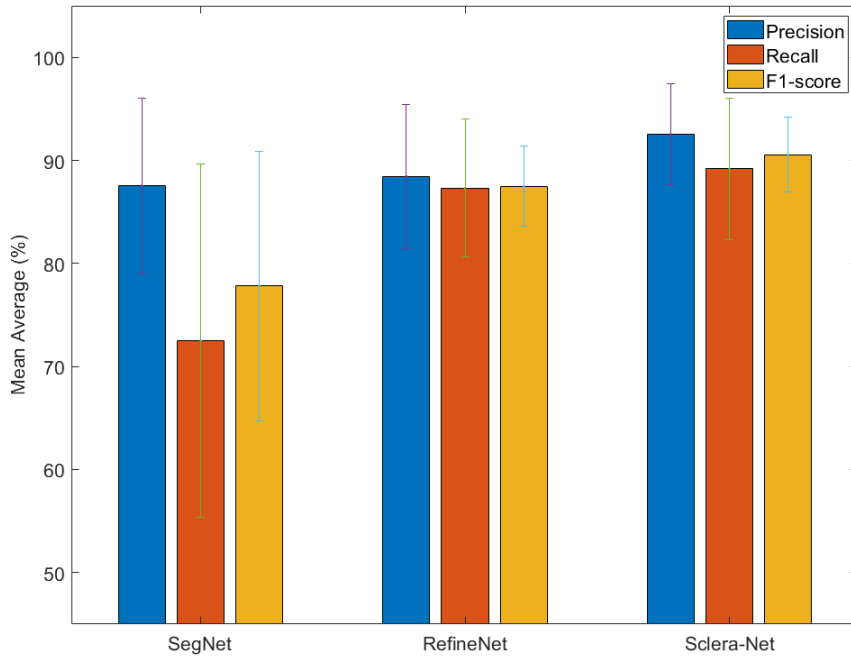


FIGURE 18. Mean and standard deviation of the proposed method and existing alternatives in terms of precision, recall and F1-score based on UBIRIS.v2 database [27].

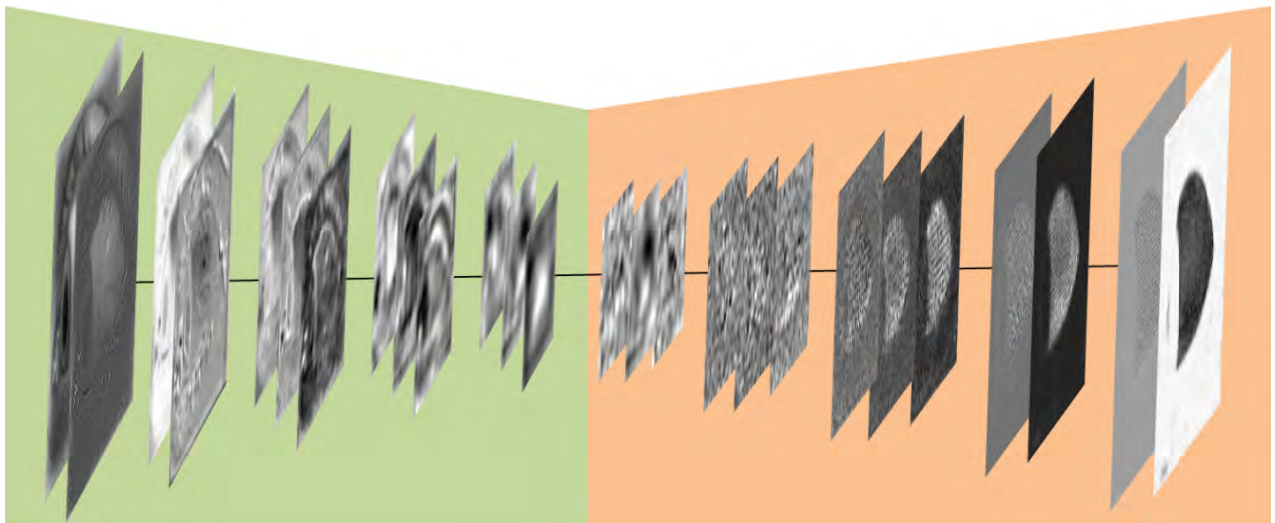


FIGURE 19. Example of best case with details on each layer of the proposed Sclera-Net model.

V. RESULT ANALYSIS AND DISCUSSION

In this study, the IM and NIM based on residual connectivity from the previous layers is used to achieve better semantic segmentation. The previous CNN based methods continuously eliminate high frequency information while the information flows through the convolution layers. This loss of information can adversely affect the performance of the network in terms of losing important and useful information. Therefore, for retaining information from the previous layers, we used the concept of IM and NIM based on RBBs, which import features from the previous layers. Figure 2 shows the

overall diagram of the encoder and decoder with IM- and NIM-based RBBs. The performance of Sclera-Net can be visualized from the results of Figures 5, 7, and 8 obtained using SBVPI [25], MICHE-I [26], and UBIRIS.v2 [27] databases, respectively. Additionally, the performance can be confirmed from the results of sclera segmentation mentioned in Tables 5, 6, and 7 obtained using SBVPI, MICHE-I, and UBIRIS.v2 datasets, respectively. Here, we have used three different metrics for performance evaluation: average segmentation error, *IoU*, and PRF. Besides correct segmentation results, we have observed some incorrect segmentation

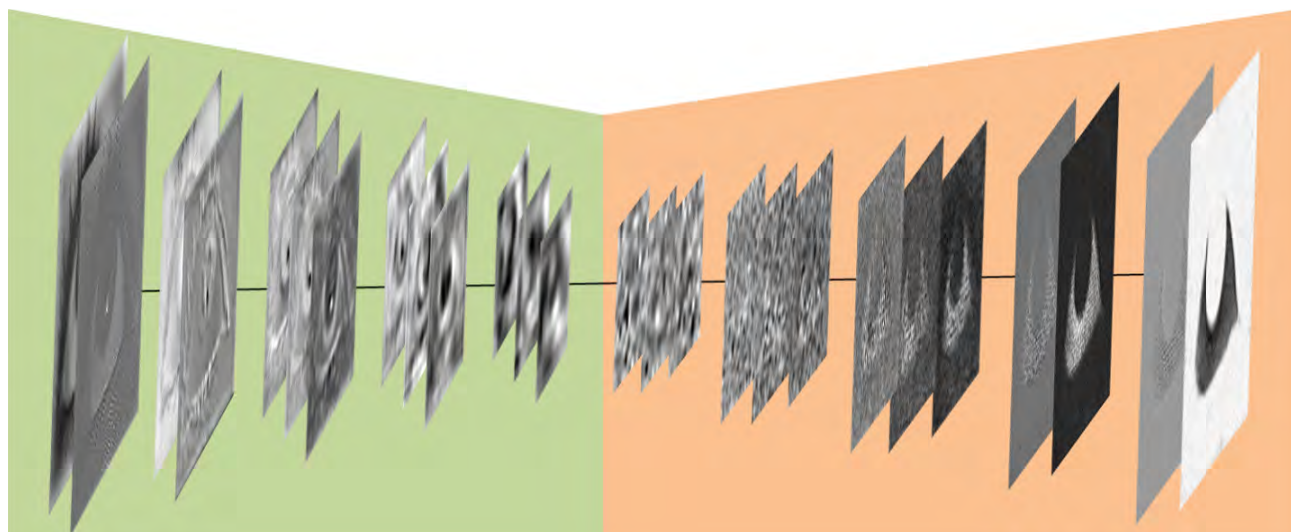


FIGURE 20. Example of medium case with details on each layer of the proposed Sclera-Net model.

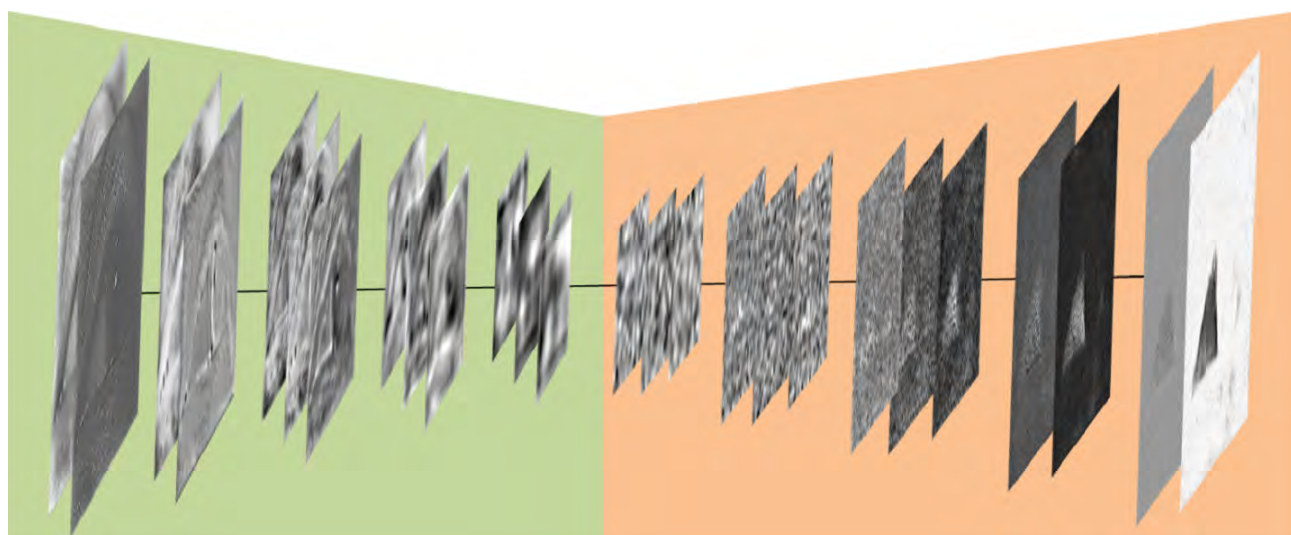


FIGURE 21. Example of worst case with details on each layer of the proposed Sclera-Net model.

results with our proposed method. It is observed that environmental light, reflections from skin, or other kind of noises, whose pixels are similar to the sclera, are the main causes of error. Our experiments have proved that the proposed sclera segmentation method is not only suitable for sclera segmentation, but also shows outstanding performance on other eye region segmentation, such as the iris. The performance of the proposed method for iris segmentation is evident from the results of Figures 9 and 10 obtained using NICE-II and CASIA distance v4.0, respectively. Additionally, the performance can be confirmed from the results of iris segmentation mentioned in Tables 8 and 9 obtained using NICE-II and CASIA distance v4.0, respectively. Here, we have used the average segmentation error metric for the performance evaluation of iris segmentation because it is considered as a

protocol by the previous methods using NICE-II and CASIA distance v4.0. From the various datasets used in different environments, it can be concluded that the proposed method showed outstanding results even on iris segmentation and outperformed other previous methods.

To explain the power of the residual skip connections, we compared the performance of the proposed residual connectivity approach with previous non-residual-based approaches through reference convolutional features. These reference convolutional features were obtained from Group 4 (Table 10) in both Sclera-Net and SegNet. Note that the output features in Group 4 after Pool-3 contains 512 channels, and for simplicity, the first 64 channels (1st to 64th) are visualized. Group 4 features are the 2nd-last pooling index features. Noticeable visual differences are presented by

TABLE 10. The Sclera-Net encoder based on residual connections.

	Layer types	Number of filters	Size of feature map	Size of kernel	Size of stride	Size of padding
	Image input layer		224 (height) × 224 (width) × 1 (channel)			
Group 1	Conv-1_1 (1st convolutional layer)	64	224 × 224 × 64	3 × 3	1 × 1	1 × 1
	ReLU-1_1		224 × 224 × 64			
	Conv-1_2 (2nd convolutional layer)	64	224 × 224 × 64	3 × 3	1 × 1	1 × 1
	ReLU-1_2		224 × 224 × 64			
	Pool-1	1	112 × 112 × 64	2 × 2	2 × 2	0 × 0
	Conv-2_1 (3rd convolutional layer)	128	112 × 112 × 128	3 × 3	1 × 1	1 × 1
	ReLU-2_1		112 × 112 × 128			
Group 2	Conv-2_2 (4th convolutional layer)	128	112 × 112 × 128	3 × 3	1 × 1	1 × 1
	ReLU-2_2		112 × 112 × 128			
	RNB-1	128	112 × 112 × 128	1 × 1		
	Add1 = Group-2-1/out + RNB-1/out	128	112 × 112 × 128			
	ReLU-2_2		112 × 112 × 128			
	Pool-2	1	56 × 56 × 128	2 × 2	2 × 2	0 × 0
	Conv-3_1 (5th convolutional layer)	256	56 × 56 × 256	3 × 3	1 × 1	1 × 1
	ReLU-3_1		56 × 56 × 256			
	Conv-3_2 (6th convolutional layer)	256	56 × 56 × 256	3 × 3	1 × 1	1 × 1
Group 3	RNB-2	256	56 × 56 × 256	1 × 1		
	Add2 = Group-3-1/out + RNB-2/out	256	56 × 56 × 256			
	ReLU-3_2		56 × 56 × 256			
	Conv-3_3 (7th convolutional layer)	256	56 × 56 × 256	3 × 3	1 × 1	1 × 1
	Add3 = Conv-3_3/out + Add2/out	256	56 × 56 × 256			
	ReLU-3_3		56 × 56 × 256			
	Pool-3	1	28 × 28 × 256	2 × 2	2 × 2	0 × 0
	Conv-4_1 (8th convolutional layer)	512	28 × 28 × 512	3 × 3	1 × 1	1 × 1
	ReLU-4_1		28 × 28 × 512			
	Conv-4_2 (9th convolutional layer)	512	28 × 28 × 512	3 × 3	1 × 1	1 × 1
Group 4	RNB-3	512	28 × 28 × 512	1 × 1		
	Add4 = Group-4-1/out + RNB-3/out	512	28 × 28 × 512			
	ReLU-4_2		28 × 28 × 512			
	Conv-4_3 (10th convolutional layer)	512	28 × 28 × 512	3 × 3	1 × 1	1 × 1
	Add5 = Conv-4_3/out + Add4/out	512	28 × 28 × 512			
	ReLU-4_3		28 × 28 × 512			
	Pool-4	1	14 × 14 × 512	2 × 2	2 × 2	0 × 0
	Conv-5_1 (11th convolutional layer)	512	14 × 14 × 512	3 × 3	1 × 1	1 × 1
	ReLU-5_1		14 × 14 × 512			
	Conv-5_2 (12th convolutional layer)	512	14 × 14 × 512	3 × 3	1 × 1	1 × 1
Group 5	Add6 = Conv-5_2/out + Pool4/out	512	14 × 14 × 512			
	ReLU-5_2		14 × 14 × 512			
	Conv-5_3 (13th convolutional layer)	512	14 × 14 × 512	3 × 3	1 × 1	1 × 1
	Add7 = Conv-5_3/out + Add6/out	512	14 × 14 × 512			
	ReLU-5_3		14 × 14 × 512			
	Pool-5	1	7 × 7 × 512	2 × 2	2 × 2	0 × 0

Note: Conv, ReLU, Pool, and Add indicate convolution, rectified linear unit, pooling, and element-wise addition layers, respectively.

these features. With a careful analysis of the output (as per Figures 14 and 15), it can be observed that the real power of the residual skip connections is evident from the visual features in Group 4 (Table 10) for both non-residual-based SegNet (Figure 14) and residual-based Sclera-Net (Figure 15). It can be seen from the figures that the Group 4 features

from SegNet are significantly noisier than those from Sclera-Net, which can reduce errors in detecting the correct sclera pixels.

To further confirm the strength of the residual skip connections, we compared the SegNet [21] sclera segmentation with the Sclera-Net results. The segmentation

TABLE 11. The Sclera-Net decoder based on residual connections.

	Layer types	Number of filters	Size of feature map	Size of kernel	Size of stride	Size of padding
DGroup 5	Unpool-5		14 × 14 × 512			
	DConv-5_3	512	14 × 14 × 512	3 × 3	1 × 1	1 × 1
	(1st convolutional layer)					
	DReLU-5_3		14 × 14 × 512			
	DConv-5_2	512	14 × 14 × 512	3 × 3	1 × 1	1 × 1
	(2nd convolutional layer)					
	Add8 = DConv-5_2/out + Unpool-5/out	512	14 × 14 × 512			
	DReLU-5_2		14 × 14 × 512			
	DConv-5_1	512	14 × 14 × 512	3 × 3	1 × 1	1 × 1
(3rd convolutional layer)						
Add9 = DConv-5_1/out + Add8/out	512	14 × 14 × 512				
DReLU-5_1		14 × 14 × 512				
DGroup 4	Unpool-4		28 × 28 × 512			
	DConv-4_3	512	28 × 28 × 512	3 × 3	1 × 1	1 × 1
	(4th convolutional layer)					
	DReLU-4_3		28 × 28 × 512			
	DConv-4_2	512	28 × 28 × 512	3 × 3	1 × 1	1 × 1
	(5th convolutional layer)					
	Add10 = DConv-4_2/out + Unpool4/out	512	28 × 28 × 512			
	DReLU-4_2		28 × 28 × 512			
	DConv-4_1	256	28 × 28 × 256	3 × 3	1 × 1	1 × 1
	(6th convolutional layer)					
	RNB-4	256	28 × 28 × 256	1 × 1		
Add11 = DGroup 4_1/out + RNB-4/out	256	28 × 28 × 256				
DReLU-4_1		28 × 28 × 256				
DGroup 3	Unpool-3	1	56 × 56 × 256			
	DConv-3_3	256	56 × 56 × 256	3 × 3	1 × 1	1 × 1
	(7th convolutional layer)					
	DReLU-3_3		56 × 56 × 256			
	DConv-3_2	256	56 × 56 × 256	3 × 3	1 × 1	1 × 1
	(8th convolutional layer)					
	Add12 = DConv-3_2/out + Unpool3/out	256	56 × 56 × 256			
	DReLU-3_2		56 × 56 × 256			
	DConv-3_1	128	56 × 56 × 128	3 × 3	1 × 1	1 × 1
	(9th convolutional layer)					
RNB-5	128	56 × 56 × 128	1 × 1			
Add13 = DConv-3_1/out + RNB-5/out	128	56 × 56 × 128				
DReLU-3_1		56 × 56 × 128				
DGroup 2	Unpool-2	1	112 × 112 × 128			
	DConv-2_2	128	112 × 112 × 128	3 × 3	1 × 1	1 × 1
	(10th convolutional layer)					
	DReLU-2_2		112 × 112 × 128			
	DConv-2_1	64	112 × 112 × 64	3 × 3	1 × 1	1 × 1
	(11th convolutional layer)					
	RNB-6	64	112 × 112 × 64	1 × 1		
Add14 = DConv-2_1/out + RNB-6/out	64	112 × 112 × 64				
DReLU-2_1		28 × 28 × 512				
DGroup 1	Unpool-1	1	224 × 224 × 64			
	DConv-1_2	64	224 × 224 × 64	3 × 3	1 × 1	1 × 1
	(12th convolutional layer)					
	DReLU-1_2		224 × 224 × 64			
	DConv-1_1	2	224 × 224 × 2	3 × 3	1 × 1	1 × 1
(13th convolutional layer)						
DReLU-1_1		224 × 224 × 2				

Note: Conv, ReLU, Pool, and Add indicate convolution, rectified linear unit, pooling, and element-wise addition layers, respectively.

results obtained with the proposed residual features show a finer and thinner boundary compared with the non-residual method, which substantially reduces the rate of error in the case of the Sclera-Net method. The proposed method is equally able to separate the thin iris boundary from sclera and robust when the sclera images are occluded with eyelashes. These important observations can be visualized from Figure 13.

VI. CONCLUSIONS AND FUTURE WORKS

In this research, Sclera-Net is proposed for the semantic segmentation of the sclera and other eye regions. It is based on the IM and NIM of the RBBs in both the encoder and decoder parts of the network. This network enhances the accuracy by enabling high-frequency information to pass through the network. This method is useful for the true boundary segmentation of different eye regions in non-ideal and challenging

situations. Preprocessing is not performed in this method as in the case of previously proposed methods. Experiments for sclera segmentation were conducted on three sclera and two iris image datasets. Results showed that our proposed method outperformed previous methods of end-to-end segmentation. For the future work, the network will be optimized and the number of parameters and layers will be reduced to enable this method to work efficiently in smart phones. In addition, its applicability in other segmentation tasks such as in crop diseases or medical images will be studied.

APPENDIX ADDITIONAL FIGURES (F) & TABLES (T)

See Figures 11–21 and Tables 10 and 11.

REFERENCES

- [1] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior, "Authentication and biometrics," in *Guide to Biometrics*, New York, NY, USA: Springer, 2004, pp. 17–29.
- [2] A. Rattani and M. Tistarelli, "Robust multi-modal and multi-unit feature level fusion of face and iris biometrics," in *Proc. IEEE ICB*, Alghero, Italy, Jun. 2009, pp. 960–969.
- [3] K.-A. Toh, H.-L. Eng, Y.-S. Choo, Y.-L. Cha, W.-Y. Yau, and K.-S. Low, "Identity verification through palm vein and crease texture," in *Proc. IEEE ICB*, Hong Kong, China, Jan. 2006, pp. 546–553.
- [4] F. Sadikoglu and S. Uzelaltinbulat, "Biometric retina identification based on neural network," *Procedia Comput. Sci.*, vol. 102, pp. 26–33, Dec. 2016.
- [5] T. A. Soomro, T. M. Khan, M. A. U. Khan, J. Gao, M. Paul, and L. Zheng, "Impact of ICA-based image enhancement technique on retinal blood vessels segmentation," *IEEE Access*, vol. 6, pp. 3524–3538, 2018.
- [6] D. A. Dharmawan, D. Li, B. P. Ng, and S. Rahardja, "A new hybrid algorithm for retinal vessels segmentation on fundus images," *IEEE Access*, vol. 7, pp. 41885–41896, Mar. 2019.
- [7] R. Derakhshani, A. Ross, and S. Crihalmeanu, "A new biometric modality based on conjunctival vasculature," in *Proc. ICANNE*, St. Louis, MO, USA, Nov. 2006, pp. 1–8.
- [8] Z. Zhou, E. Y. Du, N. L. Thomas, and E. J. Delp, "A new human identification method: Sclera recognition," *IEEE Trans. Syst. Man Cybern. Syst. Hum.*, vol. 42, no. 3, pp. 571–583, May 2012.
- [9] A. Das, U. Pal, M. Blumenstein, and M. A. F. Ballester, "Sclera recognition—A survey," in *Proc. IAPR ACPR*, Naha, Japan, Nov. 2013, pp. 917–921.
- [10] P. Tower, "The fundus oculi in monozygotic twins: Report of six pairs of identical twins," *AMA Arch. Ophthalmol.*, vol. 54, no. 2, pp. 225–239, Aug. 1955.
- [11] A. M. Joussen, "Vascular plasticity—The role of the angiopoietins in modulating ocular angiogenesis," *Graefe's Arch. Clinical Exp. Ophthalmol.*, vol. 239, no. 12, pp. 972–975, Dec. 2001.
- [12] H. Kobayashi and S. Kohshima, "Unique morphology of the human eye," *Nature*, vol. 387, no. 6635, pp. 767–768, Jun. 1997.
- [13] Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp, "Multimodal eye recognition," *Proc. SPIE*, vol. 7708, Apr. 2010, Art. no. 770806.
- [14] C. G. De Moraes, J. V. Jasien, S. Simon-Zoula, J. M. Liebmann, and R. Ritch, "Visual field change and 24-hour IOP-related profile with a contact lens sensor in treated glaucoma patients," *Ophthalmology*, vol. 123, pp. 744–753, Jan. 2016.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, San Diego, CA, USA, May 2015, pp. 1–14.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, Jun. 2016, p. 770–778.
- [17] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein, "SSBC 2015: Sclera segmentation benchmarking competition," in *Proc. BTAS*, Arlington, VA, USA, Sep. 2015, pp. 1–6.
- [18] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein, "SSRBC 2016: Sclera segmentation and recognition benchmarking competition," in *Proc. ICB*, Halmstad, Sweden, Jun. 2016, pp. 1–6.
- [19] A. Das, U. Pal, M. A. Ferrer, M. Blumenstein, D. Štepec, P. Rot, Ž. Emeršič, P. Peer, V. Štruc, S. V. A. Kumar, and B. S. Harish, "SSERBC 2017: Sclera segmentation and eye recognition benchmarking competition," in *Proc. IEEE ICB*, Denver, CO, USA, Oct. 2017, pp. 742–747.
- [20] A. Das, U. Pal, M. A. Ferrer, M. Blumenstein, D. Štepec, P. Rot, Ž. Emeršič, P. Peer, V. Štruc, "SSBC 2018: Sclera segmentation benchmarking competition," in *Proc. ICB*, Gold Coast, QLD, Australia, Feb. 2018, pp. 303–308.
- [21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [22] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 5168–5177.
- [23] PRML. (2019). *Sclera-Net CNN Model (Sclera-Net) With Algorithm*. Accessed: Apr. 19, 2019. [Online]. Available: <http://pr.gachon.ac.kr/>
- [24] P. Rot, "Deep learning methods for biometric recognition based on eye information," M.S. Thesis, FCIS, Uni-LJ, Ljubljana, Slovenia, 2018.
- [25] *SBVPI Dataset*. Accessed: Feb. 12, 2019. [Online]. Available: <http://sclera.fri.uni-lj.si/>
- [26] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols," *Pattern Recognit. Lett.*, vol. 57, pp. 17–23, May 2015.
- [27] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.
- [28] *NICE:II - Noisy Iris Challenge Evaluation-Part II*. Accessed: Apr. 26, 2019. [Online]. Available: <http://nice2.di.ubi.pt/>
- [29] *Biometrics Ideal Test*. Accessed: Feb. 26, 2019. [Online]. Available: <http://biometrics.idealtest.org/dbDetailForUser.do?id=4>
- [30] R. Derakhshani and A. Ross, "A texture-based neural network classifier for biometric identification using ocular surface vasculature," in *Proc. IJCNN*, Orlando, FL, USA, Aug. 2007, pp. 2982–2987.
- [31] S. Crihalmeanu, A. Ross, and R. Derakhshani, "Enhancement and registration schemes for Matching conjunctival vasculature," in *Proc. ICB*, Alghero, Italy, Jun. 2009, pp. 1240–1249.
- [32] N. L. Thomas, Y. Du, and Z. Zhou, "A new approach for sclera vein recognition," in *Proc. SPIE*, vol. 7708, Apr. 2010, Art. no. 770805.
- [33] K. Oh and K. A. Toh, "Extracting sclera features for cancelable identity verification," in *Proc. ICB*, New Delhi, India, Mar. 2012, pp. 245–250.
- [34] Y. Lin, E. Y. Du, Z. Zhou, and N. L. Thomas, "An efficient parallel approach for sclera vein recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 2, pp. 147–157, Feb. 2014.
- [35] Z. Zhou, E. Y. Du, and N. L. Thomas, "A comprehensive sclera image quality measure," in *Proc. ICARCV*, Singapore, Dec. 2010, pp. 638–643.
- [36] Z. Zhou, E. Y. Du, C. Belcher, N. L. Thomas, and E. J. Delp, "Quality fusion based multimodal eye recognition," in *Proc. IEEE SMC*, Seoul, South Korea, Oct. 2012, pp. 1297–1302.
- [37] M. H. Khosravi and R. Safabakhsh, "Human eye sclera detection and tracking using a modified time-adaptive self-organizing map," *Pattern Recognit.*, vol. 41, no. 8, pp. 2571–2593, Aug. 2008.
- [38] S. Crihalmeanu and A. Ross, "Multispectral scleral patterns for ocular biometric recognition," *Pattern Recognit. Lett.*, vol. 33, no. 14, pp. 1860–1869, Oct. 2012.
- [39] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein, "Sclera recognition using dense-SIFT," in *Proc. ISDA*, Bangi, Malaysia, Dec. 2013, pp. 74–79.
- [40] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein, "A new efficient and adaptive sclera recognition system," in *Proc. IEEE CIBIM*, Orlando, FL, USA, Dec. 2014, pp. 1–8.
- [41] K. V. Delna, K. A. Sneha, and R. P. Aneesh, "Sclera vein identification in real time using single board computer," in *Proc. ICNGIS*, Kottayam, India, Sep. 2016, pp. 1–5.
- [42] S. Alkassar, W. L. Woo, S. S. Dlay, and J. A. Chambers, "Enhanced segmentation and complex-sclera features for human recognition with unconstrained visible-wavelength imaging," in *Proc. ICB*, Halmstad, Sweden, Jun. 2016, pp. 1–8.
- [43] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, Jun. 2018.
- [44] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018.

- [45] P. Radu, J. Ferryman, and P. Wild, "A robust sclera segmentation algorithm," in *Proc. IEEE BTAS*, Arlington, VA, USA, Sep. 2015, pp. 1–6.
- [46] P. Rot, Ž. Emeršič, V. Struc, P. Peer, "Deep multi-class eye segmentation for ocular biometrics," in *Proc. IEEE IWOB*, San Carlos, CA, USA, Jul. 2018, pp. 1–8.
- [47] D. R. Lucio, R. Laroca, E. Severo, A. S. B. Jr, and D. Menotti, "Fully convolutional networks and generative adversarial networks applied to sclera segmentation," in *Proc. IEEE BTAS*, Los Angeles, CA, USA, Oct. 2018, pp. 1–7.
- [48] M. Teichmann, M. Weber, M. Zöllner, R. Cipolla, and R. Urtasun, "Multi-Net: Real-time joint semantic reasoning for autonomous driving," in *Proc. IEEE IV*, Changshu, China, Jun. 2018, pp. 1013–1020.
- [49] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [50] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 5967–5976.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. ECCV*, Amsterdam, Netherlands, Oct. 2016, pp. 630–645.
- [52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, Lille, France, Jul. 2015, pp. 448–456.
- [53] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. ICML*, Haifa, Israel, Jun. 2010, p. 807.
- [54] *Convolutional Neural Network*. Accessed: Dec. 19, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Convolutional_neural_network
- [55] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. AISTATS*, Fort Lauderdale, FL, USA, Apr. 2011, p. 315.
- [56] (2018). *Matlab*. Accessed: Feb. 12, 2019. [Online]. Available: <https://uk.mathworks.com/help/matlab/release-notes.html>
- [57] J. Shijie, W. Ping, J. Peiyi, and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," in *Proc. CAC*, Jinan, China, Oct. 2017, pp. 416–4170.
- [58] Z. Wang, L. Du, J. Mao, B. Liu, and D. Yang, "SAR target detection based on SSD with data augmentation and Transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 150–154, Jan. 2019.
- [59] *Optimization: Stochastic Gradient Descent*. Accessed: Feb. 12, 2019. [Online]. Available: <http://ufldl.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/>
- [60] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. IEEE ICCV*, Santiago, Chile, Dec. 2015, pp. 2650–2658.
- [61] L. Fan, H. Kong, W.-C. Wang, and J. Yan, "Semantic segmentation with global encoding and dilated decoder in street scenes," *IEEE Access*, vol. 6, pp. 50333–50343, 2018.
- [62] L. Ding, K. Zhao, X. Zhang, X. Wang, and J. Zhang, "A lightweight U-net architecture multi-scale convolutional network for pediatric hand bone segmentation in X-ray image," *IEEE Access*, vol. 7, pp. 68436–68445, 2019.
- [63] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. ICML*, Pittsburgh, PA, USA, Jun. 2006, pp. 233–240.
- [64] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in *Proc. IEEE ICB*, Halmstad, Sweden, Jun. 2016, pp. 1–8.
- [65] M. Haindl and M. Krupička, "Unsupervised detection of non-iris occlusions," *Pattern Recognit. Lett.*, vol. 57, pp. 60–65, May 2015.
- [66] Z. Zhao and A. Kumar, "An accurate iris segmentation framework under relaxed imaging constraints using total variation model," in *Proc. IEEE ICV*, Santiago, Chile, Dec. 2015, pp. 3828–3836.
- [67] M. Arsalan, H. Hong, R. Naqvi, M. Lee, M. Kim, D. Kim, C. Kim, and K. Park, "Deep learning-based iris segmentation for iris recognition in visible light environment," *Symmetry*, vol. 9, no. 11, pp. 1–25, Nov. 2017.
- [68] M. Arsalan, R. A. Naqvi, D. S. Kim, P. H. Nguyen, M. Owais, and K. R. Park, "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. vol. 5, p. 1501, May 2018.
- [69] C. Tan and A. Kumar, "Towards online iris and periocular recognition under relaxed imaging constraints," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3751–3765, Oct. 2013.



RIZWAN ALI NAQVI received the B.S. degree in computer engineering from COMSATS University, Pakistan, in 2008, the M.S. degree in electrical engineering from Karlstad University, Sweden, in 2011, and the Ph.D. degree in electronics and electrical engineering from Dongguk University, South Korea, in 2018.

From 2011 to 2012, he was Lecturer with the Computer Science Department, Sharif College of Engineering and Technology, Pakistan. He joined the Faculty of Engineering and Technology, The Superior College, Pakistan, as a Senior Lecturer, in 2012. He is currently a Postdoctoral Researcher with Gachon University, South Korea. His research interests include gaze tracking, biometrics, computer vision, artificial intelligence, machine learning, deep learning, and medical imaging analysis.



WOONG-KEE LOH received the B.S., M.S., and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1991, 1993, and 2001, respectively. He was a Visiting Scholar with the Department of Computer Science and Engineering, University of Minnesota, USA. He is currently an Associate Professor with the Department of Software, Gachon University, South Korea. His research interests include massively parallel large-scale data mining, nearest neighbor search in road networks, biometric image recognition, and fault detection and classification using sensor streaming data. He has served as a Program and an Organizing Committee Member for many international conferences, including DaWaK 2007–2008, CIKM 2009, PAKDD 2011–2014, ICDE 2015, SSTD 2015, DASFAA 2017, and BigComp 2017–2018 and 2020.

...